IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 1 May 2023, accepted 24 May 2023, date of publication 5 June 2023, date of current version 15 June 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3282876

RESEARCH ARTICLE

Multiscopic CPSS for Independent Block-Design Test Based on Hand–Object Interaction Recognition With Visual Attention

ADNAN RACHMAT ANOM BESARI[1,2], (Graduate Student Member, IEEE), AZHAR AULIA SAPUTRA[1], (Member, IEEE), TAKENORI OBO[1], (Member, IEEE), KURNIANINGSIH[3], (Senior Member, IEEE), AND NAOYUKI KUBOTA[1], (Senior Member, IEEE)

[1]Graduate School of Systems Design, Tokyo Metropolitan University, Tokyo 191-0065, Japan
[2]Department of Information and Computer Engineering, Politeknik Elektronika Negeri Surabaya, Surabaya 60111, Indonesia
[3]Department of Electrical Engineering, Politeknik Negeri Semarang, Semarang 50275, Indonesia

Corresponding authors: Naoyuki Kubota (kubota@tmu.ac.jp) and Adnan Rachmat Anom Besari (anom@pens.ac.id)

This work was partially supported by the Japan Science and Technology Agency (JST), Moonshot R&D under Grant JPMJMS2034, and the Tokyo Metropolitan University (TMU) Local 5G Research Support.

ABSTRACT This paper introduces a multiscopic cyber-physical-social system (CPSS) to bridge the gap between independent rehabilitation in physical and cognitive aspects. Specifically, we focus on hand–object interaction (HOI) recognition with visual attention for the block-design test (BDT). The proposed framework utilizes three levels which consist of microscopic, mesoscopic, and macroscopic models. In the microscopic model, a hand-tracking vision captures hand-skeletal data and finger joint angle features, enabling the estimation of physical hand postures. In the mesoscopic model, an egocentric vision with an eye tracker records to hand and eye movements, allowing for the symbolic representation of hand-eye coordination through hand gestures and visual attention focus during the test. An evaluation vision system employs color feature classification in the macroscopic model to determine whether the design matches the given task. Through the first eight designs of WAIS-IV BDT with two scenarios, the system successfully measures human behavior from the physical to the cognitive domain. The experiment involving eight healthy participants investigates the relationship between physical measurement and cognitive evaluation. Regression and correlation analyses between the dominant and non-dominant hands reveal that evaluation indices (task completion time, skewness-kurtosis of hand posture, attention to pattern and blocks) can indicate improvement during BDT. The outcomes of this study have significant implications for clinicians and researchers, providing valuable information that is typically unavailable in clinical settings. The proposed multiscopic CPSS framework holds promise for advancing independent rehabilitation practices. Code and datasets are available online at https://github.com/anom-tmu/bdt-multiscopic.

INDEX TERMS Kohs blocks design, self-rehabilitation, visual attention, hand-eye coordination.

I. INTRODUCTION

Physical and cognitive rehabilitation is crucial in recovering patients with neurological conditions like stroke. However, this rehabilitation process faces challenges, especially for hand stroke survivors who often experience eyesight impairments. These impairments, such as visual field loss and double vision, significantly affect hand movements, leading to difficulties in daily activities like reaching and grasping objects [1]. With the convenience and safety concerns brought about by the COVID-19 pandemic, patients have increasingly opted for home-based rehabilitation. However, limited

The associate editor coordinating the review of this manuscript and approving it for publication was K. C. Santosh.

TABLE 1. The research and development of technology for block design test (2018-2022).

No.	Research	Application (Sensor Types)	Methods (Contributions)
1.	Cha *et al.* (2018) [15]	Overhead video camera.	Automate classification using machine learning.
2.	Rogers *et al.* (2019) [30]	*Elements* virtual rehabilitation.	Access the efficacy of virtual rehabilitation approach using *Elements* for upper-limb skills.
3.	Averbukh *et. al.* (2019) [29]	Virtual reality.	Analyze the influence of the present experience in virtual reality on the key of intelligent tasks and the fundamentals of visualization system user activity.
4.	Cha *et al.* (2020) [28]	Overhead video camera and wearable eye tracker.	Combine scene with gaze cameras using supervised learning algorithms to measure critical behaviors automatically.
5.	Wikström *et al.* (2020) [31]	Virtual reality.	Create a collaborative block design task intended to evaluate and quantify pair performance.
6.	Dunn *et al.* (2021) [27]	Wearable eye tracker.	Detect gaze location and frequency of consulting the pattern.
7.	Shigenaga and Nagamune (2022) [32]	Virtual reality, eye tracker, and hand tracker.	Measure the eye and hand actions during the test in a VR space.
8.	Our proposed method	Hand tracker, wearable eye tracker, and over-table camera.	HOI with visual attention using a multiscopic approach to support physical-cognitive rehabilitation.

therapist availability, long distances, and privacy concerns have hindered in-person visits [2], highlighting the need for remote rehabilitation solutions. While telemedicine enables therapists to monitor patients remotely, the existing sensor technology and measurement systems do not adequately support remote rehabilitation.

Previous studies have explored cyber-physical systems (CPS) [3] to track hand therapy progress individually. Most studies have relied on contact methods where patients wear specialized devices to collect accurate data using flex, accelerometers, and hall-effect sensors [4]. However, these contact-based approaches have drawbacks, including high equipment costs and limitations in their usability. Noncontact techniques using computer vision systems for detecting human action recognition [5] have been investigated as an alternative. Nevertheless, these methods often overlook important social aspects such as privacy concerns, clinical justification, and user experience, which are critical for user acceptance. To overcome these limitations, egocentric vision, such as smart glasses, has emerged as a potential solution. Egocentric vision offers advantages such as privacy protection, mobility monitoring, and attention tracking during activities [6]. In post-stroke therapy, the egocentric vision has shown promise in hand-object interaction (HOI) recognition, outperforming full-body human interaction detection [7]. By analyzing images captured by cameras on the body, egocentric vision simplifies HOI recognition by focusing on hands and objects.

However, extending HOI recognition to 2D images poses challenges, particularly in accurately detecting hand-object contact. The lack of depth data in 2D images makes determining the precise location of hands and objects difficult. Although approaches like interaction point learning [8] have been proposed to address this issue, they are often limited to recognizing specific types of objects and may encounter difficulties detecting multiple objects. Moreover, these approaches do not consider individuals with visual impairments who face challenges performing hand movements and utilizing their recognition abilities effectively.

To overcome these challenges, we propose using cyber-physical-social systems (CPSS) [9] that seamlessly integrate physical and social spaces while harnessing data in the cyber domain. In rehabilitation, egocentric vision based on CPSS can potentially simultaneously monitor physical and cognitive aspects. This vision system utilizes hand skeleton estimation [10] and kinematic finger models [11] to assess an individual's physical condition. Additionally, an eye-tracking-equipped camera captures data on visual attention [12], [13], providing insights into cognitive abilities. Therefore, attention to hand movements and vision plays a significant role in understanding how individuals handle objects effectively.

This research focuses on the cube object used in the block design test (BDT) [14], a widely used neuropsychological assessment for evaluating visuospatial abilities. Traditionally, a neuropsychologist observes a person's accuracy, completion time, and overall problem-solving strategy and errors during the BDT [15]. However, subjective and qualitative assessments from a single perspective may be unreliable, calling for measurements from multiple perspectives. Table 1 highlights recent technological advancements related to the BDT. Therefore, we propose a new framework integrating multiple AI-based vision systems to provide additional information for neuropsychologists conducting BDT assessments. The framework incorporates HOI recognition with visual attention based on a multiscopic CPSS approach.

The study significantly contributes by applying vision systems from three perspectives to analyze hand behavior in three binding domains. (1) In the microscopic model, the feature extraction process utilizes hand-tracking vision to collect data on hand skeleton and finger joint angles, allowing for the estimation of physical hand postures. (2) In the mesoscopic model, symbolic interpretation is constructed using

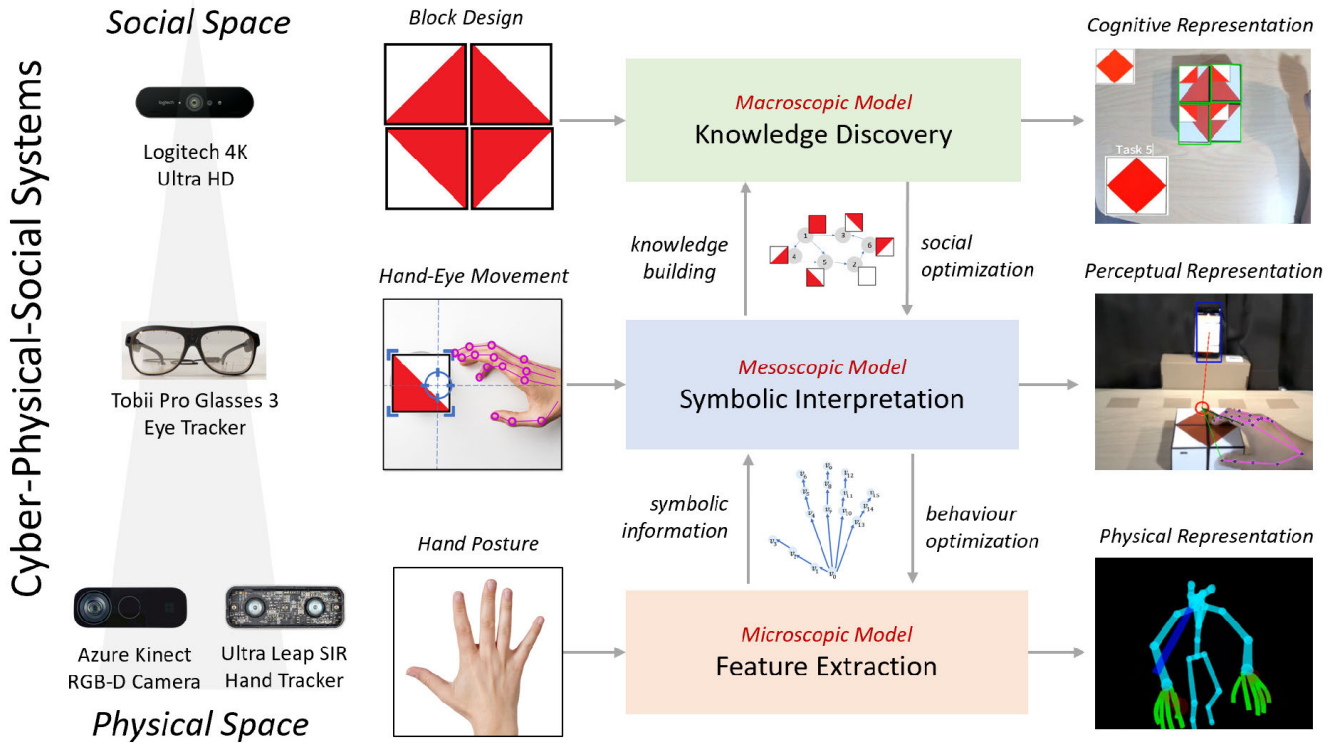

FIGURE 1. Multiscopic CPSS framework based on HOI recognition with visual attention for independent BDT.

egocentric vision with eye tracking. This model enables the analysis of hand-eye coordination by measuring the distance between the fingertips and the visual center of attention during block interaction. (3) The macroscopic model incorporates cognitive ability assessment through evaluation vision, which classifies color features in each block and determines the compatibility of the design with the given task. An eight-design scenario was conducted to validate the comprehensive approach, successfully measuring human behavior in reaching and grasping blocks from multiple perspectives. Figure 1 shows a multiscopic CPSS framework based on HOI recognition with visual attention for independent BDT.

The article is structured as follows: Section II discusses related works in monitoring the BDT for physical and cognitive assessments. Section III presents the proposed method for enhancing the current measurement approach through a multiscopic framework. Section IV discusses the findings obtained from the study and justifies the effectiveness of the proposed framework. Finally, Section V offers concluding remarks and outlines potential future directions for further research.

II. RELATED WORKS

Recent advancements in rehabilitation research have shown a growing interest in leveraging multiple vision systems to analyze hand movements. Researchers have explored various applications, including fingertip detection during therapy ball usage [16], monitoring hand movements in patients with spinal cord injuries [17], [18], and stroke patients [1], [19]. Computational challenges have also been addressed, such as the high cost of additional equipment and pixel-level observation [20] and issues like occlusion, inference, and contact [21]. Egocentric approaches using wearable cameras like GoPro and datasets such as Deeplab-VGG16, EgoHand, EPIC-ADL, and multi-datasets have been widely employed [22]. However, existing studies primarily focus on the physical evaluation of hand movements, with limited integration of cognitive abilities [23], [24]. On the other hand, hand-eye coordination research that predicts the next active object [25] rarely combines with physical assessments. To address this gap, we propose an innovative approach that combines the measurement of physical hand movements with the assessment of cognitive function using the BDT.

The BDT is a widely used neuropsychological assessment tool that evaluates visuospatial reasoning skills [26]. It is commonly included in standardized intelligence tests and is particularly valuable for describing cognition in individuals with neurological or developmental disorders. During the BDT, participants are tasked with recreating a pattern using red and white blocks. The accuracy of their reproduction and the time taken to complete the task is typically used to score their performance. However, additional measurement features derived from a person's BDT performance can provide valuable insights into their cognitive abilities [27]. Initially, the scoring system focused on tracking the number of block movements, but this approach proved challenging.

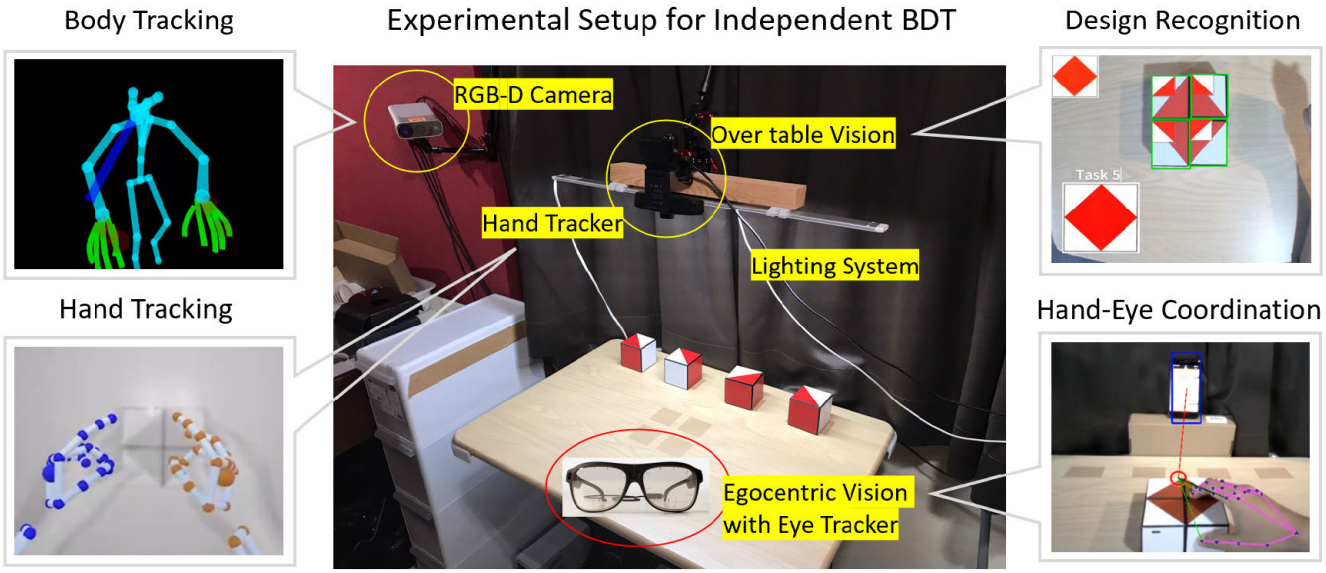

FIGURE 2. The experimental setup of multiscopic CPSS framework for independent BDT.

As a result, researchers began parameterizing BDT mistakes by analyzing the sequence of block movements, incorrect block placements, and the nature or severity of qualitative errors. Consequently, there is a need for the development of measurement technologies that can aid neuroscientists in analyzing BDT performance from various perspectives.

Cha et al. introduced a novel framework that combines vision systems and artificial intelligence (A.I.) approaches to enhance the knowledge obtained by neuropsychologists from the BDT and similar tabletop evaluations [15]. They demonstrated how machine learning techniques could automatically categorize and provide a detailed description of the person's activities and the status of the block task throughout each test. In 2020, the measurement method was further refined to capture more specific behavioral measurements [28]. Integrating scene and gaze cameras with supervised learning algorithms successfully quantified important behaviors during the block design exam, a widely used assessment of visuospatial cognitive abilities.

Averbukh et al. addressed the challenges of utilizing virtual reality in scientific visualization [29]. They experimented with evaluating how different phenomena presented in virtual reality environments affect the completion of intellectual tasks and explored user behavior in visualization systems. Rogers et al. validated these findings by developing Elements virtual rehabilitation, which incorporates goal-directed and exploratory upper-limb movement exercises to promote motor and cognitive recovery after stroke [30]. The study demonstrated significant training effects, the durability of improvements during follow-up, and the potential applicability to everyday activities, providing promising evidence for the effectiveness of virtual rehabilitation methods.

In a study by Wikstrom et al., a collaborative block design assignment was developed to assess and quantify performance in pairs [31]. The findings revealed that pair performance was normally distributed and strongly correlated with visuospatial abilities while not significantly associated with other participant-specific background variables. Dunn et al. provide a comprehensive overview of various independent and dependent variables explored in published BDT studies across multiple fields of cognitive science [27]. They also suggest areas of interest for future BDT research, especially with the availability of improved recording methods such as wearable eye trackers.

Recent research has raised concerns about using virtual reality (V.R.) in BDT. Shigenaga and Nagamune evaluated the effectiveness of BDT by recreating the test in a V.R. environment and analyzing participants' hand-eye movements during the test [32]. By administering the conventional BDT and V.R. to three healthy adult males, they concluded that the current V.R. implementation restricts gripping actions and suggests a more realistic system. However, physical simulation is considered one of the closest approaches to replicating natural human systems compared to V.R. applications. Promoting independent hand rehabilitation that encompasses both physical and cognitive aspects is crucial. This promotion is supported by studies demonstrating a relationship between the fine motor performance of the hand and cognitive abilities [23], [33].

Previous research conducted in our laboratory has focused on developing a musculoskeletal-based human physical simulation. We successfully implemented human skeleton tracking using multiple cameras to minimize occlusion [5]. The captured skeleton angles were then utilized to generate the musculoskeletal model, providing a comprehensive

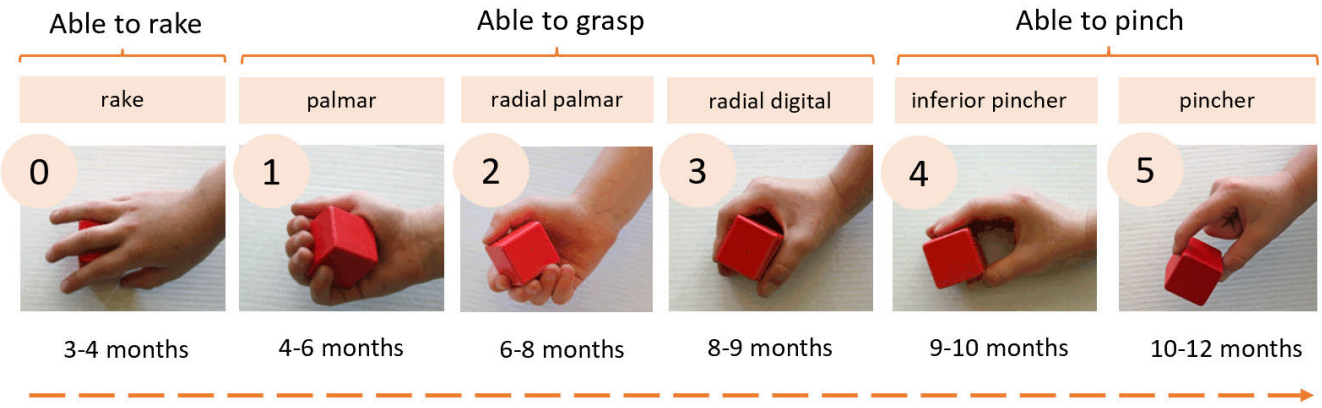

FIGURE 3. The development of human handling ability from the rake, grasp, and pinch.

understanding of human anatomy. Additionally, we conducted a study investigating a person's intentions and capabilities while reaching and grasping objects [7]. This research emphasized the significance of supporting existing systems from an egocentric perspective.

As in our previous work on the chopsticks manipulation test (CMT), we conducted a multiscopic approach to examine the importance of combining finger joint angle estimation and visual attention measurement in hand rehabilitation [34]. This study further supported the use of multiscopic methods to address dynamic locomotion in legged robots [35], cognitive memory systems for continuous gesture learning [36], and the application of CPSS for activity daily living (ADL) [37]. Building upon these insights, we propose a multiscopic approach to develop a CPSS for HOI recognition based on visual attention specifically for the BDT.

III. MULTISCOPIC CPSS

This study introduces a novel CPSS framework for advancing BDT research, utilizing a multiscopic approach. To address various technical challenges, we propose a three-level system that encompasses the following components:

1) In the microscopic model, we employ a hand-tracking optical sensor to collect hand skeleton data and finger joint angle features, enabling us to estimate the physical hand posture accurately
2) In the mesoscopic model, we leverage egocentric vision with an eye tracker to analyze hand-eye coordination. Specifically, we measure the distance between the fingertips and the visual center of attention during interactions with the block.
3) In the macroscopic model, we utilize over-table vision to evaluate cognitive ability. By classifying the color features in each block, we can determine whether the design aligns with the given task.

The experimental setup of the multiscopic CPSS framework for independent BDT is depicted in Figure 2, showcasing the integration of these three models to facilitate comprehensive assessment.

A. MICROSCOPIC MODEL: FEATURE EXTRACTION

The BDT relates to the International Classification of Functioning, Disability, and Health (ICF) in hand activities such as reach, grasp, handling, and manipulation. The BDT requires participants to rearrange blocks with specific color patterns related to ADLs, such as self-feeding [38]. Thus, the BDT results can affect drinking and eating activities. Our previous work has categorized basic motion primitives like reaching, grasping, and handling [7]. In this study, we plan to further categorize the motion primitives of object handling based on human development.

The development of a grasp is a significant milestone in infants' growth, typically occurring between six months and one year. Initially, babies use a raking approach that leads to palmar placement, allowing them to manipulate objects, explore them by bringing them to their mouths, and switch hands to grasp additional items. Around 7-8 months, the object is rotated radially, resulting in a radial palmar and radial digital grasp. The scissor grasp resembles the radial digital grasp. When the block can be grasped with the distal finger and thumb, the thumb presses the object into the side of the index finger, forming an inferior pincer grasp. Subsequently, around ten months later, the index finger and thumb joined to form a pincer grasp. Figure 3 illustrates the development of human handling abilities, including rake, grasp, and pinch.

We focus on feature extraction in the microscopic model for the experimental setup of grasp action. Feature extraction is crucial for data acquisition and attribute retrieval. We utilize the Ultraleap Stereo Infra-Red 170 (SIR170) [39], the next Leap Motion optical hand tracking sensor version, to capture hand position data in 3D coordinates. SIR170 employs cameras and infrared pattern projection to generate a three-dimensional image of the user's hand. It offers a larger field of view, extended tracking range, lower power

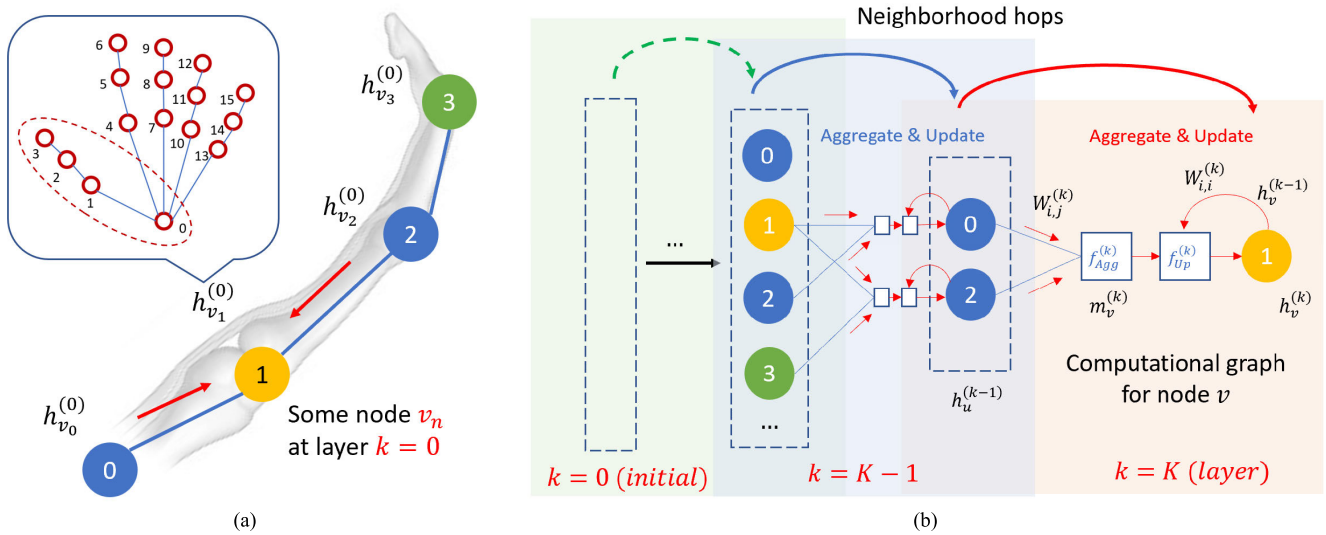

FIGURE 4. Physical finger representation: a) relationship between the joints using a directed graph; b) message-passing mechanism.

consumption, and a compact form factor. The sensor is installed on the table's top at a 15° angle facing downwards for optimal results to extract the features.

This study proposes 3D hand posture estimation using a kinematic finger model to represent the physical hand in BDT and obtain the physical embodiment [40]. A finger consists of three joints: A (metacarpal and proximal), B (proximal and intermediate phalanges), and C (distal phalanges and intermediate phalanges). As the thumb lacks metacarpals, we simplify the physical hand representation using kinematic finger models. The framework incorporates joint rotation configuration to validate finger motion specific to each individual accurately [11].

We have derived a formula to calculate the finger's joint angle based on the human hand's physical relationship [34]. To simplify the physical hand representation, we introduce a feature that replaces the joint's position with the angle formed between the joints of two adjacent bones. We standardize the position data of the joints to remove any outliers. Next, we employ a vector-to-joint-angle conversion technique to obtain the finger feature, performing this conversion for three positions. We comprehensively represent the physical hand by transforming the three-dimensional coordinate points into angles. The equation below illustrates how we calculate the angle between two vectors in three-dimensional coordinates.

$$\overrightarrow{\mathbf{AB}} = B - A \tag{1}$$

$$\overrightarrow{\mathbf{BC}} = C - B \tag{2}$$

$$\overrightarrow{\mathbf{AB}} \cdot \overrightarrow{\mathbf{BC}} = \left\|\overrightarrow{\mathbf{AB}}\right\| \left\|\overrightarrow{\mathbf{BC}}\right\| \cos\theta \tag{3}$$

$$\theta = \arccos\left(\frac{\overrightarrow{\mathbf{AB}} \cdot \overrightarrow{\mathbf{BC}}}{\left\|\overrightarrow{\mathbf{AB}}\right\| \left\|\overrightarrow{\mathbf{BC}}\right\|}\right) \tag{4}$$

To calculate the angle between vectors $\overrightarrow{\mathbf{AB}}$ and $\overrightarrow{\mathbf{BC}}$, we need the coordinates of three points, namely A, B, and C. By applying the right-hand rule and employing dot products, we can determine the angle formed by the sequence $A \rightarrow B \rightarrow C$. Additionally, the lengths of $\overrightarrow{\mathbf{AB}}$ and $\overrightarrow{\mathbf{BC}}$, denoted as $\left\|\overrightarrow{\mathbf{AB}}\right\|$ and $\left\|\overrightarrow{\mathbf{BC}}\right\|$ respectively, play a crucial role in this calculation. Using these values, we can calculate the dot product for $\overrightarrow{\mathbf{AB}} \cdot \overrightarrow{\mathbf{BC}}$. By rearranging the equation, we can solve for θ, representing the angle between the two vectors. These joint angle values serve as features in the data graph representation. We depict the physical finger representation, including the relationship between the joints, using a directed graph and a message-passing mechanism, as shown in Figure 4.

We then employ GNN to classify six different postures for cube handling. GNN is a neural network architecture widely used for learning representations of graph data and has gained popularity in prediction tasks involving nodes, graphs, and links. The underlying concept of GNN is to learn appropriate representations of graph data for neural networks. It is essential to delve into computer science's fundamental mathematical principles of graph-structured data [41]. In GNN, all the graph data is used as input, including node features and the connections stored in the adjacency matrix. A graph G can be a part of a set of attributed graphs G defined by the following equation:

$$G = (V, E, X), \quad G \in G. \tag{5}$$

Let $V = \{v_1, \ldots, v_n\}$ represent a set of nodes and $E = \{e_{a,b}, \ldots, e_{i,j}\}$ represent a set of ordered pairs that indicate connections between two nodes in V. Each node is associated with a set of node attributes, $X = \{x_v\}$, where $v \in V$. In the context of GNN, a new representation called embeddings is generated for each node, capturing the structural and

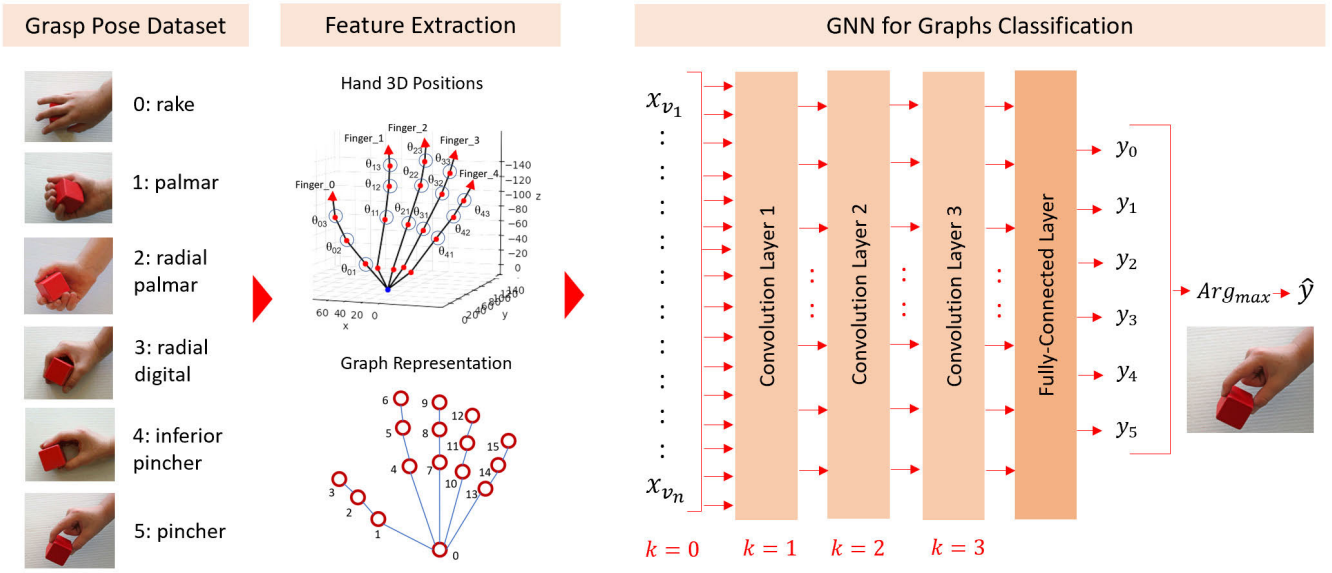

FIGURE 5. Microscopic model design using a three-layer GNN for hand posture classification.

feature information from other nodes in the graph. These node embeddings play a crucial role in making predictions. To obtain these embeddings, we perform multiple rounds of message passing. This process consists of three steps: initialization, aggregation, and update. During the initialization step, each node v at layer $k = 0$ undergoes the first round of message passing, which can be represented by the following equation:

$$h_{G,v}^{(0)} = x_v, \quad v \in V. \tag{6}$$

Let $h_{G,v}^{(k)}$ represent the node embeddings for a vertex v at the k-th layer, where the node features x_v come from all nodes $v \in V$ in graph G. In the next step, we apply the neural message passing scheme to perform aggregation on each node v using the following equations:

$$m_{G,v}^{(k)} = f_{\text{Agg}}^{(k)}\left(h_{G,u}^{(k-1)}\right), \quad 1 \le k \le K. \tag{7}$$

$$= \frac{1}{|N(v)|}\sum_{u \in N(v)} W_{i,j} h_{G,u}^{(k-1)}, \ i \ne j, \ 1 \le i,j \le |V|. \tag{8}$$

Using the neighborhood $N(v)$ of node v in graph G, we iteratively aggregate and store the node features $h_u^{(k-1)}$ of all nodes $u \in N(v)$ in $m_{G,v}^{(l)}$ using the aggregation function $f_{\text{Agg}}^{(k)}$. Here, $N(v) \subset V$ denotes the neighborhood of $v \in V$. The aggregation function is shared by all nodes within an iteration. Depending on the specific requirements, the sum operation can be replaced by average, degree-normalized sum, or coordinate-wise min or max.

In the final step of the neural message passing scheme, we update the node features $h_{G,v}^{(k-1)}$ of all nodes $v \in V$ in graph G iteratively using the aggregation results obtained from their neighbors $N(v)$. This update process is performed using the following equations:

$$h_{G,v}^{(k)} = f_{\text{Up}}^{(k)}\left(h_{G,v}^{(k-1)}, m_{G,v}^{(k)}\right) \tag{9}$$

$$= \sigma\left(W_{i,i} h_{G,v}^{(k-1)} + m_{G,v}^{(k)}\right), \quad 0 \le i \le |V|. \tag{10}$$

The update function $f_{\text{Up}}^{(k)}$ typically involves a weighted combination with learnable weight matrices. These update functions are implemented as fully connected layers, which consist of alternating linear transformations and coordinate-wise nonlinear activations σ, such as *ReLU*, *tanh,* or *sigmoid*. This allows for the incorporation of nonlinearity and the integration of information from neighboring nodes in the graph.

The final representation of a node, $h_{G,v}^{(K)}$, is obtained at the last layer, which can be concatenated with a linear classifier for specific tasks. In cases where a graph-level prediction is desired, all node embeddings can be aggregated into a single graph embedding, $H_G^{(K)}$, using the function f_{Read}. One commonly used method is to compute the average of the node embeddings. This method is achieved by summing up the node features of all nodes, $h_{G,v}^{(K)}$, in the K-th layer and dividing the sum by the total number of nodes, as shown in the following equation:

$$H_G = f_{\text{Read}}\left(h_{G,v}^{(K)}\right). \tag{11}$$

$$H_G = \frac{1}{|V|}\sum\nolimits_{v \in V} h_{G,v}^{(K)}. \tag{12}$$

Finally, to determine the class that corresponds to the input graph with the highest probability, we apply a linear transformation using a fully connected layer with W_{proj} as the weight projection. The *arg max* function is then used to select the class with the maximum probability, as shown in the

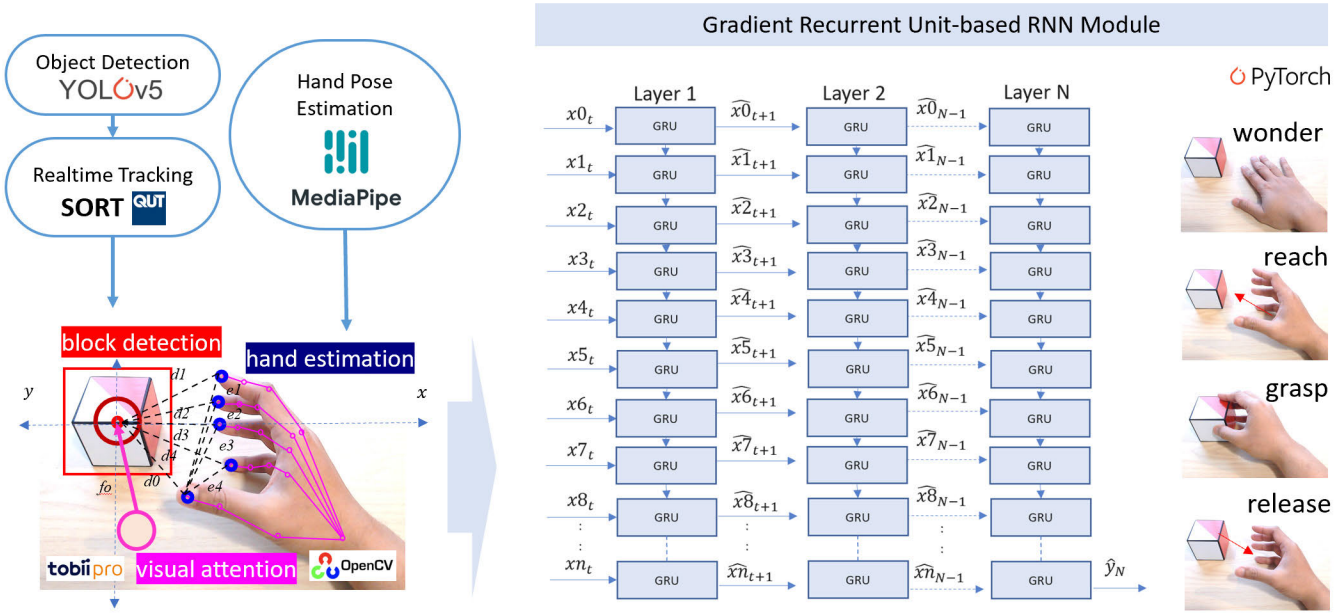

FIGURE 6. The mesoscopic model design for multivariate time-series classification.

following equation:

$$y_i = H_G W_{\text{proj}}. \tag{13}$$

$$\hat{y} = \arg\max_i y_i. \tag{14}$$

We utilize the Pytorch Geometric (PyG) [42] as the development framework for our GNN. We perform mini-batching on the small graph classification datasets to ensure efficient GPU utilization before inputting them into the GNN. PyG automatically handles the batching process by combining multiple graphs into one large graph. Figure 5 illustrates the design of our GNN with three layers for hand posture classification in the microscopic model.

We collected 1000 data samples for each hand posture in various orientations for our experiments, resulting in 6000 graphs. We divided these graphs into 4800 for training and 1200 for testing. Each data graph consists of 16 nodes connected by 15 edges. The input layer of the graph comprises 15 nodes representing joint angles and one node representing the wrist, which serves as the center of the finger connection. The output layer consists of 6 nodes representing different grasping postures. We train a final classifier on the graph embeddings obtained from the GNN.

Before applying the final classifier, we apply the Rectifier Linear Unit (ReLU) activation function to generate localized node embeddings. The GNN consists of three layers, and the training cycle involves constructing an optimizer, feeding the model inputs, computing the loss, and optimizing the model using autograd. We employ a linear transformation layer and the arg max function to classify the input and determine the grasping posture with the highest probability. The training and testing outcomes will be discussed in detail in this study's results and discussion section.

B. MESOSCOPIC MODEL: SYMBOLIC INTERPRETATION

Symbolic interpretation involves acquiring and analyzing information in the mesoscopic model, considering the interdependency of various tasks. In our previous research [43], we established the concept of task dependency, which refers to the relationship between a task and the sequence of rules that need to be executed. In the BDT activities context, this task dependency arises when hands and fingers interact with multiple blocks, picking up, rotating, and placing them back in their original positions. To assess this ability, we devised a set of task-dependent rules by combining hand features extracted from the microscopic model with visual attentional features.

To capture the visual information and get the perceptual representation, we employed egocentric vision using Tobii Pro Glasses 3 smart glasses [44]. The participant wearing these smart glasses faced the table and the blocks directly. These glasses have a high resolution of 1920 × 1080 pixels and a frame rate of 25 frames per second. With a wide field of view of 106° (95° horizontal and 63° vertical), the camera in the glasses effectively captures the hands and objects within its range. Accurate visual attention detection occurs when the hand and the block are within the camera's field of view.

We utilize the YOLOv5 model [45] to extract object locations from the image frame, which perception is represented by bounding boxes and labels. To enhance this detection process, we employ the Simple Online Real-time Tracking (SORT) technique [46]. This framework is highly effective in representation learning and has proven valuable for

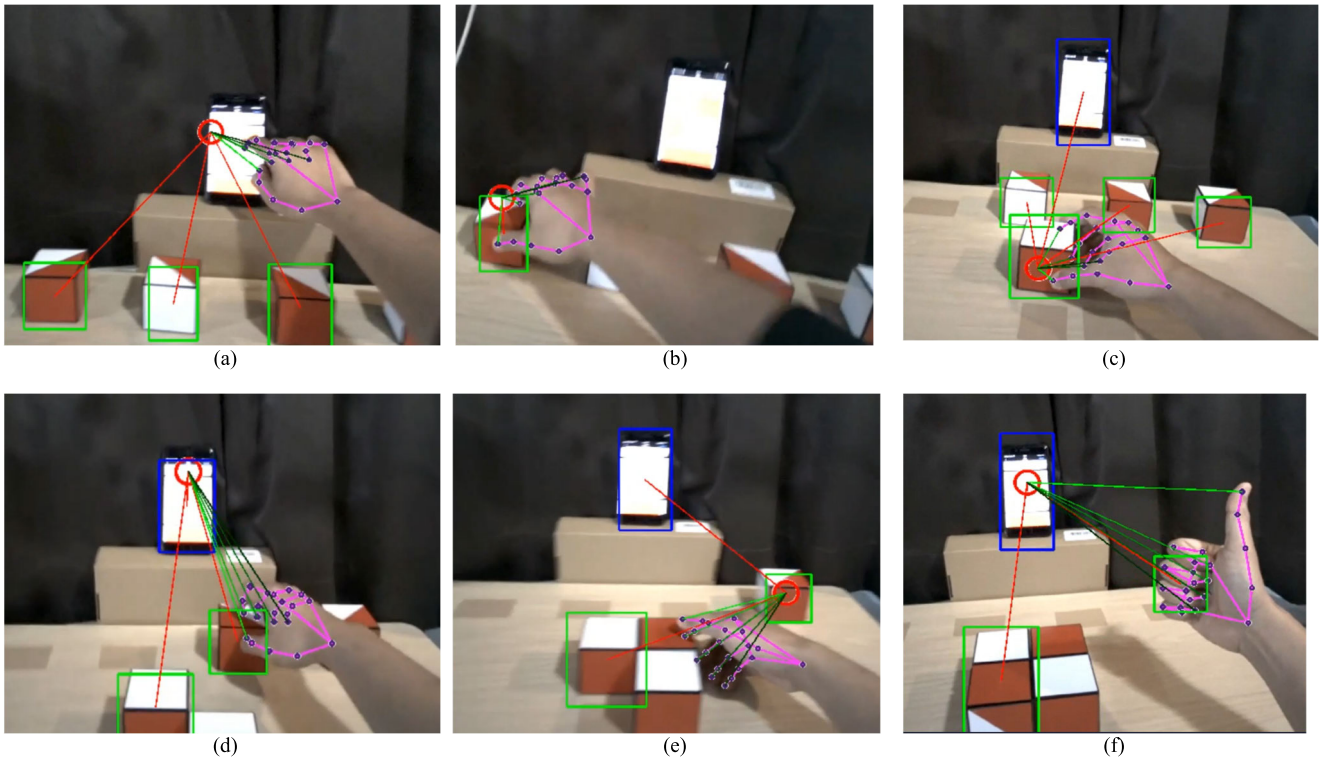

FIGURE 7. Visualization of BDT in the mesoscopic model in one task: (a) looking at the task; (b) reaching the block; (c) moving the block; (d) looking at the pattern; (e) looking at the block; (f) complete the task.

object recognition and tracking applications. Additionally, we employ MediaPipe hand tracking [47] to obtain approximate hand posture data. This method is designed for complex perceptual channels and enables fast real-time inference. By incorporating hand posture prediction as supporting data, we validate the recognition of HOI.

The combination of object detection and hand estimation provides two crucial pieces of information. Firstly, we can identify objects within a given image by searching for their presence. Secondly, we can accurately determine the hand's location and associated features in the two-dimensional image. The experimental design encompasses the system setup and implementation of an egocentric view for HOI recognition with visual attention. Figure 6 illustrates the mesoscopic model design for multivariate time-series classification, highlighting the approach of object-centered coordinates with visual attention in HOI recognition, which is compared with the traditional method.

We outline the critical steps in developing symbolic interpretation in the mesoscopic model, focusing on transforming object-centered coordinates and validating HOI recognition through the reach-to-grasp cycle specific to the task. We perform a coordinate transformation centered around the object to simplify the validation process and obtain a reduced dataset. This transformation aims to relocate the initial image coordinates (0,0) to the object's center.

For each new frame, we determine the new center and identify the position of the new coordinates. This approach applies to any pixel coordinate (x_n, y_n) in the image plane. By considering the joint finger position in the new coordinate plane and the object's length (a_0) and width (b_0), we can define the inner and outer borders of the object using the additional bounding box location information. With the help of the Pythagorean equation, we estimate the distance d_n between a point (a_n, b_n) and the center of the object coordinates (0,0). This distance d_n represents the distance between the fingertip or finger joint information, we utilize the properties a_n, b_n, and d_n. These properties enable us to confirm the interaction between the hand and the object and validate the recognition of HOI.

The validation of HOI recognition is specifically focused on the reference grasp defined in the approved hand usage section of the ICF for hand rehabilitation, particularly concerning the reach-to-grasp cycle [48]. This validation procedure consists of four distinct tasks, each defined separately. The first is the wonder task, where the user moves their hand freely without reaching for the object. This initial state serves as a baseline. The second task is the reaching task, where the subject extends their arm and opens their hand towards the object, simulating the reaching motion. Following that is the grasping task, where the participant grasps the object in any position, representing the perception when gripping. The task then transitions into the transport state when the user begins to

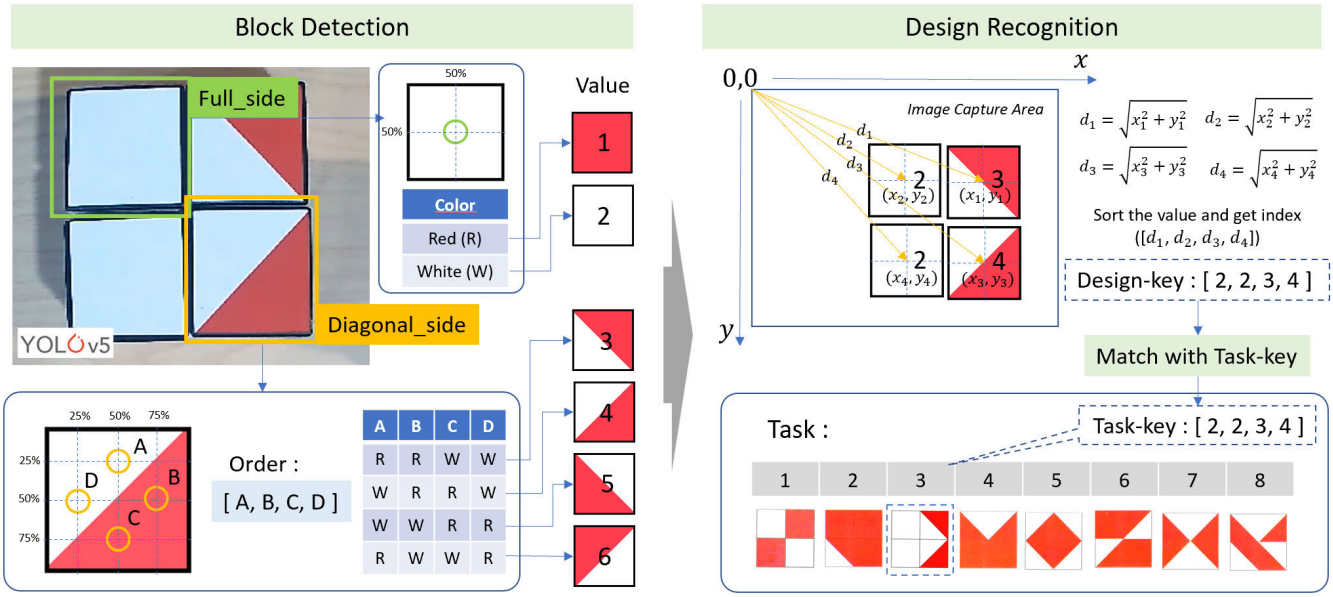

FIGURE 8. The knowledge evaluation in macroscopic model for block detection and design recogntion.

move the object they are holding. Finally, the fourth task is the release task, completed when the user withdraws their open palm away from the object, simulating the release action.

In the initial stage of our study, we focus on a single block as a reference object, which exhibits various hand postures. To capture the relevant features, we measure five elements: the distance of each fingertip to the center of the object $(d_o, d_1, d_2, d_3, d_4)$, four elements representing the distance of each fingertip to the thumb fingertip (e_1, e_2, e_3, e_4), and one visual attention feature (f_0). We collect ten frames per sample using our computer specifications, which serve as our benchmark for estimating the length of a data stream. We analyze ten data points in real-time within each picture capture series. This data is utilized as input for our neural networks in the learning system.

We employ a recurrent neural network (RNN) architecture, specifically a multi-layer gated recurrent unit (MGRU), to classify multivariate time series [49]. This design allows for flexibility in adjusting variables such as the number of layers, input size, hidden size, and recurrent layers. The MGRU-based RNN computes each element in every layer using the following functions:

$$r_t = \sigma(W_{i,r}x_t + b_{i,r} + W_{h,r}h_{(t-1)} + b_{h,r}), \quad (15)$$

$$z_t = \sigma(W_{i,z}x_t + b_{i,z} + W_{h,z}h_{(t-1)} + b_{h,z}), \quad (16)$$

$$n_t = \sigma_h(W_{i,n}x_t + b_{i,n} + r_t * (W_{h,n}h_{t-1} + b_{h,n})), \quad (17)$$

$$h_t = (1 - z_t) * n_t + z_t * h_{t-1}. \quad (18)$$

In the above equations, t represents time. The hidden states are denoted by h_t, the inputs by x_t, and the hidden states of the previous layers at time $t-1$ by h_{t-1} or the initial hidden states at time $t = 0$. Additionally, r_t, z_t, and n_t represent the reset, update, and new gates, respectively. The sigmoid function is denoted by σ, and the $*$ symbol represents element-wise multiplication. In the MGRU, the input $x_t^{(l)}$ of the l -th layer ($l >= 2$) is obtained by multiplying the hidden states $h_t^{(l-1)}$ of the previous layers by the dropout, $\delta_t^{(l-1)}$, where each $\delta_t^{(l-1)}$ is a Bernoulli random variable with a dropout probability. Figure 7 provides a visualization of the mesoscopic model of BDT in one task.

Once we have constructed the MGRU-based RNN architecture, the next step is to create a dataset for evaluating the recognition of HOI for each action. We recorded video sequences of 1-2 seconds in duration, capturing at least 50 frames per sample. Our dataset consists of 100 videos depicting hands interacting with various objects. The videos were divided as follows: 25 for the wonder task, 25 for the reaching task, 25 for the task involving grasp and transport, and 25 for the release task. We randomly split the data into training and validation sets using an 80:20 ratio to ensure a reliable evaluation. We made this division considering the subjective nature of the obtained data. In this experiment, we involved a responder who performed the actions. The training set comprised 80 videos, while the remaining 20 videos were used for testing. The results and discussion section presents the outcomes of the training and testing process.

C. MACROSCOPIC MODEL: KNOWLEDGE DISCOVERY

Knowledge discovery is the process of finding functional patterns in large datasets using techniques like data mining. It combines statistics and computer science to extract knowledge. This section focuses on the macroscopic model and evaluates cognitive skills as knowledge using the BDT. The BDT assesses an individual's ability to perform tasks

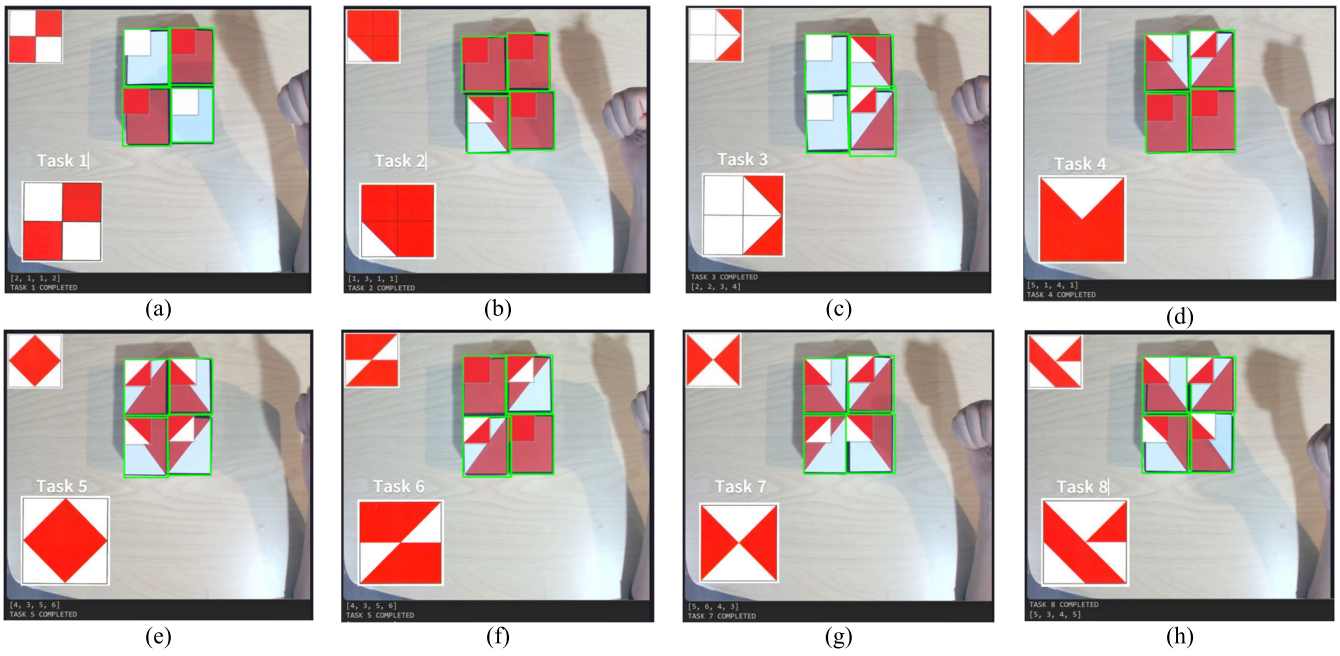

FIGURE 9. Visualization of BDT in the macroscopic model for eight tasks from easy to complex.

based on given designs. Specifically, we utilized a camera positioned on the table to capture a top-down view of the block. We then executed several steps, including detecting the top surface of the block, recognizing the design, matching the design key with the task key, and calculating the time needed to complete each task. Figure 8 illustrates the knowledge discovery in the macroscopic model, which involves block detection and design recognition.

To identify the top surface of the blocks, we utilize the YOLOv5 model, which extracts locations through bounding boxes and labels from the image frame. We classify the color features of each block into two classes: full-side and diagonal-side. However, simultaneously distinguishing all six features can be challenging since object detection may struggle with similar features and color variations caused by varying lighting intensities. To address this issue, we simplify block detection into two classes and train the model using multiple datasets. For the full-side class, we consider two block colors: red (1) and white (2). As for the diagonal-side class, we consider four combinations of red and white colors (3, 4, 5, and 6). This approach allows us to simultaneously overcome the difficulties of distinguishing all six features.

After the system successfully detects a block in the full-side category, we calculate its middle part's average color composition at coordinates (50%, 50%). If the color is closer to red (R), we assign it a value 1. Conversely, if it is closer to white (W), we assign a value 2. The approach differs when the system detects a block in the diagonal-side category. For diagonal-side blocks, we divide the block into four zones: zone A, located at the top coordinate (25%, 50%); zone B, located at the right coordinate (50%, 75%); zone C, located at the bottom coordinate (50%, 75%); and zone D, located at the left coordinate (25%, 50%).

We calculate the average color composition for each zone and store it in an ordered list based on its proximity to red or white. The order list determines the assigned value. For example, if the order list contains (R, R, W, W), the block is given a value of 3. Similarly, if the order list is (W, R, R, W), the block is assigned a value 4. Other combinations include (W, W, R, R) for a value of 5 and (R, W, W, R) for a value of 6. This process is applied to all block detection systems. By implementing this methodology, we can accurately determine the color composition and categorize each block based on its color characteristics.

Once four blocks are successfully detected, the next step is design recognition, where we aim to identify the design formed by these blocks and match it with the given task design. At this stage, we already have four values that represent the arrangement of the blocks. However, before we proceed with the matching process, we need to validate the position of each block to one another. To accomplish this, we calculate the distance between the centers of the four blocks, denoted by coordinates (x_n, y_n), and the coordinate $(0, 0)$. This distance is represented as d_n, where $n = 1, 2, 3, 4$ corresponds to the randomly assigned block detection numbers. The distance d_n can be computed using the Pythagorean theorem, which involves taking the square root of the sum of the squares of the x_n and y_n coordinates. After obtaining the four distances, we store them in a list $[d_1, d_2, d_3, d_4]$ and sort the values in descending order. The resulting order, represented by the list index, determines the arrangement of the four block

values in the essential design list. By organizing the block values in this manner, we establish a consistent cognitive representation that allows us to compare the formed design with the task design effectively.

In the final stage of the process, we compare the design key obtained through design recognition with the task key associated with each specific task. If the design key matches the task key, it indicates that the participant has completed the design for that particular task. Upon successful completion, the participant can proceed to the next task in the sequence. Additionally, we measure the time required to complete each task. The timing begins when the task starts and continues until the subsequent task is completed. It is important to note that time is not counted when the task is already detected or has been detected multiple times. Instead, the timing starts when the task begins and is undetected. This task ensures that the time measurement accurately reflects the duration of the participant's active engagement in completing the task. Figure 9 provides a visual representation of the macroscopic model of the BDT, illustrating the eight tasks arranged in order of increasing complexity, ranging from easy to challenging.

IV. RESULT AND DISCUSSION

The experiment conducted in this study involved participants engaging in tabletop activities using the Wechsler Adult Intelligence Scale-IV (WAIS-IV) BDT [26]. Our framework was designed to assess both physical hand function and cognitive function simultaneously. Before testing the system for rehabilitation, we obtained ethical approval from the Ethical, Legal, and Social Issues (ELSI) committee at Tokyo Metropolitan University (TMU), Japan. The initial testing was conducted on healthy student participants in selected laboratories. Eight healthy students participated in the study, aged 20 to 40 years. Before their involvement, all participants provided informed consent for the study. The following outlines the standard operating procedure that was followed during the BDT experiments:

1) Experimental Ethics: The researcher informs the participants about the data collection process, including the recording and processing egocentric vision, upper table video, and eye-tracking data. It is emphasized that this will be done under the data transparency policy [50].
2) BDT Trial: Once participants have consented, they are given one minute to familiarize themselves with the WAIS-IV BDT No.1 guidelines with eight designs commonly used to assess cognitive function in elderly individuals.
3) Preparation and Setup: The preparation and setup process involves several steps: (i) positioning the blocks in the designated ''all-white'' construction area, (ii) having participants wear the eye trackers, with the option of requesting additional lenses if needed, and (iii) wirelessly connecting the computer system to the eye tracker and calibrating it using a calibration card.
4) Experiment: With participants ready for the investigation, they are tasked with completing two scenarios: one using their dominant hand and the other using their non-dominant hand. In each scenario, participants arrange four blocks according to a given pattern. This results in a total of eight patterns/designs created. The system records a video of the participants as they complete each design.
5) Closing: After the session, participants remove the eye tracker and return it to its original location while the researchers reset the blocks to their initial ''all-white'' design, preparing the equipment for the next participant.

It is important to note that throughout the experiment, the privacy and confidentiality of the participant's personal information and data were strictly maintained under ethical guidelines and regulations. The results obtained at the microscopic, mesoscopic, and macroscopic models are then described, followed by a discussion of critical points that must be emphasized in developing the multiscopic system for future improvements.

A. MICROSCOPIC MODEL EVALUATION

Initially, we focused on developing the microscopic model of our system, which involved designing a feature extraction method using hand-tracker vision to analyze hand and finger postures. To estimate these features, we utilized the kinematics of the finger [34] and hand model [51]. However, not all feature extractions yielded accurate results due to the varying angles at which fingers are positioned relative to the camera. To address this challenge, we devised a solution combining the hand estimation results from hand-tracker vision and a dual Kinect camera [5]. We achieved more reliable outcomes by incorporating the data obtained from the skeleton Kinect. This approach created a cohesive hand skeleton model as a directed graph structure. The graph data encompassed the connections between the joint angles of the hand and fingers, providing a comprehensive physical representation of the hand.

The GNN learning process focused on classifying six hand postures within a directed graph structure. Initially, the acquisition system gathered input data from the hand-tracker vision and stored it in the database. This dataset was then utilized to train the GNN model using 1000 epochs. The training phase demonstrated that the GNN model, when applied to supervised classification, yielded promising results worthy of discussion. The testing phase was conducted to validate the performance of the classification system. The testing dataset was integrated into the GNN model, and the accuracy of the model's predictions was compared to that of traditional models, specifically the multi-layer perceptron (MLP). The GNN model achieved an impressive testing accuracy of 97.5% for the six hand postures. In contrast, the MLP network achieved a testing accuracy of 82.0%. This comparison highlights the

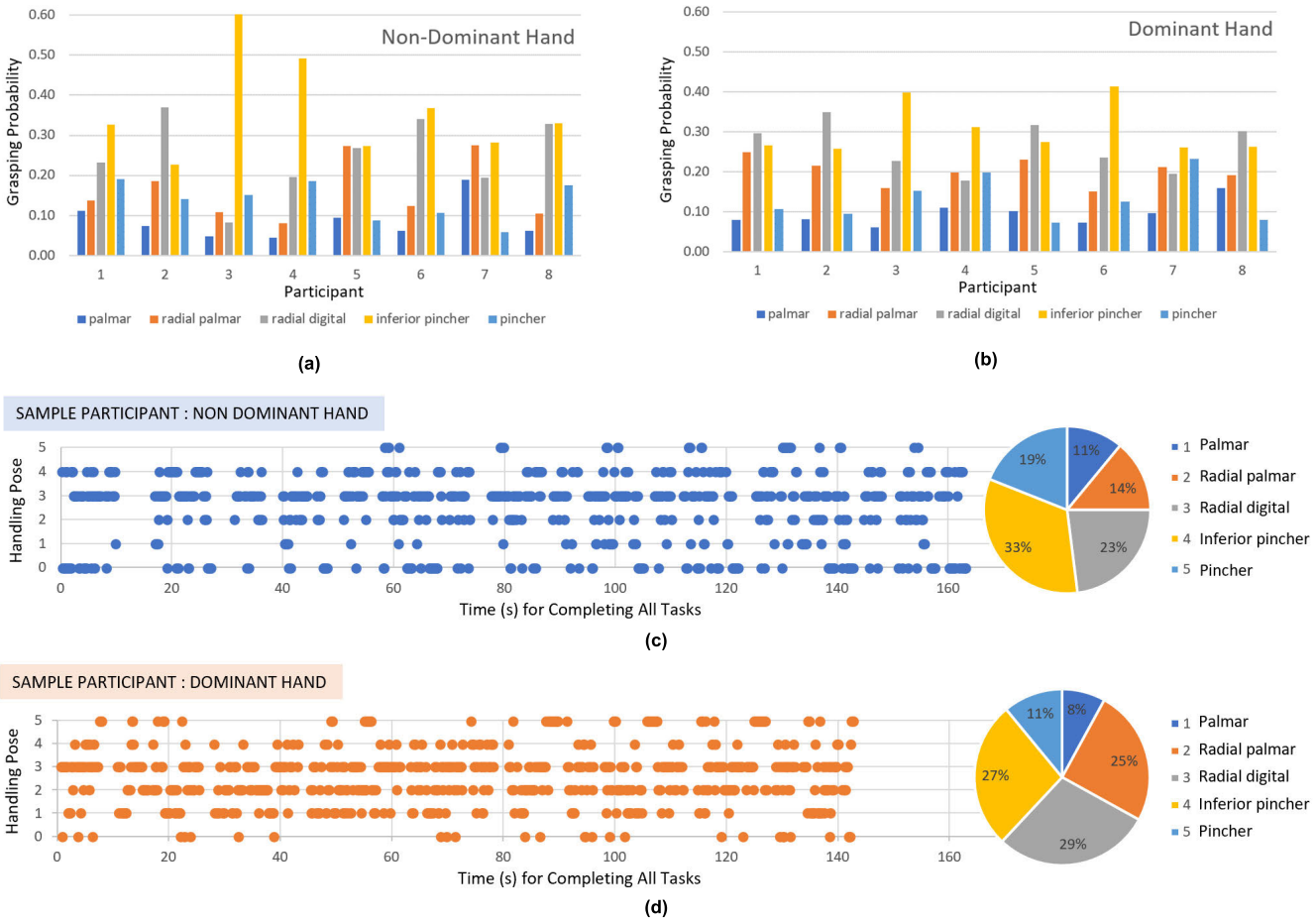

FIGURE 10. Result of eight participants during BDT in the mesoscopic model for completing all tasks: a) Non-dominant hand, b) Dominant hand, c) Sample data of one participant with non-dominant hand in time series, d) Sample data of one participant with dominant hand in time series.

superior performance of the GNN model in accurately recognizing hand postures.

Furthermore, the results indicate that the dataset of hand postures used in the experiment contains crucial characteristics for adequate classification. Specifically, the MLP network struggled to distinguish between similar hand postures, such as radial digital grasp and inferior pincher. However, adding a three-layer GNN to the MLP significantly improved the accuracy, as demonstrated by the higher testing accuracy achieved by the GNN model.

After establishing the accuracy of the feature extraction system, we proceeded to classify the hand postures in the BDT activity. This classification was tested on eight healthy participants. Figure 10 displays the results of the eight participants in the microscopic model as they completed the eight tasks in the BDT. The figure illustrates the variations in hand postures between different individuals. In the dominant-hand scenario, hand postures 2 (palmar radial), 3 (digital radial), and 4 (inferior pincher) were observed to be more frequently used compared to postures 1 (palmar) and 5 (pincher). However, in the non-dominant hand scenario, hand postures varied across the eight healthy participants. The comparison of hand postures in the dominant hand showed a more diverse distribution compared to the non-dominant hand scenario.

In our statistical analysis, we introduce two main parameters to assess the non-normality of the hand posture distribution: skewness of hand posture (SHP) and kurtosis of hand posture (KHP). [52]. SHP quantifies the asymmetry of the distribution of the five hand postures, including grasping and pinching. An SHP value of zero indicates a symmetric distribution around the mean, characteristic of a normal distribution. Positive SHP indicates a longer right tail in the distribution, while negative SHP indicates a longer left tail.

On the other hand, KHP measures the peakedness of the hand posture distribution. A KHP value of 3 corresponds to a distribution that is neither flat nor highly peaked, indicative of a normal distribution. Higher KHP values indicate a more peaked distribution, while lower KHP values indicate a flatter distribution. Further analysis, such as regression and correlation, will be conducted in the statistical analysis sub-chapter to delve into these parameters and explore their relationships in more detail.

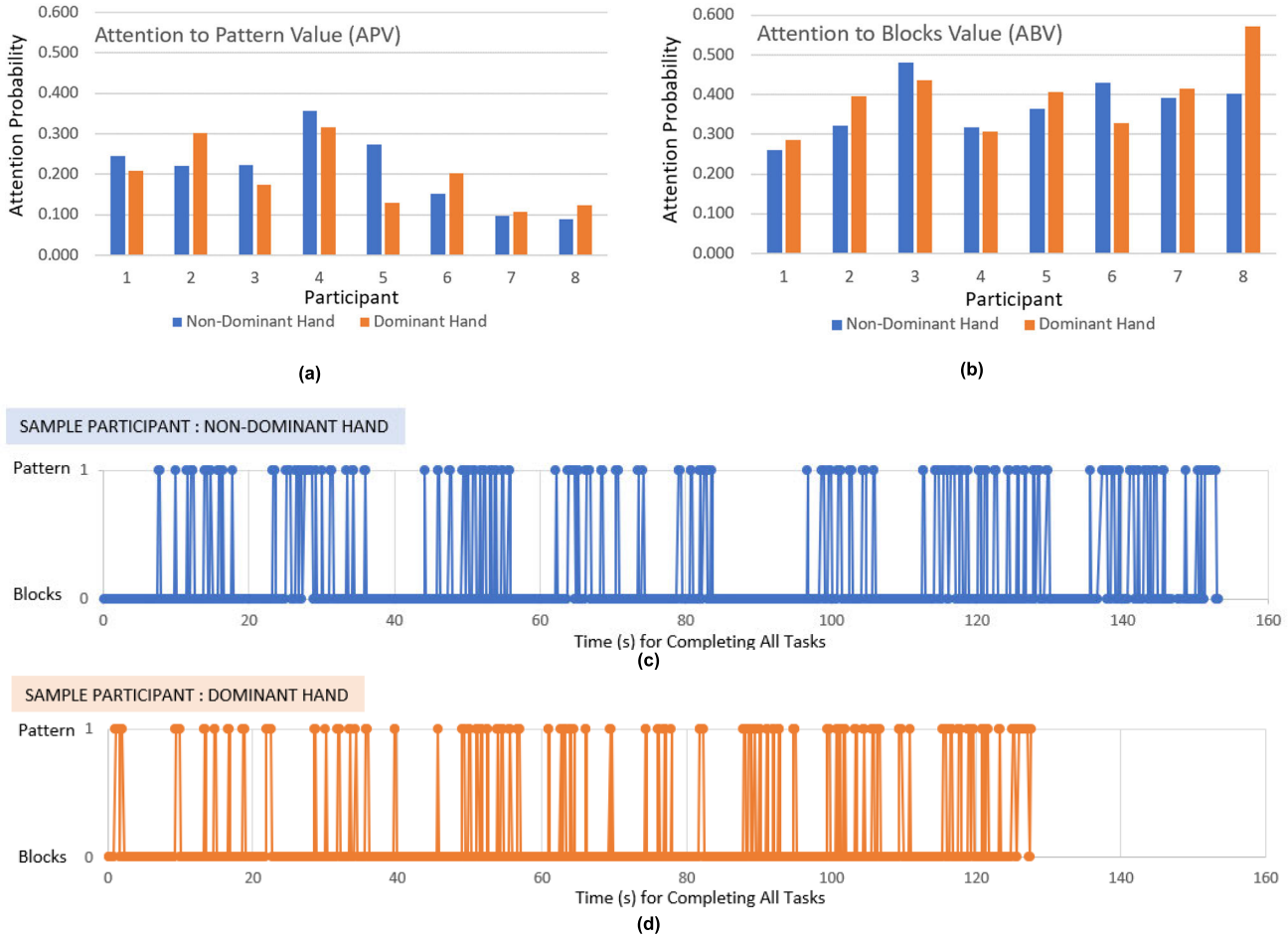

FIGURE 11. Result of eight participants during BDT in the mesoscopic model for completing all tasks: a) Attention to Pattern Value (APV), b) Attention to Block Value (APV) c) Sample data of one participant with non-dominant hand in time series, d) Sample data of one participant with dominant hand in time series.

B. MESOSCOPIC MODEL EVALUATION

We have also focused on the mesoscopic model, which involves investigating the symbolic interpretation using egocentric vision. We aim to understand the progress in behavioral performance and the visual strategies employed in acquiring grasp skills [34], [53]. To achieve this, we conducted a task-specific reach-to-grasp cycle experiment and built upon the egocentric vision capabilities discovered in our previous research [51]. However, during the system's deployment, we encountered significant technical challenges with the feature extraction process. Firstly, we observed that MediaPipe's hand position evaluation and YOLOv5 object identification showed lower accuracy when dealing with specific hand postures and occlusions. To address this issue, we developed a multi-vision system [5] that combines different visual techniques to enhance accuracy.

The second issue was the lack of depth information in the collected data, represented in 2D pixel units. Since the egocentric approach requires three-dimensional understanding, we believe using an RGB-D camera can provide more consistent and detailed depth perception results [54]. Further research on incorporating sensory depth data could improve the effectiveness and applicability of this low-cost application. By addressing these technical challenges and exploring additional depth sensory research, we aim to enhance the capabilities and performance of the egocentric vision system in the mesoscopic model.

We evaluated the testing results of the symbolic interpretation in the task-specific reach-to-grasp action. For this purpose, we trained two recurrent neural networks (RNN) models: MGRU and long short-term memory (LSTM) [7]. The goal was to assess the system's accuracy in recognizing the actions performed. The evaluation revealed that the MGRU model achieved the best recognition results, with an average accuracy of 97.0%, while the LSTM model achieved an accuracy of 94.0%. Additionally, the MGRU model exhibited a shorter training time than LSTM, with MGRU taking approximately 13.03 seconds and LSTM taking 20.48 seconds.

To further explore the symbolic interpretation, we utilized an MGRU-based RNN to address a multivariate time-series classification problem. The MGRU model outperformed the

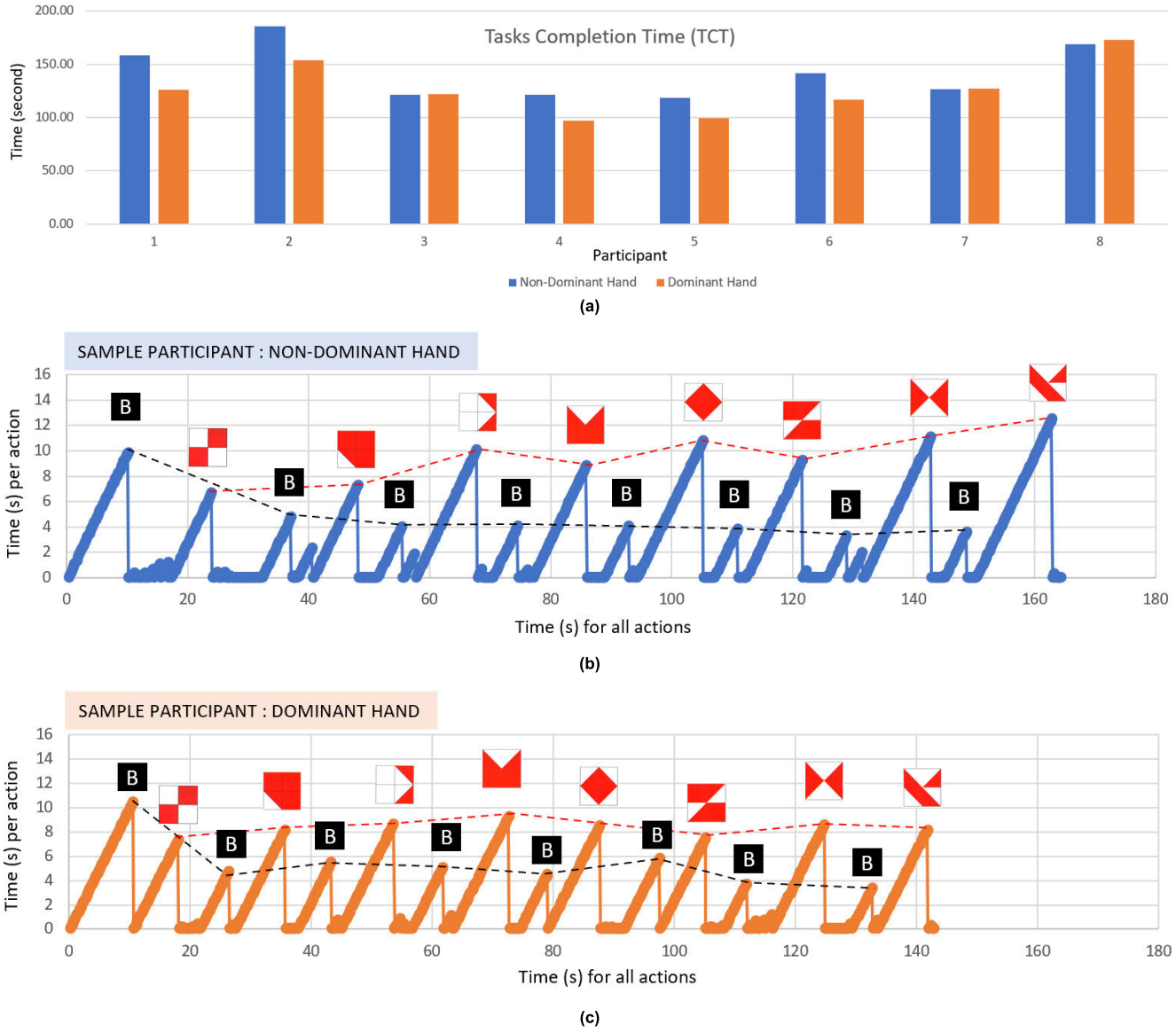

FIGURE 12. Result of eight participants during BDT in the macroscopic model for completing all tasks: a) Task Completion Time (TCT) for Non-dominant hand and Dominant hand, b) Sample data of one participant with non-dominant hand in time series, c) Sample data of one participant with dominant hand in time series.

vanilla-RNN and LSTM models in terms of accuracy. Previous studies have demonstrated that the MGRU model can integrate information rapidly and improve time-series identification performance compared to the basic model [55]. Despite using a limited number of features for training, we achieved satisfactory accuracy in our experiments. These results highlight the effectiveness of the MGRU-based RNN model in capturing the dynamics of the task-specific reach-to-grasp action and its potential for advancing symbolic interpretation capabilities.

After confirming the reasonable accuracy of the symbolic interpretation, our focus shifted to classifying the reach-to-grasp behavior and eye movements in the BDT activity. We collected data from eight healthy participants, explicitly capturing the timing of reaching, grasping, moving, manipulating, and releasing the blocks. This data provided insights into the participants' gaze patterns and block interaction during the design process. Figure 11 illustrates the results of the eight participants during the BDT activity in the mesoscopic model while completing all the tasks. The figure highlights the variations in symbolic interpretation among individuals. Each participant exhibited different durations of focus on the patterns and blocks during block manipulation.

To quantify human attention during the BDT activity, we propose two main parameters for statistical analysis: attention to pattern value (APV) and attention to block value (ABV), building upon previous research [28], [56]. APV represents the frequency of a person's gaze on the pattern during

the test. It can be calculated by dividing the accumulated frames where the total number of frames detected the pattern. Similarly, ABV indicates the gaze frequency on the blocks during the test. ABV is obtained by dividing the frames where the total number of frames detected blocks. These two values, APV and ABV, will be further analyzed for regression and correlation relationships in the subsequent statistical analysis subsection.

C. MACROSCOPIC MODEL EVALUATION

In addition to the microscopic and mesoscopic models, we developed a macroscopic model to evaluate cognitive ability using upper table vision. Our objective was to assess cognitive skills in the BDT activity, where participants perform tasks based on the eight patterns provided. To enable comprehensive knowledge discovery, we extended the capabilities of YOLOv5. We incorporated functionalities such as full-side and diagonal-side classifications, detecting red or white colors, and block orientations on the diagonal side. This enhancement allowed us to simultaneously recognize and analyze six color features across multiple blocks.

However, we conducted the recognition process in stages due to limitations in the current object detection system. This process was necessary because the system faced challenges in accurately determining the similarity of features and color differences, primarily influenced by variations in lighting intensity [57]. By addressing these technical issues, we aimed to provide a more robust cognitive assessment of the macroscopic model of our framework.

After conducting several real-time experiments, we implemented the BDT upper-side classification method. Once we confirmed our evaluation system's accuracy, we measured the participants' activity during the BDT activity. We captured the duration for each pattern to be completed by the eight healthy participants. Additionally, we recorded the information regarding the return of all the blocks to the block bank (B), which can be further analyzed using the Block and Box Test [58]. Figure 12 illustrates the results obtained from the eight participants during the macroscopic model of the BDT activity, encompassing the completion of all tasks.

The result highlights the variations in cognitive skills among the participants. Notably, all participants could complete the eight BDT patterns within a relatively short time frame. Participants who predominantly used their dominant hand exhibited slightly better performance than those using their non-dominant hand, with a noticeable difference in completion time [59]. Conversely, participants who excelled with their non-dominant hand exhibited slightly lower completion times than those using their dominant hand. These findings shed light on the interplay between cognitive skills and hand dominance during the BDT activity.

The macroscopic results obtained from this study offer valuable preliminary insights for cognitive therapists, enabling them to conduct further analysis. Mainly, these results serve as a foundation for understanding hand postures in individuals undergoing rehabilitation. Gathering personal datasets, especially from rehabilitation patients, is crucial to recognize and assessing various hand postures accurately [60]. By employing the proposed technology, therapists can gain valuable insights into a person's behavior during the grasping process, which can inform the design and implementation of tailored rehabilitation systems.

Considering individual constraints and specific issues when developing rehabilitation systems is essential. Each person may have unique needs and limitations, and the rehabilitation system should be customized to address these factors effectively. Incorporating personalization and individualized approaches can optimize the rehabilitation process to promote better patient outcomes.

D. STATISTICAL ANALYSIS

We have successfully evaluated the multiscopic CPSS [37] as a novel framework for BDT applications, demonstrating high accuracy across the microscopic, mesoscopic, and macroscopic models. However, it is essential to note that the system's accuracy may vary when applied to different datasets or under varying conditions. This variability could be attributed to factors such as variations in hand size, non-standardized handling techniques, or environmental changes.

We recruited eight healthy participants to conduct a comprehensive system analysis and collected data from them. The participants were trained to grasp the block using various hand postures and correctly position their fingers using dominant and non-dominant hands. These trial sessions were designed to facilitate error-free learning for participants unfamiliar with the BDT. By ensuring participants underwent practice sessions, we aimed to establish a linear relationship with time-related variables. This assumption was made to ensure that all participants completed their initial tasks, as any failures would render their data invalid. Table 2 compares parameters for eight healthy participants in two scenarios.

Next, we conducted a correlation analysis to assess the strength of the relationships between each evaluation index. The Pearson correlation coefficient (R) was employed to analyze the relationships. Figure 13 displays the R-values for the four relationships, depicting the regression and correlation analysis results between BDT tasks performed using the non-dominant hand and those performed using the dominant hand. The findings align with our prediction, as the Task Completion Time (TCT) values for the dominant hand are predominantly more minor than those for the non-dominant hand. Moreover, both scenarios exhibit similar correlations and display a linear relationship. Table 3 shows the squared Pearson correlation coefficient (R^2) for eight healthy participants in two scenarios.

Figure 13(a) indicates no correlation between SHP and the increase in TCT. Similarly, Figure 13(b) demonstrates that KHP does not correlate significantly with the increase in TCT. These initial results suggest no substantial differences in hand posture among healthy participants that would impact

TABLE 2. The comparison of parameters for eight healthy participants in two scenarios.

Scenario	Subject	Parameters				
		Skewness of Hand Posture (SHP)	Kurtosis of Hand Posture (KHP)	Attention to Pattern Value (APV)	Attention to Blocks Value (ABV)	Tasks Completion Time (TCT) in second
A: Non-Dominant Hand	P1	-0.442	-0.769	0.246	0.261	158.44
	P2	-0.070	-0.680	0.221	0.323	185.75
	P3	-0.755	-0.367	0.223	0.481	121.25
	P4	-0.691	-0.068	0.357	0.319	121.51
	P5	-0.001	-0.866	0.274	0.364	118.38
	P6	-0.408	-0.198	0.153	0.429	141.70
	P7	0.063	-1.166	0.098	0.393	126.44
	P8	-0.452	-0.224	0.088	0.403	168.82
	Average	**-0.345**	**-0.542**	**0.207**	**0.372**	**142.79**
B: Dominant Hand	P1	-0.031	-0.784	0.209	0.286	125.89
	P2	-0.070	-0.650	0.302	0.396	153.86
	P3	-0.383	-0.649	0.175	0.436	122.32
	P4	-0.304	-1.058	0.317	0.308	97.18
	P5	-0.121	-0.699	0.130	0.406	99.22
	P6	-0.401	-0.594	0.203	0.328	116.97
	P7	-0.237	-1.112	0.107	0.416	127.17
	P8	-0.117	-0.870	0.124	0.572	172.90
	Average	**-0.208**	**-0.802**	**0.196**	**0.393**	**126.94**
Difference of scenario A and B.	P1	-0.411	+0.015	+0.037	-0.024	+32.55
	P2	-0.001	-0.030	-0.081	-0.073	+31.89
	P3	-0.372	+0.282	+0.047	+0.045	-1.07
	P4	-0.387	+0.990	+0.040	+0.011	+24.33
	P5	0.119	-0.167	+0.145	-0.042	+19.16
	P6	-0.007	+0.396	-0.050	+0.101	+24.73
	P7	0.300	-0.054	-0.009	-0.023	-0.73
	P8	-0.335	+0.647	-0.035	-0.169	-4.08
	Average	**-0.137**	**+0.260**	**+0.012**	**-0.022**	**15.84**

TABLE 3. The squared pearson correlation coefficient (R^2) for eight healthy participants in two scenarios.

Scenario	Pearson Correlation Coefficient (R^2)			
	R^2 (SHP, ATP)	R^2 (KHP, TCT)	R^2 (APV, TCT)	R^2 (ABV, TCT)
A: Non-Dominant Hand	0.0248	0.0015	0.1568	0.1042
B: Dominant Hand	0.1782	0.0066	0.4554	0.0248

their scores on the BDT. Therefore, gathering data from individuals with hand function impairments is necessary to investigate this aspect further. However, we observed that the distribution of SHP and KHP on the non-dominant hand exhibits more variability than the dominant hand, as indicated by the dashed red box in Figures 13(a) and 13(b). Assuming that the non-dominant hand represents a dominant hand with reduced function, we can conclude that SHP and KHP values may be an initial step in distinguishing a normally functioning dominant hand from one with decreased functionality.

Figure 13(c) illustrates a negative correlation between APV and the increase in TCT. This finding suggests that individuals focusing more on patterns tend to complete the design in less time, resulting in higher BDT scores. On the other hand, Figure 13(d) demonstrates a positive correlation between ABV and the increase in TCT. This correlation indicates that individuals focusing more on blocks take longer to complete the design, resulting in lower BDT scores. These results suggest that the improvement in BDT performance is influenced by visual attention, particularly the emphasis on patterns rather than blocks.

Compared to other methods, such as BDT with different approaches [15], [28], [56] or the use of virtual reality [31], [32], [58], the multiscopic approach offers several advantages.

1) First, the feature extraction using the hand tracker in the microscopic model provides accurate estimations of the participant's hand and finger kinematics. This model enables us to analyze the postures used in handling blocks and study the interaction between the hand and the objects in the reach-to-grasp cycle. Future work on this model involves improving the accuracy of hand estimation measurements using cameras, compared to the contact-based methods that involve physically touching the patient's hand.

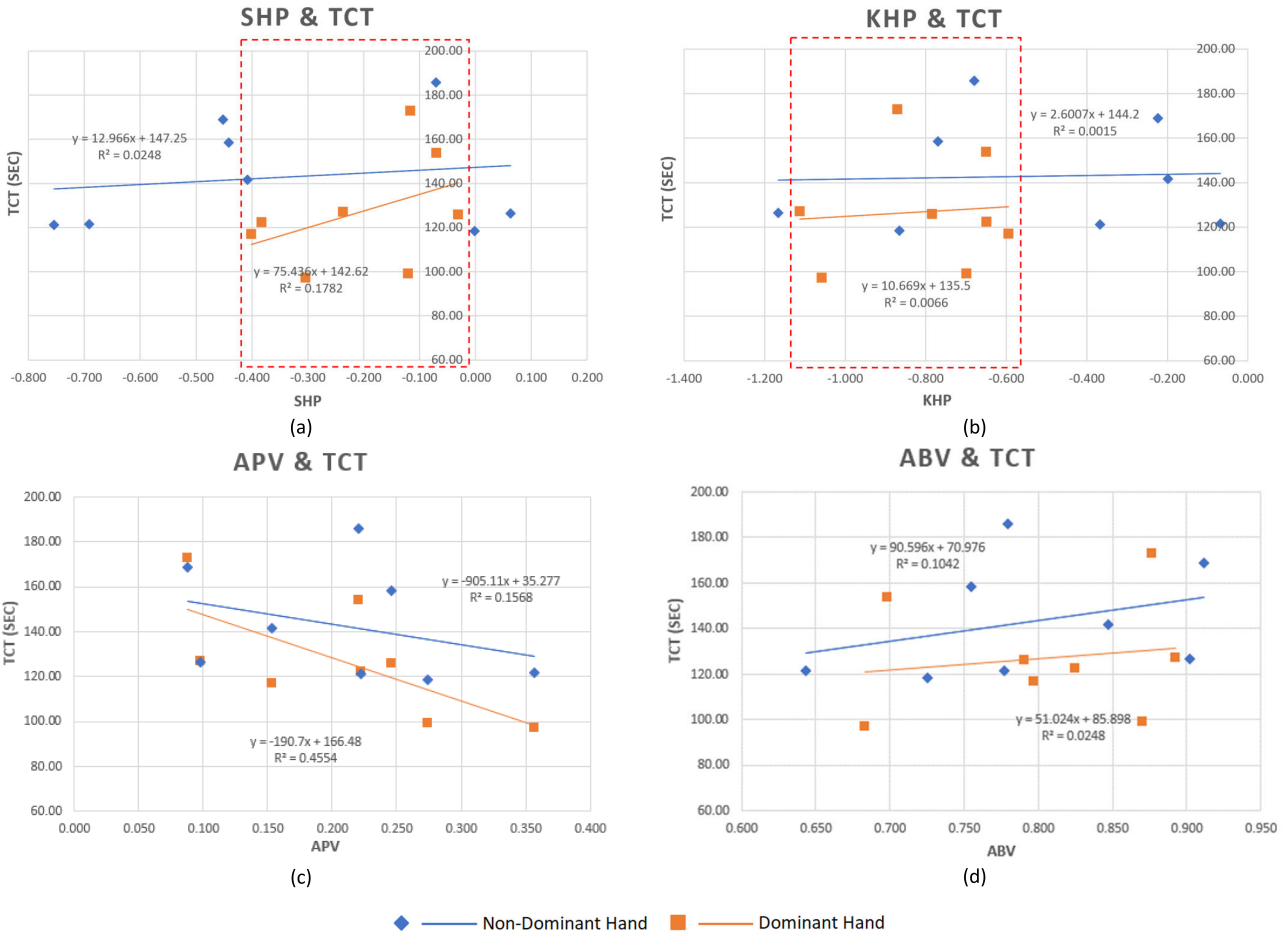

FIGURE 13. Regression and correlation analysis between BDT using non-dominant hand and dominant hand: (a) SHP and TCT, (b) KHP and TCT, (c) APV and TCT, (d) ABV and TCT.

2) Second, in the mesoscopic model, the symbolic interpretation utilizing egocentric vision allows us to capture the behavior model of the participants. This model enables us to analyze the interaction between hands and objects and study the visual attention characteristics of the subjects. Future work on this model could involve exploring applications for self-rehabilitation, where individuals can use the system to improve their motor skills.
3) Third, in the macroscopic model, the knowledge discovery using upper table vision successfully captures the cognitive model of the participants. This model allows us to assess an individual's ability to solve problems ranging from simple to complex. Future work on this model could involve developing a system to predict a person's endurance and ability to concentrate on repetitive tasks.

The multiscopic approach comprehensively analyzes human behavior and performance during BDT, providing valuable insights for rehabilitation and cognitive assessment.

In conclusion, integrating physical measurements and cognitive evaluations in BDT assessment offers numerous advantages that are impossible with conventional methods. To further enhance the understanding of participants' performance in BDT, gathering more detailed information about the factors contributing to success or failure is essential. This method simultaneously analyzes the dominant hand and eye movements to assess the improvement or decline in hand-eye coordination. Such information can significantly assist therapists in developing tailored rehabilitation plans. However, there is a physical limitation in the current data collection process, as it only captures the 2D positions of hands and objects. It is necessary to implement a data acquisition strategy that incorporates 3D information. This strategy will provide a more comprehensive understanding of participants' movements and interactions during the BDT.

Furthermore, validating this proposed method by comparing it with existing BDT practices in rehabilitation facilities is crucial. By conducting such validation, we can confirm the added value of the three models of the multiscopic CPSS in BDT measurements. The goal is to provide therapists and researchers with valuable information typically unavailable in a clinical setting. The next step involves collecting patient samples to validate further and refine this technology for

rehabilitation. The developments resulting from this work will contribute to future cognitive rehabilitation efforts. Ultimately, we aim to utilize these technologies to improve the effectiveness and outcomes of rehabilitation programs.

V. CONCLUSION AND FUTURE WORKS

This paper introduces a multiscopic CPSS framework for supporting independent rehabilitation through HOI recognition with visual attention. The framework effectively integrates the physical and cognitive aspects of the BDT application by incorporating multiple vision systems across three levels. The hand-tracking vision system accurately collects hand-skeletal data and finger joint angle features at the microscopic model. This model enables the classification of physical hand postures into six categories, providing valuable insights into the predominant postures used during grasping and pinching blocks. In the mesoscopic model, the egocentric vision system combined with an eye tracker captures hand and eye movements. The symbolic interpretation successfully categorizes hand-eye coordination during the reach-to-grasp cycle. This analysis sheds light on the influence of hand actions and visual focus on patterns and blocks on the success rate of BDT. In the macroscopic model, the upper table vision system classifies color features in each block. The knowledge discovery accurately assesses whether the design matches the given task, comprehensively understanding participants' cognitive abilities.

The conducted eight-pattern BDT with two scenarios demonstrates the framework's capability to measure participant behavior from multiple perspectives. Results indicate slightly better performance in the dominant hand scenario than in the non-dominant hand scenario. Furthermore, regression and correlation analysis reveals the relationship between physical measurement and cognitive evaluations. This research is expected to benefit significantly therapists and researchers by offering valuable insights not readily available in clinical settings. To further validate and implement the framework in rehabilitation, it is essential to conduct testing on actual patients as part of future research efforts. By implementing this approach, we can augment the practicality and efficacy of the proposed physical and cognitive rehabilitation framework.

ACKNOWLEDGMENT

This work was partially supported by Japan Science and Technology Agency (JST), Moonshot R&D, under grant number JPMJMS2034, and Tokyo Metropolitan University (TMU) Local 5G research support. The authors greatly appreciate the scholarship support from the Japan Ministry of Education, Culture, Sports, Science, and Technology (MEXT).

REFERENCES

[1] T. Singh, C. M. Perry, S. L. Fritz, J. Fridriksson, and T. M. Herter, "Eye movements interfere with limb motor control in stroke survivors," *Neurorehabil. Neural Repair*, vol. 32, no. 8, pp. 724–734, Aug. 2018.

[2] M. Szekeres and K. Valdes, "Virtual health care telehealth: Current therapy practice patterns," *J. Hand Therapy*, vol. 35, no. 1, pp. 124–130, Jan. 2022.

[3] A. Laghari, Z. A. Memon, S. Ullah, and I. Hussain, "Cyber physical system for stroke detection," *IEEE Access*, vol. 6, pp. 37444–37453, 2018.

[4] A. Rashid and O. Hasan, "Wearable technologies for hand joints monitoring for rehabilitation: A survey," *Microelectron. J.*, vol. 88, pp. 173–183, Jun. 2019.

[5] A. A. Saputra, A. R. A. Besari, and N. Kubota, "Human joint skeleton tracking using multiple Kinect Azure," in *Proc. Int. Electron. Symp. (IES)*, Aug. 2022, pp. 430–435.

[6] M. Dousty and J. Zariffa, "Towards clustering hand grasps of individuals with spinal cord injury in egocentric video," in *Proc. 42nd Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Jul. 2020, pp. 2151–2154.

[7] A. R. A. Besari, A. A. Saputra, W. H. Chin, N. Kubota, and Kurnianingsih, "Hand-object interaction detection based on visual attention for independent rehabilitation support," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Padua, Italy, Jul. 2022, pp. 1–6.

[8] T. Wang, T. Yang, M. Danelljan, F. S. Khan, X. Zhang, and J. Sun, "Learning human-object interaction detection using interaction points," 2020, *arXiv:2003.14023*.

[9] P. Wang, L. T. Yang, J. Li, J. Chen, and S. Hu, "Data fusion in cyber-physical-social systems: State-of-the-art and perspectives," *Inf. Fusion*, vol. 51, pp. 42–57, Nov. 2019.

[10] J. Likitlersuang, E. R. Sumitro, T. Cao, R. J. Visée, S. Kalsi-Ryan, and J. Zariffa, "Egocentric video: A new tool for capturing hand use of individuals with spinal cord injury at home," *J. NeuroEng. Rehabil.*, vol. 16, no. 1, p. 83, Dec. 2019.

[11] J.-S. Won and S. Lee, "Geometry-based finger kinematic models for joint rotation configuration and parameter estimation," *Int. J. Adv. Robot. Syst.*, vol. 17, no. 4, Jul. 2020, Art. no. 172988142093057.

[12] L. Lévêque, H. Bosmans, L. Cockmartin, and H. Liu, "State of the art: Eye-tracking studies in medical imaging," *IEEE Access*, vol. 6, pp. 37023–37034, 2018.

[13] B. Banire, D. Al-Thani, M. Qaraqe, K. Khowaja, and B. Mansoor, "The effects of visual stimuli on attention in children with autism spectrum disorder: An eye-tracking study," *IEEE Access*, vol. 8, pp. 225663–225674, 2020.

[14] S. C. Kohs, "The block-design tests," *J. Exp. Psychol.*, vol. 3, no. 5, pp. 357–376, Oct. 1920.

[15] S. Cha, J. Ainooson, and M. Kunda, "Quantifying human behavior on the block design test through automated multi-level analysis of overhead video," 2018, *arXiv:1811.07488*.

[16] D. Qurratu'aini, A. Sophian, W. Sediono, H. Md Yusof, and S. Sudirman, "Visual-based fingertip detection for hand rehabilitation," *Indonesian J. Electr. Eng. Comput. Sci.*, vol. 9, no. 2, p. 474, Feb. 2018.

[17] J. Likitlersuang, R. J. Visée, S. Kalsi-Ryan, and J. Zariffa, "Capturing hand use of individuals with spinal cord injury at home using egocentric video: A feasibility study," *Spinal Cord Ser. Cases*, vol. 7, no. 1, p. 17, Mar. 2021.

[18] A. Bandini, M. Dousty, S. L. Hitzig, B. C. Craven, S. Kalsi-Ryan, and J. Zariffa, "Measuring hand use in the home after cervical spinal cord injury using egocentric video," *J. Neurotrauma*, vol. 39, nos. 23–24, pp. 1697–1707, Dec. 2022.

[19] M.-F. Tsai, R. H. Wang, and J. Zariffa, "Identifying hand use and hand roles after stroke using egocentric video," *IEEE J. Transl. Eng. Health Med.*, vol. 9, pp. 1–10, 2021.

[20] Y. Li, L. Jia, Z. Wang, Y. Qian, and H. Qiao, "Un-supervised and semi-supervised hand segmentation in egocentric images with noisy label learning," *Neurocomputing*, vol. 334, pp. 11–24, Mar. 2019.

[21] Y. Lee, W. Do, H. Yoon, J. Heo, W. Lee, and D. Lee, "Visual-inertial hand motion tracking with robustness against occlusion, interference, and contact," *Sci. Robot.*, vol. 6, no. 58, p. 1315, Sep. 2021.

[22] G. Kapidis, R. Poppe, and R. C. Veltkamp, "Multi-dataset, multitask learning of egocentric vision tasks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 6, pp. 6618–6630, Jun. 2023.

[23] E. B. Fauth, S. Y. Schaefer, S. H. Zarit, M. Ernsth-Bravell, and B. Johansson, "Associations between fine motor performance in activities of daily living and cognitive ability in a nondemented sample of older adults: Implications for geriatric physical rehabilitation," *J. Aging Health*, vol. 29, no. 7, pp. 1144–1159, Oct. 2017.

[24] Y.-H. Chen and J.-H. Chang, "Establishment and discussion of the design criteria for training chopsticks for children," in *Proc. 21st Congr. Int. Ergonom. Assoc. (IEA)*, N. L. Black, W. P. Neumann, and I. Noy, Eds. Cham, Switzerland: Springer 2021, pp. 63–70.

[25] J. Jiang, Z. Nan, H. Chen, S. Chen, and N. Zheng, "Predicting short-term next-active-object through visual attention and hand position," *Neurocomputing*, vol. 433, pp. 212–222, Apr. 2021.
[26] H. Joung, D. Yi, M. S. Byun, J. H. Lee, Y. Lee, H. Ahn, and D. Y. Lee, "Functional neural correlates of the WAIS-IV block design test in older adult with mild cognitive impairment and Alzheimer's disease," *Neuroscience*, vol. 463, pp. 197–203, May 2021.
[27] A. C. Dunn, A. Qiao, M. R. Johnson, and M. Kunda, "Measuring more to learn more from the block design test: A literature review," in *Proc. 43rd Annu. Meeting Cogn. Sci. Soc.*, 2021, pp. 611–617.
[28] S. Cha, J. Ainooson, E. Chong, I. Soulieres, J. M. Rehg, and M. Kunda, "Enhancing cognitive assessment through multimodal sensing: A case study using the block design test," in *Proc. 42nd Annu. Meeting Cogn. Sci. Soc.*, Jan. 2020, pp. 2546–2552.
[29] V. L. Averbukh, N. V. Averbukh, and D. V. Semenischev, "Activity approach in design of specialized visualization systems," *Sci. Vis.*, vol. 11, no. 3, pp. 1–16, 2019.
[30] J. M. Rogers, J. Duckworth, S. Middleton, B. Steenbergen, and P. H. Wilson, "Elements virtual rehabilitation improves motor, cognitive, and functional outcomes in adult stroke: Evidence from a randomized controlled pilot study," *J. Neuroeng. Rehabil.*, vol. 16, no. 1, p. 56, Dec. 2019.
[31] V. Wikström, S. Martikainen, M. Falcon, J. Ruistola, and K. Saarikivi, "Collaborative block design task for assessing pair performance in virtual reality and reality," *Heliyon*, vol. 6, no. 9, Sep. 2020, Art. no. e04823.
[32] K. Shigenaga and K. Nagamune, "A development of kohs block design test in virtual reality with eye tracking and hand tracking," in *Proc. Int. Conf. Mach. Learn. Cybern. (ICMLC)*, Sep. 2022, pp. 271–275.
[33] K. Hesseberg, G. G. Tangen, A. H. Pripp, and A. Bergland, "Associations between cognition and hand function in older people diagnosed with mild cognitive impairment or dementia," *Dementia Geriatric Cogn. Disorders Extra*, vol. 10, no. 3, pp. 195–204, Dec. 2020.
[34] A. R. A. Besari, A. A. Saputra, W. H. Chin, Kurnianingsih, and N. Kubota, "Finger joint angle estimation with visual attention for rehabilitation support: A case study of the chopsticks manipulation test," *IEEE Access*, vol. 10, pp. 91316–91331, 2022.
[35] A. A. Saputra, K. Wada, S. Masuda, and N. Kubota, "Multi-scopic neurocognitive adaptation for legged locomotion robots," *Sci. Rep.*, vol. 12, no. 1, p. 16222, Sep. 2022.
[36] W. Dou, W. Chin, and N. Kubota, "Multi-scopic cognitive memory system for continuous gesture learning," *Biomimetics*, vol. 8, no. 1, p. 88, Feb. 2023.
[37] A. R. A. Besari, A. A. Saputra, W. H. Chin, Kurnianingsih, and N. Kubota, "Hand–object interaction recognition based on visual attention using multiscopic cyber-physical-social system," *Int. J. Adv. Intell. Informat.*, vol. 9, no. 2, p. 187, Jul. 2023.
[38] K. D. Nguyen, L. A. Corben, P. N. Pathirana, M. K. Horne, M. B. Delatycki, and D. J. Szmulewicz, "The assessment of upper limb functionality in friedreich ataxia via self-feeding activity," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 28, no. 4, pp. 924–933, Apr. 2020.
[39] A. Ganguly, G. Rashidi, and K. Mombaur, "Comparison of the performance of the leap motion controller^{TM} with a standard marker-based motion capture system," *Sensors*, vol. 21, no. 5, p. 1750, Mar. 2021.
[40] C. Dai and X. Hu, "Finger joint angle estimation based on motoneuron discharge activities," *IEEE J. Biomed. Health Informat.*, vol. 24, no. 3, pp. 760–767, Mar. 2020.
[41] S. Jegelka, "Theory of graph neural networks: Representation and learning," 2022, *arXiv:2204.07697*.
[42] M. Fey and J. Eric Lenssen, "Fast graph representation learning with PyTorch geometric," 2019, *arXiv:1903.02428*.
[43] A. R. A. Besari, W. H. Chin, N. Kubota, and Kurnianingsih, "Ecological approach for object relationship extraction in elderly care robot," in *Proc. 21st Int. Conf. Res. Educ. Mechtron. (REM)*, Kraków, Poland, Dec. 2020, pp. 1–6.
[44] H. R. Nasrabadi and J.-M. Alonso, "Modular streaming pipeline of eye/head tracking data using Tobii pro glasses 3," *Animal Behav. Cogn.*, vol. 2022, pp. 1–17, Sep. 2022.
[45] C.-Y. Wang, A. Bochkovskiy, and H.-Y. Mark Liao, "YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors," 2022, *arXiv:2207.02696*.
[46] H. Fu, L. Wu, M. Jian, Y. Yang, and X. Wang, "MF-SORT: Simple online and realtime tracking with motion features," in *Image Graphics* (Lecture Notes in Computer Science), vol. 11901, Y. Zhao, N. Barnes, B. Chen, R. Westermann, X. Kong, and C. Lin, Eds. Cham, Switzerland: Springer, 2019, pp. 157–168.
[47] F. Zhang, V. Bazarevsky, A. Vakunov, A. Tkachenka, G. Sung, C.-L. Chang, and M. Grundmann, "MediaPipe hands: On-device real-time hand tracking," 2020, *arXiv:2006.10214*.
[48] Q. Cai, J. Li, and J. Long, "Effect of physical and virtual feedback on reach-to-grasp movements in virtual environments," *IEEE Trans. Cogn. Develop. Syst.*, vol. 14, no. 2, pp. 708–714, Jun. 2022.
[49] S. Shin and W. Kim, "Skeleton-based dynamic hand gesture recognition using a part-based GRU-RNN for gesture-based interface," *IEEE Access*, vol. 8, pp. 50236–50243, 2020.
[50] E. P. Larsen, J. M. Kolman, F. N. Masud, and F. Sasangohar, "Ethical considerations when using a mobile eye tracker in a patient-facing area: Lessons from an intensive care unit observational protocol," *Ethics Hum. Res.*, vol. 42, no. 6, pp. 2–13, Nov. 2020.
[51] A. R. A. Besari, A. A. Saputra, W. H. Chin, N. Kubota, and Kurnianingsih, "Feature-based egocentric grasp pose classification for expanding human-object interactions," in *Proc. IEEE 30th Int. Symp. Ind. Electron. (ISIE)*, Kyoto, Japan, Jun. 2021, pp. 1–6.
[52] M. K. Cain, Z. Zhang, and K.-H. Yuan, "Univariate and multivariate skewness and kurtosis for measuring nonnormality: Prevalence, influence and estimation," *Behav. Res. Methods*, vol. 49, no. 5, pp. 1716–1735, Oct. 2017.
[53] T. J. Bosch, T. Hanna, K. A. Fercho, and L. A. Baugh, "Behavioral performance and visual strategies during skill acquisition using a novel tool use motor learning task," *Sci. Rep.*, vol. 8, no. 1, p. 13755, Sep. 2018.
[54] M. Yani, A. R. A. Besari, N. Yamada, and N. Kubota, "Ecological-inspired system design for safety manipulation strategy in home-care robot," in *Proc. Int. Symp. Community-Centric Syst. (CcS)*, Hachioji, Tokyo, Japan, Sep. 2020, pp. 1–6.
[55] S. Wang, J. Zhao, C. Shao, C. Dong, and C. Yin, "Truck traffic flow prediction based on LSTM and GRU methods with sampled GPS data," *IEEE Access*, vol. 8, pp. 208158–208169, 2020.
[56] M. Kunda, M. E. Banani, and J. M. Rehg, "A computational exploration of problem-solving strategies and gaze behaviors on the block design task," in *Proc. 38th Annu. Meeting Cogn. Sci. Soc.* Philadelphia, PA, USA, 2016, pp. 235–240.
[57] S. Li, Y. Li, Y. Li, M. Li, and X. Xu, "YOLO-FIRI: Improved YOLOv5 for infrared image object detection," *IEEE Access*, vol. 9, pp. 141861–141875, 2021.
[58] N. A. Hashim, N. A. A. Razak, and N. A. A. Osman, "Comparison of conventional and virtual reality box and blocks tests in upper limb amputees: A case-control study," *IEEE Access*, vol. 9, pp. 76983–76990, 2021.
[59] J. Négyesi, P. Négyesi, T. Hortobágyi, S. Sun, J. Kusuyama, R. M. Kiss, and R. Nagatomi, "Handedness did not affect motor skill acquisition by the dominant hand or interlimb transfer to the non-dominant hand regardless of task complexity level," *Sci. Rep.*, vol. 12, no. 1, Oct. 2022, Art. no. 18181.
[60] L. Wang, J. Liu, and J. Lan, "Feature evaluation of upper limb exercise rehabilitation interactive system based on kinect," *IEEE Access*, vol. 7, pp. 165985–165996, 2019.

ADNAN RACHMAT ANOM BESARI (Graduate Student Member, IEEE) received the Bachelor of Applied Science degree in electronics engineering from Politeknik Elektronika Negeri Surabaya (PENS), in 2008, and the Master of Science degree in robotics and automation from the Faculty of Manufacturing Engineering, Universiti Teknikal Malaysia Melaka (UTeM), in October 2011. He is currently pursuing the Ph.D. degree with Tokyo Metropolitan University (TMU), Japan. He was the Head of the Real-Time Programming Laboratory, from 2017 to 2019. He is a Lecturer with the Department of Informatics and Computer Engineering, PENS. His research interests include computer vision, educational robotics, the Internet of Things, and elderly care systems.

AZHAR AULIA SAPUTRA (Member, IEEE) received the Bachelor of Applied Science degree in electronic engineering from Politeknik Elektronika Negeri Surabaya, Indonesia, in March 2014, the Master of Engineering and Doctoral of Philosophy degrees from Tokyo Metropolitan University, Japan, in March 2018 and 2021, respectively, and the Ph.D. degree from Tokyo Metropolitan University under JSPS Research Fellow–DC1, in 2021. He is currently an Assistant Professor with Tokyo Metropolitan University. His research interests include intelligent control systems, neural-based locomotion systems, and robotics. He received the Bronze Award from the Capstone Design Fair International Session, South Korea, and several achievements in national and international robotic competitions. He received the Excellent Graduate School Research Award from the Japan Society of Automotive Engineers (JSAE), in 2020.

TAKENORI OBO (Member, IEEE) received the Ph.D. degree from the Department of Human Mechatronics Systems, Graduate School of System Design, Tokyo Metropolitan University, in 2014. In April 2014, he became a specially appointed Assistant Professor with the Human Mechatronics Systems Course, Faculty of System Design, Tokyo Metropolitan University. In October 2015, he was specially appointed as an Assistant Professor with the Intelligent Mechanical Systems Course, Faculty of System Design, Tokyo Metropolitan University. Since April 2017, he has been an Assistant Professor with the Department of Computer Applications, Faculty of Engineering, Tokyo Polytechnic University. In April 2022, he became an Assistant Professor with the Department of Mechanical Systems Engineering, Faculty of System Design, Tokyo Metropolitan University. He is currently with the Department of Mechanical Systems Engineering, Faculty of System Design, Tokyo Metropolitan University. He is a member of the Society of Instrument and Control Engineers, the Japan Society of Intelligent Information Fuzzy, the Robotics Society of Japan, and the Institute of Systems, Control and Information Engineers.

KURNIANINGSIH (Senior Member, IEEE) received the B.Eng. degree in informatics engineering from Telkom University, Indonesia, the M.Eng. degree in electrical engineering from North Sumatera University, Indonesia, the Ph.D. degree in electrical engineering from Universitas Gadjah Mada, Indonesia, and the engineering degree from Institut Teknologi Bandung, Indonesia. She is currently an Assistant Professor with the Department of Electrical Engineering, Politeknik Negeri Semarang, Indonesia. Her current research interests include sensor networks, machine learning, and computational intelligence. She has been an Executive Committee Member of the IEEE Region 10 (Asia-Pacific Region), since 2018, appointed as the Information Management Committee Chair. She is a member of the IEEE Computational Intelligence Society (IEEE CIS) and the IEEE Systems, Man, and Cybernetics Society (IEEE SMCS). She received the IEEE MGA Achievement Award, in 2020, and the IEEE Region 10 Young Professionals Award in Academician, in 2018. She was the past Vice-Chair of the IEEE Indonesia Section.

NAOYUKI KUBOTA (Senior Member, IEEE) received the degree from Osaka Kyoiku University, in 1992, the M.E. degree from Hokkaido University, in 1994, and the D.E. degree from Nagoya University, Japan, in 1997. He was a Visiting Professor with the University of Portsmouth, U.K., in 2007 and 2009. He was an Invited Visiting Professor with Seoul National University, from 2009 to 2012. He is currently a Professor with the Department of Mechanical Systems Engineering, Graduate School of Systems Design, Tokyo Metropolitan University, Japan. His current research interests include coevolutionary computing, fuzzy computing, topological mapping, cognitive robotics, social robotics, and informationally structured space. He has published more than 500 refereed journals and conference papers in the above research fields. He received the Best Paper Award from IEEE IECON 1996 and the Best Paper Award from IEEE CIRA 1997, amongst others. He serves as the Vice Director for the Tokyo Biomarker Innovation Research Association, Japan, and the Chair for the IEEE Society on Systems, Man, and Cybernetics, Japan Chapter. He was an Associate Editor of the IEEE TRANSACTIONS ON FUZZY SYSTEMS, from 1999 to 2010, the IEEE CIS Intelligent Systems Applications Technical Committee, Robotics Task Force Chair, from 2007 to 2014, the IEEE Systems, Man, and Cybernetics Society, Japan Chapter Chair, since 2018, to name a few.