Article

Prediction Models of Collaborative Behaviors in Dyadic Interactions: An Application for Inclusive Teamwork Training in Virtual Environments

Ashwaq Zaini Amat [1,*], Abigale Plunk [1], Deeksha Adiani [2], D. Mitchell Wilkes [1] and Nilanjan Sarkar [1,2,3]

1   Department of Electrical and Computer Engineering, Vanderbilt University, Nashville, TN 37240, USA
2   Department of Computer Science, Vanderbilt University, Nashville, TN 37240, USA
3   Department of Mechanical Engineering, Vanderbilt University, Nashville, TN 37240, USA
*   Correspondence: ashwaq.zaini.amat.haji.anwar@vanderbilt.edu

Abstract: Collaborative virtual environment (CVE)-based teamwork training offers a promising avenue for inclusive teamwork training. The incorporation of a feedback mechanism within virtual training environments can enhance the training experience by scaffolding learning and promoting active collaboration. However, an effective feedback mechanism requires a robust prediction model of collaborative behaviors. This paper presents a novel approach using hidden Markov models (HMMs) to predict human behavior in collaborative interactions based on multimodal signals collected from a CVE-based teamwork training simulator. The HMM was trained using k-fold cross-validation, achieving an accuracy of 97.77%. The HMM was evaluated against expert-labeled data and compared against a rule-based prediction model, demonstrating the superior predictive capabilities of the HMM, with the HMM achieving 90.59% accuracy compared to 76.53% for the rule-based model. These results highlight the potential of HMMs to predict collaborative behaviors that could be used in a feedback mechanism to enhance teamwork training experiences despite the complexity of these behaviors. This research contributes to advancing inclusive and supportive virtual learning environments, bridging gaps in cross-neurotype collaborations.

Keywords: human behavior recognition; human–computer interaction; probabilistic modeling; collaborative virtual environment; cross-neurotype collaboration

1. Introduction

Human–computer interaction (HCI) technologies have become prevalent tools for facilitating skill learning, offering engaging interactions and replicable solutions to benefit and enhance learning experiences [1–3]. These systems teach a range of skills, including cognitive abilities such as visual–spatial, auditory, and verbal skills [4]; affective learning such as emotion regulation [5]; and collaboration skills [6–8]. Real-time prompts and feedback through visual, audio, and tactile cues are integral features of these systems that help in enhancing user engagement and learning experiences [9,10].

In our prior work, we developed a series of virtual collaborative tasks as a teamwork training simulator within a collaborative virtual environment (CVE), focusing on facilitating dyadic interaction [11]. While participants found the training paradigms engaging, we identified a need for real-time feedback mechanisms to support participants during these collaborative tasks. Recognizing human behavior in feedback mechanisms is crucial for effective training outcomes [12]. Recent studies reported that incorporating human behavior recognition, along with task performance, in adaptive training paradigms can significantly enhance participants' engagement and learning outcomes [7,8,13]. Current methods for behavior recognition rely heavily on behavioral experts to observe and evaluate human behavior either in real-time during experimental sessions [14] or through analysis of video recordings post-experiment [15–17]. While manual labeling of human behavior is reliable, it

lacks real-time accessibility, is resource-intensive, and is prone to bias [18–20], highlighting the necessity for an automated human behavior prediction model.

Various machine learning methods have been employed to predict human behavior in computer-based interactions [21,22], aiming to address the inherent uncertainty in recognizing human behaviors [19]. Traditional machine learning approaches have been pivotal, with several studies achieving notable success [23–26]. For instance, Abdelrahman et al. [27] used deep learning and neural network methods to predict engagement and disengagement in human–robot interaction by extracting and scoring engagement-related features from human participants, such as gaze, head pose, and body posture, to infer engagement. The model achieved a 93% F1 score. Many of these studies used multimodal signals for the classification of behavior to improve the classification accuracy due to the complex nature of human behavior. According to a meta-analysis review of 30 studies that compared affect detection accuracy between multimodal and unimodal signals, researchers reported that accuracies based on multimodal data fusion were consistently better than those based on unimodal signals [28]. This is further supported by the findings in a study by Mallol-Ragolta et al. [29], where they reported the best agreement score using multimodal signals compared to a unimodal signal in a robotic empathy recognition system. In another study by Okada et al. [30], signals from speech and head movement were captured and both verbal and non-verbal features were extracted from the signals to assess the collaborative behavior in discussions.

To handle the stochastic nature of human behavior, hidden Markov models (HMMs) have been widely used [31,32], offering robustness and flexibility in analyzing temporal patterns which are necessary for predicting behavioral patterns (e.g., emotions, activities, learning behaviors such as engagement, disengagement, confusion, frustration, distress, etc.) [33–38]. Mihoub et al. [39] demonstrated the effectiveness of an incremental discrete hidden Markov model (IDHMM) to recognize and generate multimodal joint actions in face-to-face interactions. The results reported that the classification accuracy of IDHMM was 92%, while a support vector machine (SVM)'s classification accuracy was 81%. Another study compared HMM performance against traditional classification models that included support vector machine (SVM), random forest (RF), linear regression (LR), and deep neural network (DNN) in predicting students learning behavior in e-learning environments [37]. The study utilized early assessment data and results indicated that HMM outperformed the other models for 5 of the 6 courses with accuracies above 90%. Similarly, Sharma et al. [34], utilized an HMM to predict effortful behavior in adaptive learning environments using both performance and physiological data, enabling real-time feedback based on predicted behavioral patterns. In this work, our first contribution is the design of an HMM-based model to recognize collaborative behaviors in dyadic interactions using multimodal signals, aiming to provide real-time feedback and scaffold learning in computer-based interactions. As a baseline for comparison, we implemented a rule-based behavior recognition method designed in consultation with experienced behavior analysts. We compared the HMM model against the rule-based method with expert labeled data as the ground truth.

While previous works have focused on behavior recognition in single-user interactions, recognizing human behavior in dyadic interactions, such as teamwork, remains underexplored. Teamwork, defined as collaborative work between individuals toward a common goal, has garnered increased research attention in recent years, particularly within organizational contexts [40,41]. In addition to the benefits teamwork brings to a company, teamwork also leads to increased satisfaction in the workplace, which can fulfill personal growth [42]. As society embraces inclusive workplaces with neurodiverse individuals, research interest on cross-neurotype collaboration, i.e., collaboration between neurotypical (people with "normal" neurotypes) and neurodiverse individuals, has increased [43]. Studies have shown that cross-neurotype collaboration can be less effective than collaborations between individuals within the same neurotype [44]. This phenomenon, driven by the double empathy problem [45], underscores that challenges in cross-neurotype communication and social interaction are the responsibilities of both neurotypical and autistic individu-

als [46,47]. Our second contribution is the implementation of the prediction model in our previously designed teamwork training simulator to support cross-neurotype teamwork training.

Additionally, we contribute to the field by creating a novel dataset containing multimodal signals from dyadic interactions labeled with collaborative behaviors by expert annotators. The outcome of this work can motivate future research to incorporate a robust behavioral prediction model in a feedback mechanism within virtual training environments that could enhance training experiences by scaffolding learning and promoting active collaboration. By creating a novel prediction model and a multimodal dyadic interaction dataset, our work seeks to advance the integration of robust behavioral prediction mechanisms into computer-based teamwork training, ultimately enhancing collaborative skill development.

2. Materials and Methods

2.1. Experimental Design

We conducted a preliminary study to gather multimodal signals from participants interacting with each other to complete various collaborative tasks in a CVE. The signals were then processed, analyzed, and labeled based on defined collaborative behaviors. The labeled signals were used to (i) train an HMM prediction model, (ii) design a rule-based prediction model, and (iii) evaluate both models.

2.1.1. Collaborative Tasks Description

The collaborative task selection was driven by employment-related studies for autistic individuals [48]. We then designed the activities within the tasks based on input from stakeholders, including human resource personnel from several companies, certified behavioral analysts, career counselors, and autistic adults. They provided suggestions and feedback to encourage teamwork in a workplace environment between an autistic individual and a neurotypical (non-autistic) partner, which was discussed in detail in our previous work [11]. Multiple discussion sessions with the stakeholders were conducted to select tasks that are collaborative and include interactions that were translatable to workplace environments.

The first task was a PC assembly task in which two participants were located on opposite ends of a table in the virtual environment, giving them different views of the workspace. They both were given written instructions and different hardware to collaboratively build a single computer. They would use a keyboard and mouse to move the components into the correct location within a set amount of time. Participants were required to take turns and communicate with each other when assembling the PC. The next task was a furniture assembly task in which participants were placed in a virtual living room and worked together to assemble various furniture pieces within a set amount of time. They used a haptic device to move and assemble the furniture parts to the target area. The final task was a fulfillment center task in which participants would drive virtual forklifts with varying height capacities to transport crates from a warehouse to a drop-off location. Participants used a gamepad to drive the forklift in this task. These collaborative tasks, as illustrated in Figure 1, were designed in Unity, a multi-platform game development software [49].

Three design strategies were embedded within the tasks to encourage communication and collaboration between the participants: (a) PC assembly: incomplete installation instructions were given to each participant to encourage them to exchange information to progress in the task; (b) furniture assembly: participants were given only an image of the assembled furniture, without written instruction, to encourage them to divide the task and coordinate their actions; and (c) fulfillment center: the list of crates for each participant were different and the location of the crates varied to allow participants to practice turn-taking.

Step-by-step instruction to build a PC

PC ASSEMBLY

Image of a bookshelf to be assembled

FURNITURE ASSEMBLY

Map of the warehouse to locate the palate

FULFILLMENT CENTER

Figure 1. Collaborative tasks to support collaborative interaction between autistic individuals and neurotypical partners.

2.1.2. Participants and Protocol

We recruited six autistic (ASD) and six neurotypical (NT) participants to form six cross-neurotype (ASD-NT) participant pairs. The demographics for the participants are shown in Table 1. Participants with ASD were recruited through an existing university-based clinical research registry and the NT participants were recruited from the local community through regional advertisement. All study procedures were approved by the Vanderbilt University's Institutional Review Board (IRB) with associated procedures for informed assent and consent. Figure 2 illustrates the setup of the experiment.

Table 1. Participants demographic information.

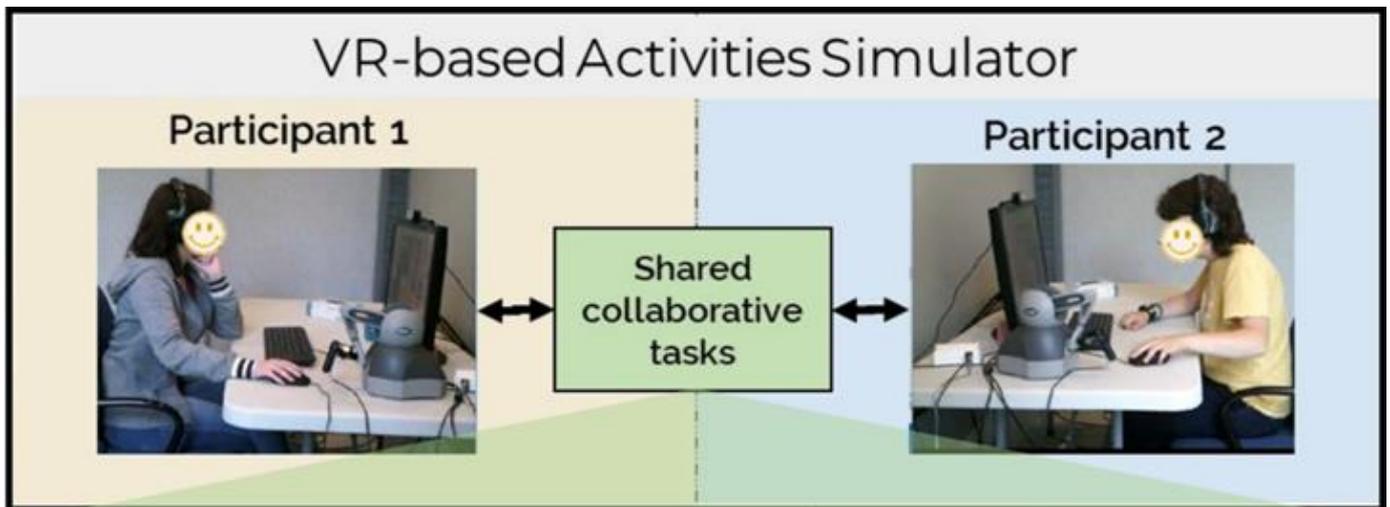| Participants	ASD (N = 6)	NT (N = 6)
	Mean (SD)	Mean (SD)
Age	20.5 (2.8)	22.8 (3.6)
Gender (% male-female)	50%-50%	50%-50%
Race (% White, % African American)	100%	83%, 0%
Ethnicity (% Hispanic)	0%	17%

Figure 2. System setup where two participants in separate rooms perform virtual collaborative tasks together.

2.2. Prediction Models Workflow

We describe four main processes involved in designing, training, and evaluating behavior prediction models in collaborative interactions, as seen in Figure 3. First, we captured multimodal data from both participants and performed signal processing to design the prediction models. Then, the multimodal signals together with video recordings were used by annotators to label the participants' collaborative behavior, which we defined as either *"Engaged"*, *"Waiting"*, or *"Struggling"*, to establish ground truth. These labeled data were used to design a rule-based prediction model and train an HMM. We then evaluated both prediction models' performances. The following subsections explain each step of the process in detail.
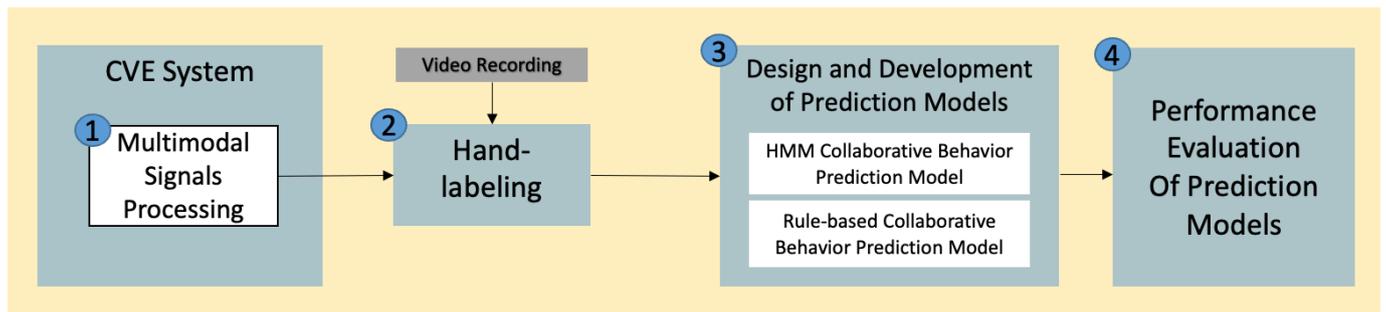
Figure 3. Workflow for the design and evaluation of the prediction models.

2.2.1. Multimodal Signal Processing

The multimodal signals were captured from three devices integrated into the collaborative system. We used the signals from (i) task-dependent controllers, (ii) a microphone headset, and a (iii) Tobii EyeX eye tracker that were set up for each participant to extract seven binary features used to recognize the behavior of the participants in collaborative interactions. As an initial approach to analyze the signals, we chose to represent the features as binary as it allows for simplified analysis with still dependable results [4,50]. The diagram in Figure 4 shows that we derived one feature, *Speech Presence*, from the microphone headset as a measure of communication between the participants. Then, from the eye tracker, we extracted the *Gaze Presence* feature and *Gaze on Object* feature to measure participant's focus in the task. Finally, we extracted four features from the controllers based on the presence of the controller, represented by the *Controller Presence* and *Controller Manipulation*, and the distance of the virtual object from a target location as a measure of task progression, as *Object Move Closer* and *Object Move Away.* The feature values were either 1 or 0 representing the presence or absence of the feature. We describe the selection of the feature values in more detail in Table 2.

Table 2. Description of binary features determination from input devices.

Device	Binary Feature	Feature Description
Microphone headset	Speech Presence	Feature is set to "1" when participant is speaking and "0" otherwise.
Tobii EyeX eye tracker	Gaze Presence	Feature is set to "1" when participant's gaze detected on screen and "0" otherwise.
	Gaze On Object	Feature is set to "1" when gaze is on a virtual object or within the defined "focus area" as depicted in Figure 5.
Task-dependent controller (keyboard, haptic, or game controller)	Controller Presence	Feature is set to "1" when an input is detected from the controller (keyboard button, mouse clicks, haptic presses) and "0" otherwise.
	Controller Manipulation	Feature is set to "1" when controller is actively moving an object, and "0" otherwise.
	* Object Move Closer	Feature is set to "1" when the distance of the object from the target location is decreasing, and "0" otherwise.
	* Object Move Away	Feature is set to "1" when the distance of the object from the target location is increasing, and "0" otherwise.

* When both *Object Move Closer* and *Object Move Away* are set to "0" at the same time, it means that the object is stationary.
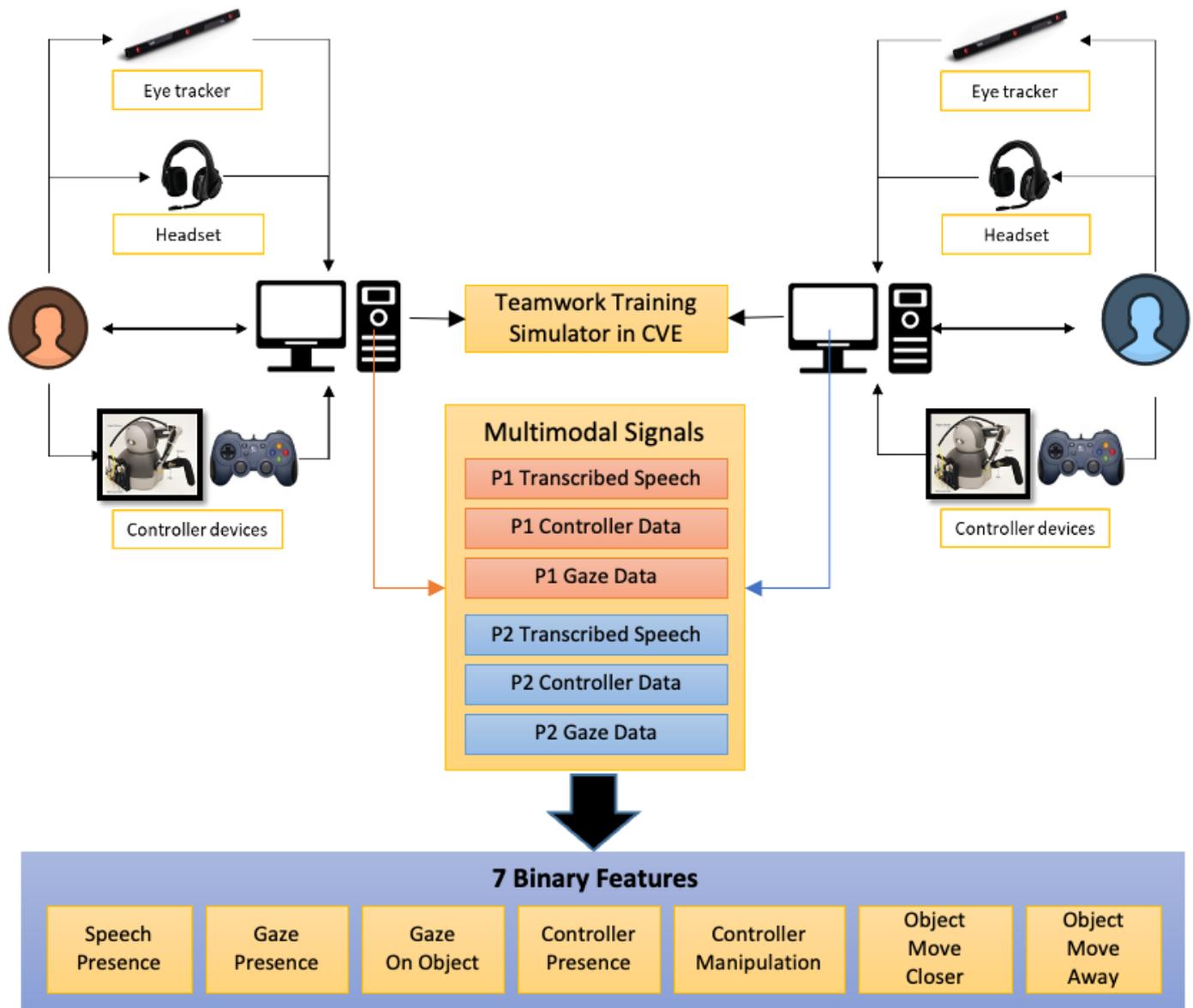
Figure 4. Feature extraction from multimodal signals coming from three peripheral devices.

All the features were collected with a sampling rate of 1 Hz. These binary features were concatenated to form a feature vector (e.g., [0 1 0 1 0 1 0]) for the HMM design, while individual binary values were used as input to the rule-based model. A similar concatenation of the features was also used by Khamparia et al. [51] in their HMM application to investigate psychological and environmental factors to help improve learners' performance. As an example, based on the description in Table 2, the combination of the features [0 1 0 1 0 1 0] would represent *speech absence*, *gaze presence*, *controller manipulated*, and *object moving closer to the target*.

Figure 5. Example of virtual objects and focus areas defined for the eye tracker.

2.2.2. Collaborative Behavior Coding Scheme

A literature review on collaborative learning showed that the most frequent and prevalent behavior that could influence collaborative interactions were *engagement* [52], *struggling* [53,54], and *boredom* [55,56]. Engagement could represent positive collaborative interactions while struggling and boredom could indicate a negative collaborative experience that would require intervention. Using this literature review and discussions with the stakeholders and behavioral analysts, we chose the following three behaviors that would be the most useful in recognizing the initial collaboration level in our teamwork training simulator, which will henceforth be referred to as collaborative behaviors: *Engaged*, *Struggling*, and *Waiting*. Note that boredom was replaced with waiting in our application as it can be indicative of a negative collaborative experience. By focusing on when participants are waiting or struggling, the focus can be shifted to prevent boredom or disengagement in the system. These three behaviors represent essential stages of teamwork, allowing the system to provide informed and meaningful feedback. *Engaged* captures the behavior of the participant when performing the task and collaborating with their partner [57], allowing the system to provide positive feedback, such as "Good job!" or "Keep up the good work!". *Struggling* represents the behavior of the participant when they were not progressing in the task (e.g., the task object was moving away from the target), were not interacting with their partner, or were disengaged with the task (e.g., looking outside the focus area for some time) [58]. The system would then use the *Struggling* behavior as an indicator to prompt the participants to collaborate—for example, "Ask your partner to help you with the task" and to the other participant "Your partner seems to be struggling, offer them help". Turn-taking is part of teamwork and collaborative interaction. As such, we used the *Waiting* behavior to represent the behavior when the participant was on standby while their partner was performing a task [59]. This *Waiting* behavior is different from when a participant is not progressing in the task due to being distracted or disinterested (which is categorized under *Struggling*). In the *Waiting* behavior, the system would allocate some time for the participants to wait without prompting the participants. Although there are only three behaviors of collaboration discussed in this work, other behaviors could be

added in the future based on the need and understanding of collaboration and teamwork. A definition of the collaborative behaviors was defined in consultation with a certified behavioral analyst, to ensure the consistency of the manual labeling, shown in Table 3.

Table 3. Definition of Participant's Collaborative Behavior.

#	Collaborative Behavior	Definition	Condition
1	Engaged	The participant is focused on the task, communicating, and progressing well.	Participant could be talking to their partner. Participant is using the controller and virtual object is moving closer to the target. *Engaged = Speech Presence ∪ (Controller Manipulation ∩ Object Move Closer)*
2	Struggling	The participant is not progressing with the task due to difficulty performing the task, not communicating with their partner, distracted, or disinterested with the task.	Participant is not talking to their partner while: i. manipulating the controller but virtual object moving away from the target, or ii. not manipulating the controller and not looking at the screen (virtual objects, focused area). *Struggling = ¬Speech Presence ∩ ((Controller Manipulation ∩ Object Move Away) ∪ (¬Controller Manipulation ∩ ¬Gaze))*
3	Waiting	The participant is on standby for their partner in a turn-taking task, not moving.	Participant is not talking to their partner, not using the controller, and not moving virtual objects, but is looking at an object or focus area. *Waiting = ¬Speech Presence ∩ ¬Controller Manipulation ∩ ¬Object Move Away ∩ ¬Object Move Closer ∩ Gaze*

2.2.3. Hand Labelling to Establish Ground Truth

Two annotators trained by a certified behavioral analyst used the collaborative behaviors defined in Table 3 to label the participants' behavior as either Engaged, Struggling, or Waiting based on the extracted features discussed in Section 2.2.1 and from watching video recording of the sessions. The annotators labeled 10 min of interactions from each session, for all six experimental sessions individually, resulting in 4976 hand-labeled datapoints. They achieved 98% agreement, and the remaining 2% disagreement was reconciled where both annotators decided on an agreed label through discussion. From the labeled data, the class distributions of the three behaviors were as follows: *Engaged*—19.9%, *Struggling*—28.0%, and *Waiting*—52.1%. The hand-labeled data distribution shows that the behaviors are not equally distributed, and the majority of the labeled behavior was Waiting since the tasks mainly involved turn-taking. When designing the HMM prediction model, the imbalance in the data distribution was taken into consideration to avoid overfitting and bias to the prediction model. To achieve this, we used *k*-fold cross-validation to minimize the data imbalanced. This is explained in more detailed in Sections 2.2.5 and 3.1.

2.2.4. Rule-Based Prediction Model Design

We gathered the inputs and feedback from the behavioral analyst when developing the coding scheme in Section 2.2.3 into a set of rules for each collaborative behavior based on the binary features. The rules were constructed to closely replicate the role of human annotators. The seven binary features that were discussed in Section 2.2.1 were used to drive the categorization of the collaborative behaviors defined in Table 3. In the rule-based model, we begin by checking the presence of speech. Since speech data contributes between

20–30% of the entire collaborative interaction, in this initial design, we assumed that any utterances while performing the task as an indication of engagement. Further analysis of the speech in future work would allow us to categorize the behavior more accurately (i.e., positive utterances as *Engaged*, and negative utterances as *Struggling*). If speech was detected, the rule would assign the collaborative behavior as being *Engaged*. If speech was not detected, the second rule was to check for keypresses, which were based on controller manipulation features. If controller manipulation was present, it would set the keypresses as true, and move to the next rule to check the object distance based on the object move closer and object move away features. If the object move away feature was true, then it means the object was moving away from the target, and the rule would assign the collaborative behavior as *Struggling*, If the object move closer feature was true, then it means the object is moving closer to the target, the rule would assign the collaborative behavior as *Engaged*, and if neither the object move closer and object move away are true, then it means that the object was not moving, so the rule would assign the collaborative behavior as *Waiting*. In the case when no speech was present and no keypresses were present, the rule would check for eye gaze presence. If eye gaze was present, participant collaborative behavior was assigned as *Waiting*. However, if eye gaze was absent, the collaborative behavior was assigned as *Struggling*. We then consolidated these rules into a rule-based model as shown by the flow chart in Figure 6.
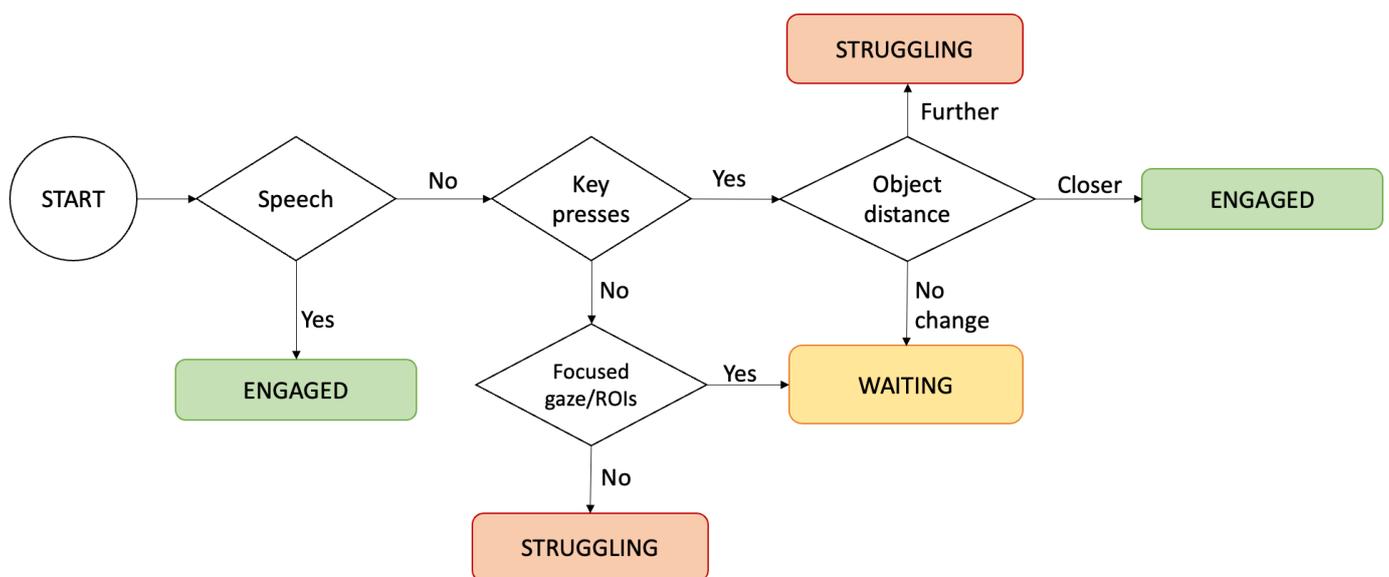

Figure 6. Flowchart for the rule-based prediction model.

2.2.5. HMM Design and Training

A Hidden Markov Model (HMM) is a probabilistic graphical model used to represent systems that evolve over time offering flexibility and scalability compared to deterministic predictive models [60]. It comprises of five main elements [61] shown in Table 4. The first is the set of hidden states (N), which signifies the unobservable underlying conditions or states within the system. In our application, that is our three defined collaborative behaviors: *Engaged*, *Struggling*, and *Waiting*. Second, there are observations (M) associated with each state. In our application these observations are the seven binary features explained in Section 2.2.1.

Table 4. Definition of HMM elements.

Symbol	Definition	Values
N	Number of hidden states in the model.	$\{Engaged,\ Struggling, Waiting\}$
M	Number of distinct observations.	We are using a 7-digit binary vector based on the extracted features from the multi-modal data. $\{Obs_1,\ Obs_2, Obs_3, \ldots, Obs_M\}$ Example values: 1101010, 0010100
A	State transition probability distribution—Probability matrix of transition from one state to another.	Matrix size is $N \times N$, in our case 3×3. The values of the matrix are generated from training the model. $$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$$
B	Emission probability distribution—Probability matrix of observing a particular observation in the current state.	Matrix size is $N \times M$. The values of the matrix are generated from training the model. $$\begin{bmatrix} b_{11} & \cdots & b_{1M} \\ \vdots & \ddots & \vdots \\ b_{31} & \cdots & b_{3M} \end{bmatrix}$$
π	Initial state probability distribution.	Initial state probability matrix, usually equally distributed. $$\begin{bmatrix} 0.3 & 0.3 & 0.3 \end{bmatrix}$$

The model's dynamics are governed by state transition probabilities, represented as a state transition matrix (A). The matrix encodes the probabilities of transitioning from one state to another at each time step, reflecting how the system evolves over time. This matrix is generated when training the model. In addition to the transition matrix, there is an emission matrix (B) that defines the likelihood of generating a particular observation given the current state. Finally, the model requires an initial probability distribution (π) which specifies the initial likelihood of beginning the sequence in each hidden state.

Mathematically, HMMs address two fundamental problems: the evaluation problem, solved using the Forward Algorithm [62], which quantifies the likelihood that the HMM generated a specific sequence of observations, and the decoding problem, solved using the Viterbi Algorithm [63] or the Baum-Welch Algorithm [64], which determines the most probable sequence of hidden states that generated a given sequence of observations. In our application, we train the HMM by calculating the maximum likelihood estimate of the transition (A) and emission (B) probabilities for a sequence of distinct observations (M) with known states (N). Using the estimated transition and emission probabilities, we use the Baum Welch Algorithm to determine the most probable sequence of hidden states for the remainder of our observations. An ergodic state transition model was designed for our model as we assumed that the collaboration state can change from one state to any of the other states. Figure 7 shows a possible diagram of the HMM.

The HMM training was done in MATLAB [65] using the Statistics and Machine Learning Toolbox [66]. The MATLAB function hmmestimate was used to generate estimated transition and emission matrices for the model by calculating the maximum likelihood, and the MATLAB function hmmviterbi was used to predict the collaborative behaviors. We used the *k*-fold cross-validation method to enhance the training outcome of the HMM to achieve optimal performance. As illustrated in Figure 8, we split 70% of the hand-labeled data as the HMM training set, while the remaining 30% of the hand-labeled data was used as the hold-out test set to evaluate the selected HMM.
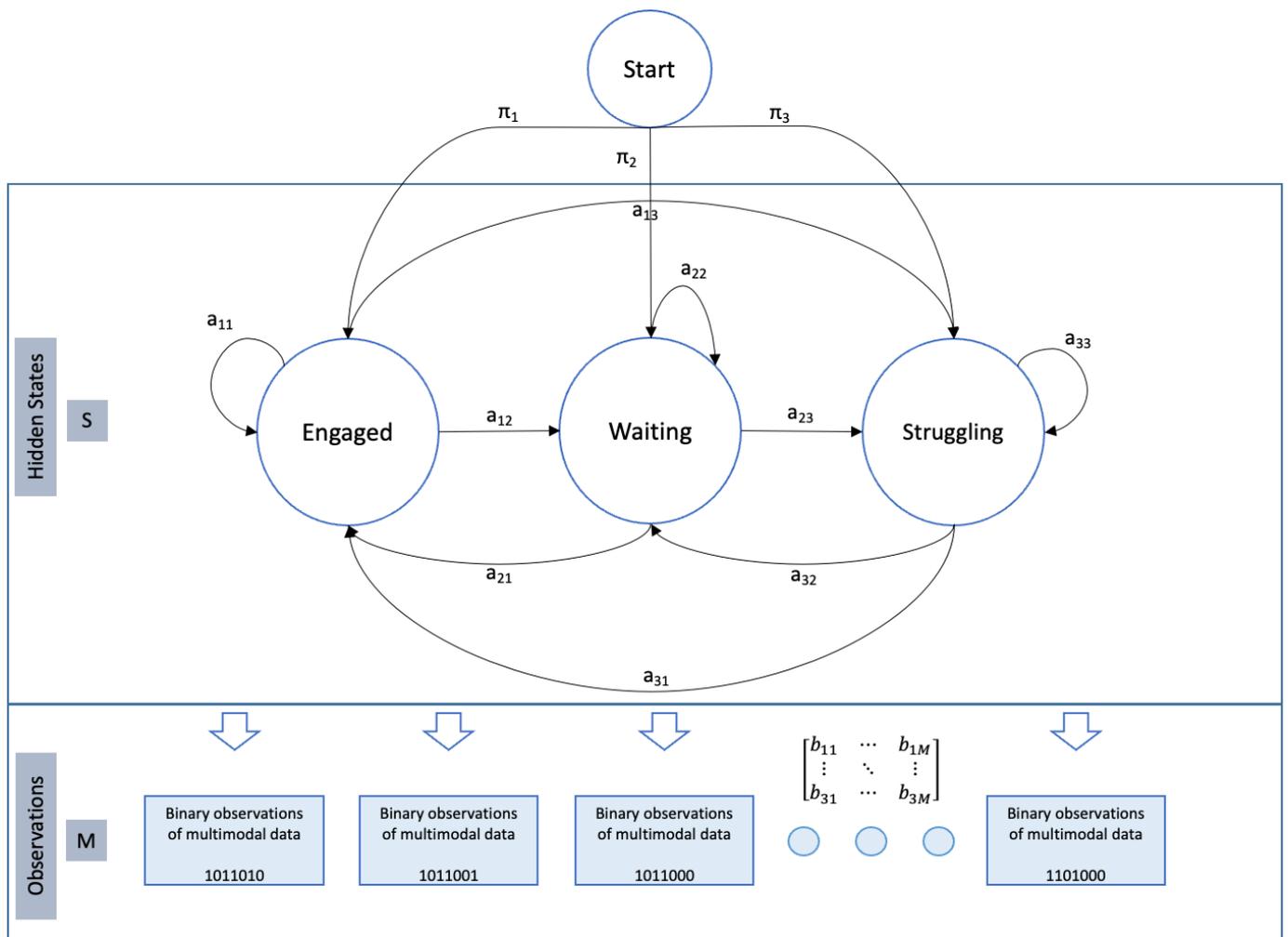
Figure 7. HMM diagram of all elements.

In the *k*-fold cross-validation, due to the imbalance in the labeled collaborative behaviors, we needed to split the data points to address the imbalance by including all possible observations and states in each training instances of the *k*-fold cross-validation. Since the sequence of the observation datapoints is important in generating the transition and emission matrices, we could not use random selection of the datapoints or stratification of the datapoints. As such, we opted to treat the data continuously by splitting the data into certain percentages instead of fixed datapoints for each fold. We found that by splitting the fold into 80% for training and 20% for validation, we were able to generate sets with all observations and behaviors included in each split. The validation set was then shifted upwards in each instance of the *k*-fold until every datapoint from the training set is used for validation in the same sequence without shuffling them into random positions. This is illustrated in the top portion of Figure 8. As part of the splitting of the data, we would label each datapoint to either training or validation to ensure that the same data point would not end up in both training and validation.

The function hmmestimate used the maximum likelihood estimate to generate the state transition and emission matrices based on the binary observation sequences and hand-labeled hidden states. The matrices were then optimized using the function hmmtrain, where a Baum-Welch Algorithm was used to improve the probabilities. With the optimized matrices, we validated the model with the 20% remaining datapoints using the function hmmviterbi. This function used the Viterbi Algorithm to predict the most likely collaborative behavior based on the sequence of observations and probability matrices. The

predicted collaborative behaviors were compared to the actual hand-labeled behavior to find the accuracies of each fold and the average accuracies for all *k* values from 5 through 10.
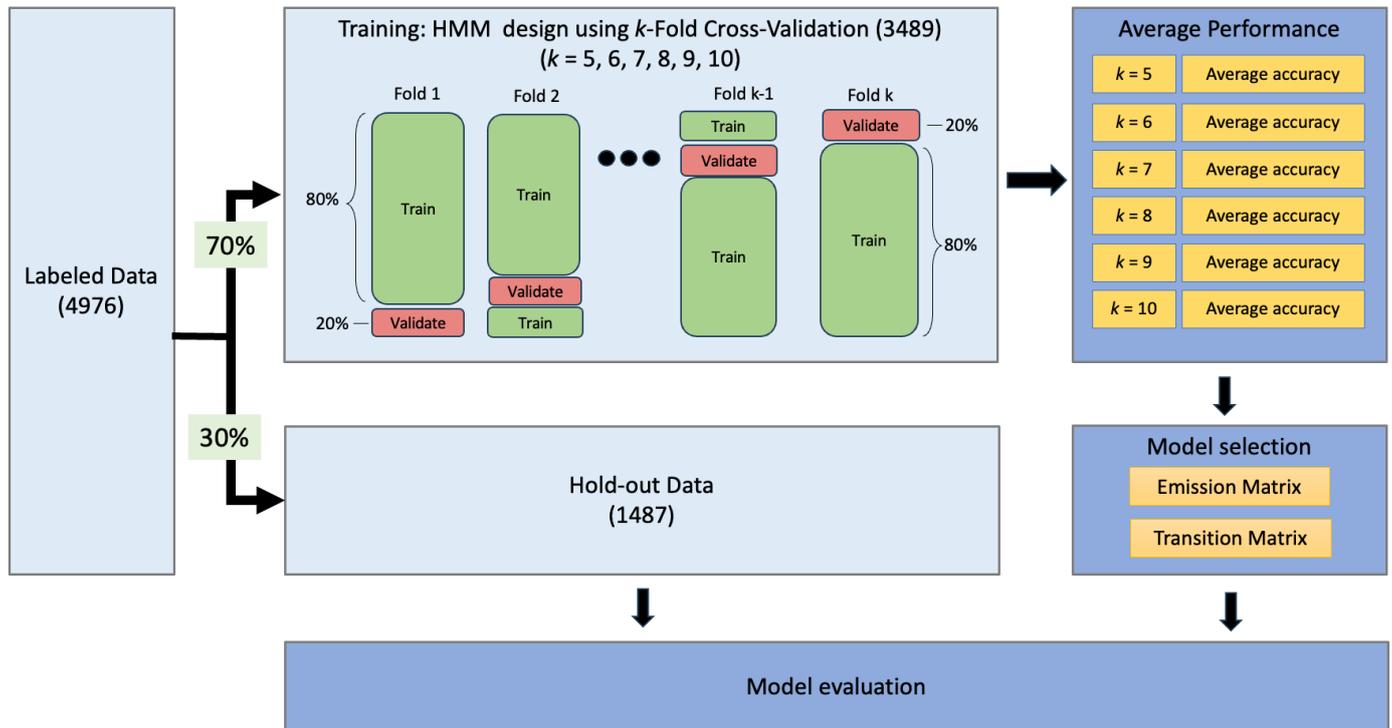

Figure 8. HMM design using *k*-fold cross-validation.

2.2.6. Evaluating Prediction Models Performance

The evaluation of the prediction models was conducted using the remaining 30% of the hand-labeled data that were assigned as hold-out test set. For the rule-based prediction model evaluation, the collaborative behaviors were predicted using the defined rules from Section 2.2.4 As for the HMM prediction model evaluation, we selected the transition and emission matrices from the HMM with the highest accuracy in the *k*-fold cross-validation and used the MATLAB function hmmviterbi to generate the predicted collaborative behaviors using the observations in the hold-out test set. We then compared the predicted collaborative behaviors to the hand-labeled collaboration behaviors. The results are presented in the next section.

3. Results and Discussion

3.1. HMM Training and Validation Results

We trained HMM models using 70% of the hand-labeled data using *k*-fold cross-validation method to optimize the training output. As mentioned in Section 2.2.3, due to the imbalance in the data, we use a 80%-20% split for the training and validation of the HMM, respectively, to avoid missing observations and behaviors when training the HMM.

We present the accuracies for all folds with different *k*-fold values as a boxplot in Figure 9. The boxplot shows the distribution of the accuracy for different *k*-fold values and the scatter plot overlaid to represent the individual accuracy of each fold within the *k*-fold that generated an HMM model. From the plot, we can see that the ranges of the accuracies across all *k*-folds were between 90.35% to 97.77%, and the average accuracies across all folds were around 93%. The highest accuracy occurred when *k* = 9 and Fold = 6. We chose the optimized transition and emission matrices from this fold for evaluation against the hold-out test set.
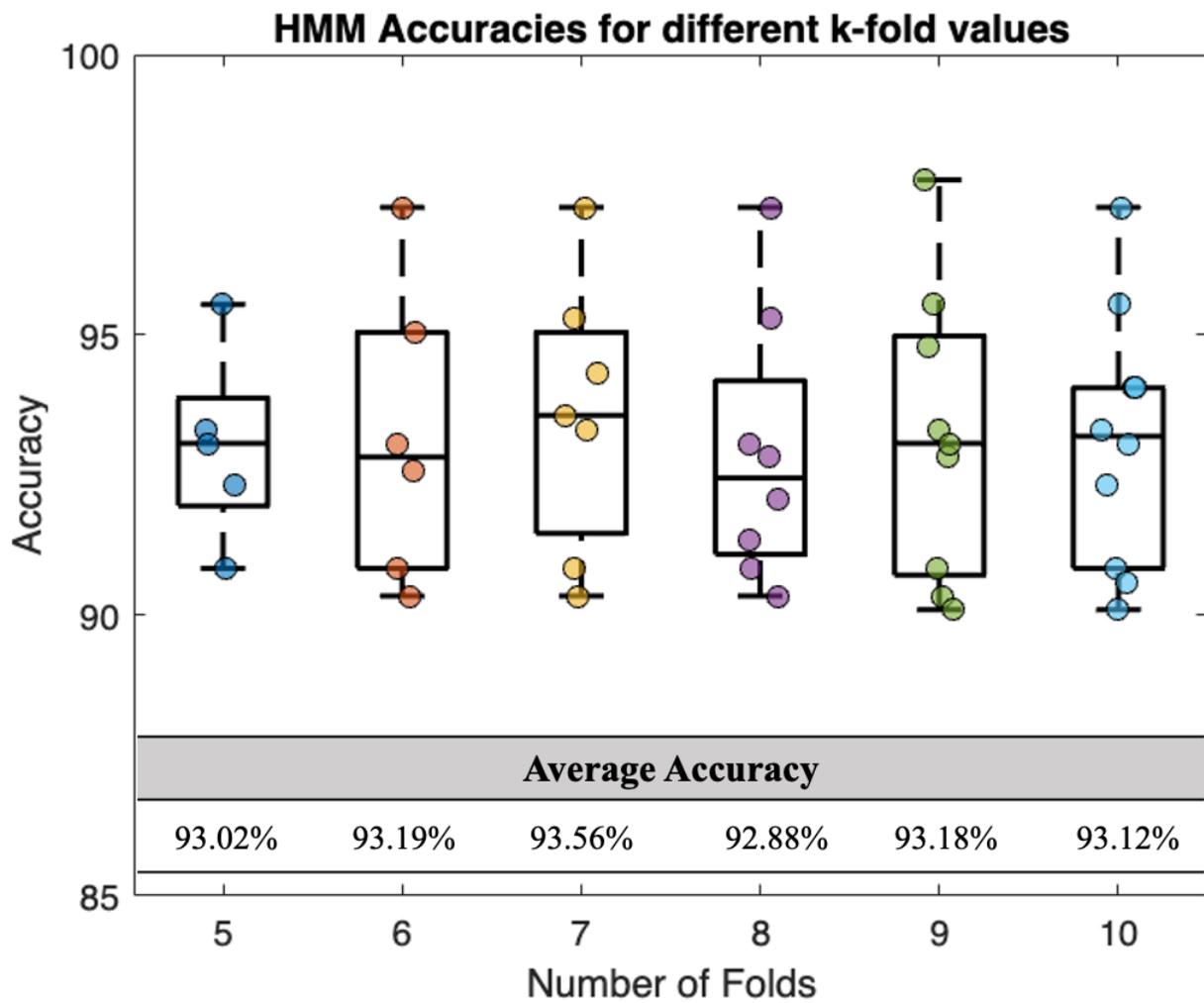
Figure 9. HMM accuracies for different *k*-fold values.

3.2. Prediction Models Evaluation Results

The evaluation was performed for both the rule-based prediction model and the HMM prediction model using the data from the hold-out test set. In the rule-based prediction model, the collaborative behaviors were predicted by evaluating the observations using the defined rules in Section 2.2.4.

As for the HMM prediction model, we chose the HMM that was generated with the highest accuracy of 97.77%. The collaborative behaviors were predicted in MATLAB using hmmviterbi by using the Viterbi Algorithm to generate the most likely collaborative behavior based on the sequence of observations.

In both cases, the predicted collaborative behaviors were compared against the hand-labeled collaborative behaviors. Table 5 compares the performance of rule-based prediction model and HMM prediction model and Figure 10 illustrates the confusion matrix of both prediction models.

Table 5. Evaluation results for rule-based and HMM prediction models.

	Rule-Based (%)	HMM (%)
Accuracy	76.53	90.58
Precision	71.81	89.55
Recall	68.93	87.94

HMM					
	Engaged	**Struggling**	**Waiting**	**Recall**	
Engaged	218	19	0	**91.98%**	8.02%
Struggling	0	231	67	**77.52%**	22.48%
Waiting	2	52	898	**94.33%**	5.67%
Precision	99.09%	76.49%	93.06%	**Accuracy**	
	0.91%%	23.51%	6.94%	**90.59%**	

Rule-based Model					
	Engaged	**Struggling**	**Waiting**	**Recall**	
Engaged	215	11	11	**90.72%**	9.28%
Struggling	2	83	213	27.85%	**72.15%**
Waiting	13	99	840	**88.24%**	11.76%
Precision	93.48%	43.01%	78.95%	**Accuracy**	
	6.52%	56.99%	21.05%	**76.53%**	

(**a**)   (**b**)

Figure 10. Confusion matrix for: (**a**) HMM; (**b**) rule-based model. For both confusion matrices, the column on the left for the collaborative behaviors represent the real labels, while the labels at the top represent the predicted behaviors. The boxes highlighted in yellow indicate the highest number of predictions for each of the behaviors, the boxes highlighted in green represent a positive performance measure of either recall, precision, or accuracy. The boxes highlighted in orange represent a negative performance measure of recall, precision, or accuracy.

Overall, the HMM provided higher accuracy, precision, and recall of the participants' collaborative behavior compared to a rule-based model. When we look at the behaviors as shown in Figure 10, both models performed the best for the *Engaged* behavior since the conditions for the *Engaged* behavior were quite simple and straightforward where both models could provide a reliable prediction. However, for *Waiting* and *Struggling* behaviors, the rule-based model performed quite poorly where the model predicted most of the *Struggling* behavior as *Waiting*. The inflexibility of the rule-based model could have caused this. Rule-based models only allow one behavior for one set of conditions, whereas real hand-labeled data would have instances where the same condition produced different outcomes based on the context of the task (or previous sequence of events). For such cases, if we keep the rule-based model to predict participants' collaborative behavior, the feedback that the participants were to receive would not be true to their actual behavior. A participant that is *Struggling* would not be prompted to seek assistance as the system would assume they are *Waiting* for their partner to complete a turn. On the other hand, the HMM prediction results for *Waiting* and *Struggling* were reliable since the temporal information that was learned from the training was embedded within the state transition and emission probability matrices. This is consistent with the results reported by another study that implemented a semi-supervised model using the same dataset as this study [67]. The study compared the performance of the developed semi-supervised automated labeling of behaviors to supervised and unsupervised models. In this study, a fully supervised support vector machine (SVM) achieved 86.1% accuracy, while a semi-supervised self-training model with 2.5% of the data labelled achieved 84.5% accuracy. Based on this observation, the deterministic nature of HMMs would fit better in a dynamic interaction as it offers more flexibility than a rule-based model and traditional machine learning algorithms, without requiring a large amount of labeled data for training.

4. Conclusions and Future Work

HCI technologies have become integral in skills learning, offering engaging interactions and replicable solutions that enhance learning experiences. These HCI-based systems often incorporate real-time feedback based on user performance to boost user engagement and learning outcomes. By including human behavior in addition to user performance into the feedback mechanism, we can further increase skills learning and engagement. Manual labeling or annotation of data by experts is often used for offline analysis, however evaluating human behavior using this method can be resource-intensive, hindering real-time feedback capability. Thus, the ability to accurately and autonomously recognize

human behavior using computational classification methods is crucial for effective feedback mechanisms.

Among various machine learning methods used to predict human behavior [68], probabilistic models, such as Hidden Markov Models (HMM), have been commonly used to predict or extract behavior patterns [34,36–38]. Leveraging probabilistic classification methods, HMMs analyze temporal patterns, making them suitable for predicting human behavior in collaborative interactions [33]. Our work contributes to this field by developing an HMM-based model tailored to recognize collaborative behaviors in dyadic interactions using multimodal signals. We additionally designed a rule-based method of behavior prediction for a baseline comparison.

Building on our previous work which developed a teamwork training simulator in a CVE [11], which showed acceptability in dyadic interactions between autistic and neurotypical participants, we extend our work by designing prediction models that could be used to recognize collaborative behaviors in dyadic teamwork interactions. In future work, we want to explore the impact of utilizing the predicted collaborative behaviors as an input for a real-time feedback mechanism and how that could improve collaborative interactions in dyadic interactions, specifically cross-neurotype collaboration.

The results of the preliminary study indicated that our HMM prediction model was able to recognize collaborative behaviors with 90.59% accuracy, outperforming the rule-based model. While both models excelled in predicting *Engaged* behavior, the HMM demonstrated greater flexibility, particularly in predicting the *Waiting* and *Struggling* behaviors, due to its ability to leverage temporal information learned during training. This underscores the advantage of the HMM without requiring extensive labeled data for training.

Additionally, our creation of a novel dataset, comprised of multimodal signals from dyadic interactions labeled with collaborative behaviors by expert annotators, opens avenues for further research and experimentation in this domain. By integrating robust behavioral prediction mechanisms into computer-based teamwork training, we anticipate a significant enhancement in collaborative skills development, ultimately advancing the efficacy of virtual training environments.

Although the results are promising, it is important to acknowledge limitations in the HMM design and suggest key improvements for future studies. First, extracting more complex features from the multimodal data would allow researchers to better understand and observe a wider range of human behavior related to collaborative interactions. For example, adding a dialogue act classification [69] for the speech feature would better inform whether the participant said something because they needed help or sharing information indicating different behaviors. Second, the number of human behaviors used to capture participants' collaborative behaviors were limited. Expanding the range of behaviors, particularly for *Waiting*, into more distinguishable behaviors to allow the researchers to better understand what is taking place in the collaboration. Third, the imbalance on the distribution of the collaborative behavior introduced challenges in training the HMM. Using *K*-fold cross-validation was a preliminary method in generating an HMM capable of predicting collaborative behaviors. Future work would benefit from exploring other methods that could include supervised, or semi-supervised learning methods. Despite these limitations, results from the evaluation highlight the advantages of HMMs over rule-based prediction models in a dyadic collaborative interaction between autistic and neurotypical individuals, even with a small labeled dataset. Future work can continue to bridge the gap in effective teamwork training, ensuring a more inclusive and supportive learning experience.

Author Contributions: Conceptualization, A.Z.A., N.S. and D.A.; methodology, A.Z.A., A.P., D.A. and D.M.W.; software, A.Z.A., A.P. and D.A.; validation, N.S. and D.M.W.; formal analysis, A.Z.A. and A.P.; investigation, A.Z.A.; resources, N.S.; data curation, A.Z.A. and A.P.; writing—original draft preparation, A.Z.A.; writing—review and editing, A.Z.A., A.P., D.A. and N.S.; visualization,

References

1. Parsons, S.; Mitchell, P. The potential of virtual reality in social skills training for people with autistic spectrum disorders. *J. Intellect. Disabil. Res.* **2002**, *46*, 430–443. [CrossRef] [PubMed]
2. Slovák, P.; Fitzpatrick, G. Teaching and Developing Social and Emotional Skills with Technology. *ACM Trans. Comput.-Hum. Interact.* **2015**, *22*, 1–34. [CrossRef]
3. Al Mahdi, Z.; Rao Naidu, V.; Kurian, P. Analyzing the Role of Human Computer Interaction Principles for E-Learning Solution Design. In *Smart Technologies and Innovation for a Sustainable Future*; Al-Masri, A., Curran, K., Eds.; Advances in Science, Technology & Innovation; Springer International Publishing: Cham, Switzerland, 2019; pp. 41–44. ISBN 978-3-030-01658-6. [CrossRef]
4. Delavarian, M.; Bokharaeian, B.; Towhidkhah, F.; Gharibzadeh, S. Computer-based working memory training in children with mild intellectual disability. *Early Child. Dev. Care* **2015**, *185*, 66–74. [CrossRef]
5. Fernández-Aranda, F.; Jiménez-Murcia, S.; Santamaría, J.J.; Gunnard, K.; Soto, A.; Kalapanidas, E.; Bults, R.G.A.; Davarakis, C.; Ganchev, T.; Granero, R.; et al Video games as a complementary therapy tool in mental disorders: PlayMancer, a European multicentre study. *J. Ment. Health* **2012**, *21*, 364–374. [CrossRef] [PubMed]
6. Bernardini, S.; Porayska-Pomsta, K.; Smith, T.J. ECHOES: An intelligent serious game for fostering social communication in children with autism. *Inf. Sci.* **2014**, *264*, 41–60. [CrossRef]
7. Zhang, L.; Amat, A.Z.; Zhao, H.; Swanson, A.; Weitlauf, A.; Warren, Z.; Sarkar, N. Design of an Intelligent Agent to Measure Collaboration and Verbal-Communication Skills of Children with Autism Spectrum Disorder in Collaborative Puzzle Games. *IEEE Trans. Learn. Technol.* **2021**, *14*, 338–352. [CrossRef] [PubMed]
8. Zhao, H.; Zaini Amat, A.; Migovich, M.; Swanson, A.; Weitlauf, A.S.; Warren, Z.; Sarkar, N. INC-Hg: An Intelligent Collaborative Haptic-Gripper Virtual Reality System. ACM Transactions on Accessible Computing. Available online: https://dl.acm.org/doi/10.1145/3487606 (accessed on 11 October 2023).
9. Zheng, Z.K.; Sarkar, N.; Swanson, A.; Weitlauf, A.; Warren, Z.; Sarkar, N. CheerBrush: A Novel Interactive Augmented Reality Coaching System for Toothbrushing Skills in Children with Autism Spectrum Disorder. *ACM Trans. Access. Comput.* **2021**, *14*, 1–20. [CrossRef]
10. Amat, A.Z.; Zhao, H.; Swanson, A.; Weitlauf, A.S.; Warren, Z.; Sarkar, N. Design of an Interactive Virtual Reality System, InViRS, for Joint Attention Practice in Autistic Children. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2021**, *29*, 1866–1876. [CrossRef] [PubMed]
11. Amat, A.Z.; Adiani, D.; Tauseef, M.; Breen, M.; Hunt, S.; Swanson, A.R.; Weitlauf, A.S.; Warren, Z.E.; Sarkar, N. Design of a Desktop Virtual Reality-Based Collaborative Activities Simulator (ViRCAS) to Support Teamwork in Workplace Settings for Autistic Adults. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2023**, *31*, 2184–2194. [CrossRef]
12. Awais Hassan, M.; Habiba, U.; Khalid, H.; Shoaib, M.; Arshad, S. An Adaptive Feedback System to Improve Student Performance Based on Collaborative Behavior. *IEEE Access* **2019**, *7*, 107171–107178. [CrossRef]
13. Green, C.S.; Bavelier, D. Learning, attentional control and action video games. *Curr. Biol. CB* **2012**, *22*, R197–R206. [CrossRef] [PubMed]
14. Kotov, A.; Bennett, P.N.; White, R.W.; Dumais, S.T.; Teevan, J. Modeling and analysis of cross-session search tasks. In Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval, Beijing, China, 24–28 July 2011. Available online: https://dl.acm.org/doi/10.1145/2009916.2009922 (accessed on 11 October 2023).
15. Vondrick, C.; Patterson, D.; Ramanan, D. Efficiently Scaling up Crowdsourced Video Annotation: A Set of Best Practices for High Quality, Economical Video Labeling. *Int. J. Comput. Vis.* **2013**, *101*, 184–204. [CrossRef]

16. Hagedorn, J.; Hailpern, J.; Karahalios, K.G. VCode and VData: Illustrating a new framework for supporting the video annotation workflow. In Proceedings of the Working Conference on Advanced Visual Interfaces; AVI '08. Association for Computing Machinery: New York, NY, USA, 2008; pp. 317–321.

17. Gaur, E.; Saxena, V.; Singh, S.K. Video annotation tools: A Review. In Proceedings of the 2018 International Conference on Advances in Computing, Communication Control and Networking (ICACCCN), Greater Noida (UP), India, 12–13 October 2018; pp. 911–914.

18. Fredriksson, T.; Mattos, D.I.; Bosch, J.; Olsson, H.H. Data Labeling: An Empirical Investigation into Industrial Challenges and Mitigation Strategies. In *Proceedings of the Product-Focused Software Process Improvement*; Morisio, M., Torchiano, M., Jedlitschka, A., Eds.; Springer International Publishing: Cham, Switzerland, 2020; pp. 202–216.

19. Vinciarelli, A.; Esposito, A.; André, E.; Bonin, F.; Chetouani, M.; Cohn, J.F.; Cristani, M.; Fuhrmann, F.; Gilmartin, E.; Hammal, Z.; et al. Open Challenges in Modelling, Analysis and Synthesis of Human Behaviour in Human–Human and Human–Machine Interactions. *Cogn. Comput.* **2015**, *7*, 397–413. [CrossRef]

20. Salah, A.A.; Gevers, T.; Sebe, N.; Vinciarelli, A. Challenges of Human Behavior Understanding. In *Human Behavior Understanding*; Salah, A.A., Gevers, T., Sebe, N., Vinciarelli, A., Eds.; Lecture Notes in Computer Science; Springer: Berlin/Heidelberg, Germany, 2010; Volume 6219, pp. 1–12. ISBN 978-3-642-14714-2.

21. Shiyan, A.A.; Nikiforova, L. Model of Human Behavior Classification and Class Identification Method for a Real Person. Supplement. *PsyArXiv SSRN* **2022**. [CrossRef]

22. Dzedzickis, A.; Kaklauskas, A.; Bucinskas, V. Human Emotion Recognition: Review of Sensors and Methods. *Sensers* **2020**, *20*, 592. [CrossRef] [PubMed]

23. Ravichander, A.; Black, A.W. An Empirical Study of Self-Disclosure in Spoken Dialogue Systems. In *Proceedings of the 19th Annual SIGdial Meeting on Discourse and Dialogue*; Association for Computational Linguistics: Melbourne, Australia, 2018; pp. 253–263.

24. Liu, P.; Glas, D.F.; Kanda, T.; Ishiguro, H. Data-Driven HRI: Learning Social Behaviors by Example From Human–Human Interaction. *IEEE Trans. Robot.* **2016**, *32*, 988–1008. [CrossRef]

25. Sturman, O.; Von Ziegler, L.; Schläppi, C.; Akyol, F.; Privitera, M.; Slominski, D.; Grimm, C.; Thieren, L.; Zerbi, V.; Grewe, B.; et al. Deep learning-based behavioral analysis reaches human accuracy and is capable of outperforming commercial solutions. *Neuropsychopharmacology* **2020**, *45*, 1942–1952. [CrossRef] [PubMed]

26. Song, S.; Shen, L.; Valstar, M. Human Behaviour-Based Automatic Depression Analysis Using Hand-Crafted Statistics and Deep Learned Spectral Features. In Proceedings of the 2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018), Xi'an, China, 15–19 May 2018; pp. 158–165.

27. Abdelrahman, A.A.; Strazdas, D.; Khalifa, A.; Hintz, J.; Hempel, T.; Al-Hamadi, A. Multimodal Engagement Prediction in Multiperson Human–Robot Interaction. *IEEE Access* **2022**, *10*, 61980–61991. [CrossRef]

28. D'Mello, S.; Kory, J. Consistent but modest: A meta-analysis on unimodal and multimodal affect detection accuracies from 30 studies. In Proceedings of the 14th ACM International Conference on Multimodal Interaction, Santa Monica, CA, USA, 22–26 October 2012; pp. 31–38.

29. Mallol-Ragolta, A.; Schmitt, M.; Baird, A.; Cummins, N.; Schuller, B. Performance Analysis of Unimodal and Multimodal Models in Valence-Based Empathy Recognition. In Proceedings of the 2019 14th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2019), Lille, France, 14–18 May 2019; pp. 1–5.

30. Okada, S.; Ohtake, Y.; Nakano, Y.I.; Hayashi, Y.; Huang, H.-H.; Takase, Y.; Nitta, K. Estimating communication skills using dialogue acts and nonverbal features in multiple discussion datasets. In Proceedings of the 18th ACM International Conference on Multimodal Interaction, Tokyo Japan, 12–16 November 2016; pp. 169–176.

31. Huang, W.; Lee, G.T.; Zhang, X. Dealing with uncertainty: A systematic approach to addressing value-based ethical dilemmas in behavioral services. *Behav. Interv.* **2023**, *38*, 1–15. [CrossRef]

32. Asghari, P.; Soleimani, E.; Nazerfard, E. Online human activity recognition employing hierarchical hidden Markov models. *J. Ambient. Intell. Hum. Comput.* **2020**, *11*, 1141–1152. [CrossRef]

33. Tang, Y.; Li, Z.; Wang, G.; Hu, X. Modeling learning behaviors and predicting performance in an intelligent tutoring system: A two-layer hidden Markov modeling approach. *Interact. Learn. Environ.* **2023**, *31*, 5495–5507. [CrossRef]

34. Sharma, K.; Papamitsiou, Z.; Olsen, J.K.; Giannakos, M. Predicting learners' effortful behaviour in adaptive assessment using multimodal data. In Proceedings of the Tenth International Conference on Learning Analytics & Knowledge, Frankfurt, Germany, 23–27 March 2020; pp. 480–489.

35. Sánchez, V.G.; Lysaker, O.M.; Skeie, N.-O. Human behaviour modelling for welfare technology using hidden Markov models. *Pattern Recognit. Lett.* **2020**, *137*, 71–79. [CrossRef]

36. Soleymani, M.; Stefanov, K.; Kang, S.-H.; Ondras, J.; Gratch, J. Multimodal Analysis and Estimation of Intimate Self-Disclosure. In Proceedings of the 2019 International Conference on Multimodal Interaction, Suzhou China, 14–18 October 2019; pp. 59–68.

37. Gupta, A.; Garg, D.; Kumar, P. Mining Sequential Learning Trajectories With Hidden Markov Models For Early Prediction of At-Risk Students in E-Learning Environments. *IEEE Trans. Learn. Technol.* **2022**, *15*, 783–797. [CrossRef]

38. Zhao, M.; Eadeh, F.R.; Nguyen, T.-N.; Gupta, P.; Admoni, H.; Gonzalez, C.; Woolley, A.W. Teaching agents to understand teamwork: Evaluating and predicting collective intelligence as a latent variable via Hidden Markov Models. *Comput. Hum. Behav.* **2023**, *139*, 107524. [CrossRef]

39. Mihoub, A.; Bailly, G.; Wolf, C. Social Behavior Modeling Based on Incremental Discrete Hidden Markov Models. In *Human Behavior Understanding*; Salah, A.A., Hung, H., Aran, O., Gunes, H., Eds.; Lecture Notes in Computer Science; Springer International Publishing: Cham, Switzerland, 2013; Volume 8212, pp. 172–183. ISBN 978-3-319-02713-5.

40. Salas, E.; Cooke, N.J.; Rosen, M.A. On Teams, Teamwork, and Team Performance: Discoveries and Developments. *Hum. Factors* **2008**, *50*, 540–547. [CrossRef] [PubMed]

41. McEwan, D.; Ruissen, G.R.; Eys, M.A.; Zumbo, B.D.; Beauchamp, M.R. The Effectiveness of Teamwork Training on Teamwork Behaviors and Team Performance: A Systematic Review and Meta-Analysis of Controlled Interventions. *PLoS ONE* **2017**, *12*, e0169604. [CrossRef] [PubMed]

42. Hanaysha, J.; Tahir, P.R. Examining the Effects of Employee Empowerment, Teamwork, and Employee Training on Job Satisfaction. *Procedia—Soc. Behav. Sci.* **2016**, *219*, 272–282. [CrossRef]

43. Jones, D.R.; Morrison, K.E.; DeBrabander, K.M.; Ackerman, R.A.; Pinkham, A.E.; Sasson, N.J. Greater Social Interest Between Autistic and Non-autistic Conversation Partners Following Autism Acceptance Training for Non-autistic People. *Front. Psychol.* **2021**, *12*, 739147. [CrossRef] [PubMed]

44. Crompton, C.J.; Sharp, M.; Axbey, H.; Fletcher-Watson, S.; Flynn, E.G.; Ropar, D. Neurotype-Matching, but Not Being Autistic, Influences Self and Observer Ratings of Interpersonal Rapport. *Front. Psychol.* **2020**, *11*, 586171. [CrossRef]

45. Milton, D.E.M.; Heasman, B.; Sheppard, E. Double Empathy. In *Encyclopedia of Autism Spectrum Disorders*; Volkmar, F.R., Ed.; Springer: New York, NY, USA, 2018; pp. 1–8. ISBN 978-1-4614-6435-8.

46. Edey, R.; Cook, J.; Brewer, R.; Johnson, M.H.; Bird, G.; Press, C. Interaction takes two: Typical adults exhibit mind-blindness towards those with autism spectrum disorder. *J. Abnorm. Psychol.* **2016**, *125*, 879–885. [CrossRef]

47. Heasman, B.; Gillespie, A. Perspective-taking is two-sided: Misunderstandings between people with Asperger's syndrome and their family members. *Autism* **2018**, *22*, 740–750. [CrossRef] [PubMed]

48. Bozgeyikli, L.; Raij, A.; Katkoori, S.; Alqasemi, R. A Survey on Virtual Reality for Individuals with Autism Spectrum Disorder: Design Considerations. *IEEE Trans. Learn. Technol.* **2018**, *11*, 133–151. [CrossRef]

49. Juliani, A.; Berges, V.-P.; Teng, E.; Cohen, A.; Harper, J.; Elion, C.; Goy, C.; Gao, Y.; Henry, H.; Mattar, M.; et al. Unity: A General Platform for Intelligent Agents. *arXiv* **2018**. [CrossRef]

50. Li, S.; Zhang, B.; Fei, L.; Zhao, S.; Zhou, Y. Learning Sparse and Discriminative Multimodal Feature Codes for Finger Recognition. *IEEE Trans. Multimed.* **2023**, *25*, 805–815. [CrossRef]

51. Khamparia, A.; Gia Nhu, N.; Pandey, B.; Gupta, D.; Rodrigues, J.J.P.C.; Khanna, A.; Tiwari, P. Investigating the Importance of Psychological and Environmental Factors for Improving Learner's Performance Using Hidden Markov Model. *IEEE Access* **2019**, *7*, 21559–21571. [CrossRef]

52. Sanghvi, J.; Castellano, G.; Leite, I.; Pereira, A.; McOwan, P.W.; Paiva, A. Automatic analysis of affective postures and body motion to detect engagement with a game companion. In Proceedings of the 6th International Conference on Human-Robot Interaction, Lausanne, Switzerland, 6–9 March 2011; pp. 305–312.

53. Isohätälä, J.; Järvenoja, H.; Järvelä, S. Socially shared regulation of learning and participation in social interaction in collaborative learning. *Int. J. Educ. Res.* **2017**, *81*, 11–24. [CrossRef]

54. Lobczowski, N.G. Bridging gaps and moving forward: Building a new model for socioemotional formation and regulation. *Educ. Psychol.* **2020**, *55*, 53–68. [CrossRef]

55. Camacho-Morles, J.; Slemp, G.R.; Oades, L.G.; Morrish, L.; Scoular, C. The role of achievement emotions in the collaborative problem-solving performance of adolescents. *Learn. Individ. Differ.* **2019**, *70*, 169–181. [CrossRef]

56. Sarason, I.G.S.; Gregory, R.P.; Barbara, R. (Eds.) *Cognitive Interference: Theories, Methods, and Findings*; Routledge: New York, NY, USA, 2014; ISBN 978-1-315-82744-5.

57. D'Mello, S.; Olney, A.; Person, N. Mining Collaborative Patterns in Tutorial Dialogues. *JEDM* **2010**, *2*, 2–37. [CrossRef]

58. Schmidt, M.; Laffey, J.M.; Schmidt, C.T.; Wang, X.; Stichter, J. Developing methods for understanding social behavior in a 3D virtual learning environment. *Comput. Hum. Behav.* **2012**, *28*, 405–413. [CrossRef]

59. Basden, B.H.; Basden, D.R.; Bryner, S.; Thomas III, R.L. A comparison of group and individual remembering: Does collaboration disrupt retrieval strategies? *J. Exp. Psychol. Learn. Mem. Cogn.* **1997**, *23*, 1176–1189. [CrossRef] [PubMed]

60. Adam, T.; Langrock, R.; Weiß, C.H. Penalized estimation of flexible hidden Markov models for time series of counts. *Metron* **2019**, *77*, 87–104. [CrossRef]

61. Rabiner, L.; Juang, B. An introduction to hidden Markov models. *IEEE ASSP Mag.* **1986**, *3*, 4–16. [CrossRef]

62. Nadas, A. Hidden Markov chains, the forward-backward algorithm, and initial statistics. *IEEE Trans. Acoust. Speech Signal Process.* **1983**, *31*, 504–506. [CrossRef]

63. Tao, C. A generalization of discrete hidden Markov model and of viterbi algorithm. *Pattern Recognit.* **1992**, *25*, 1381–1387. [CrossRef]

64. Baum, L.E.; Petrie, T.; Soules, G.; Weiss, N. A Maximization Technique Occurring in the Statistical Analysis of Probabilistic Functions of Markov Chains. *Ann. Math. Stat.* **1970**, *41*, 164–171. [CrossRef]

65. Schreiber, R. MATLAB. *Scholarpedia* **2007**, *2*, 2929. [CrossRef]

66. Matlab, R. *MATLAB Statistics and Machine Learning Toolbox*, Version R2017a 2017; The MathWorks Inc.: Natick, MA, USA.

67. Plunk, A.; Amat, A.Z.; Tauseef, M.; Peters, R.A.; Sarkar, N. Semi-Supervised Behavior Labeling Using Multimodal Data during Virtual Teamwork-Based Collaborative Activities. *Sensors* **2023**, *23*, 3524. [CrossRef] [PubMed]

68. Hiatt, L.M.; Narber, C.; Bekele, E.; Khemlani, S.S.; Trafton, J.G. Human modeling for human–robot collaboration. *Int. J. Robot. Res.* **2017**, *36*, 580–596. [CrossRef]

69. Stolcke, A.; Ries, K.; Coccaro, N.; Shriberg, E.; Bates, R.; Jurafsky, D.; Taylor, P.; Martin, R.; Ess-Dykema, C.V.; Meteer, M. Dialogue Act Modeling for Automatic Tagging and Recognition of Conversational Speech. *Comput. Linguist.* **2000**, *26*, 339–373. [CrossRef]