2024 33rd IEEE International Conference on
Robot and Human Interactive Communication (RO-MAN)
August 26-30, 2024. Pasadena, California, USA.

2024 33rd IEEE International Conference on Robot and Human Interactive Communication (ROMAN) | 979-8-3503-7502-2/24/$31.00 ©2024 IEEE | DOI: 10.1109/RO-MAN60168.2024.10731360

# Identifying Worker Motion Through a Manufacturing Plant: A Finite Automaton Model

Shaoze Yang†, Shreyas Bhat†, Yutong Ren,
Paul Pridham, and X. Jessie Yang*
†co-first authors *corresponding author
Department of Industrial and Operations Engineering
The University of Michigan, Ann Arbor, MI 48109
Email: {shaozey, shreyasb, rentony, pridham, xijyang}@umich.edu

Terra Stroup and Al Salour
The Boeing Company, Arlington, VA, 22202
Email: {terra.a.stroup, al.salour}@boeing.com

***Abstract*— Autonomous Guided Vehicles (AGVs) are becoming increasingly common in industrial environments to transport heavy equipment around warehouses. Within the idea of Industry 5.0, these AGVs are expected to work alongside humans in the same shared workspace. To enable smooth and trustworthy interaction between workers and AGVs, the AGVs must be able to model the workers' behavior and plan their trajectories around it. In this paper, we introduce a Finite Automaton Model (FAM) to model worker motion in such a context. We conduct a human subject experiment using a Virtual Reality (VR) environment and an omnidirectional treadmill to collect data about worker trajectories to tune our model. We show that not only is our model more interpretable, but also outperforms machine learning models at classifying worker motion behavior with limited training data. Future research can use our model to modify AGV behavior to promote trustworthy human-AGV interaction.**

## I. Introduction

Transitioning from Industry 4.0 to Industry 5.0 signifies a pivotal shift towards a more human-centric approach in industrial and manufacturing settings. Industry 4.0 introduced us to major technologies such as Smart Factories, Autonomous Guided Vehicles (AGVs), the Internet of Things (IoT), highlighting the integration of advanced automation and data exchange in manufacturing technologies [1]. As we transition into Industry 5.0, the focus extends beyond automation to emphasize the synergy between humans and machines, fostering an environment where humans and machines collaborate seamlessly [2].

To facilitate smooth and trustworthy interaction between Autonomous Guided Vehicles (AGVs) and workers within manufacturing settings, we envision a future where autonomous and robotic agents can predict and adapt to workers' behaviors [3], [4]. In this paper, we tackle one part of the human-AGV interaction problem: How should the AGV model human motion in manufacturing environments? While existing research on Pedestrian-Autonomous Vehicle (AV) interactions modeled pedestrian behaviors at crosswalks [5], factory settings present unique challenges due to their more chaotic, unstructured nature absent of clear pedestrian pathways. In addition, previous studies (e.g., [6], [7]) on Worker-AGV interactions have not fully leveraged the structured patterns of human movement within these environments, where workers frequently navigate between various workstations. Our model seeks to bridge these gaps by offering a better understanding of human motion in factory settings.

We develop a Finite Automaton Model (FAM) with an intuitive state space to model worker motion in the presence of AGVs. To evaluate the model, we conduct a human-subject study with 19 participants in a Virtual Reality (VR) environment that mirrors a typical manufacturing environment, where AGVs are responsible for transporting goods, and workers move toolboxes between stations.

Results show that our model outperforms other approaches including neural networks, support vector machines, and other classification models in the presence of limited training data. Our model also includes an error feedback loop that allows it to self-correct itself in case of unexpected behavior from the worker. Future studies could use our model to perform adaptive behavior modifications for the AGVs to enable trustworthy interaction between workers and AGVs.

The rest of the paper is organized as follows: Section II discusses literature in pedestrian-AV interaction and worker-AGV interaction that motivates our work. Section III provides details of our Finite Automaton Model including the states, features, the error feedback loop, and the constraints. Section IV details the human subject study conducted. Section V discusses major results from the study. Finally, Section VI concludes the paper and discusses limitations and possible directions for future work.

## II. Related Work

Our work is related to two major areas of research in human-robot interaction, namely, pedestrian-AV interaction and worker-AGV interaction. In this section we discuss literature that motivates our approach.

### A. Pedestrian-AV interaction

Research in pedestrian-AV interaction can be broadly classified into 3 categories: 1) physics-based approaches [8], [9], 2) pattern-based approaches [10], [11], and 3) planning-based approaches [12], [13] (see [14] for a detailed review). Physics-based approaches simulate the human's trajectory

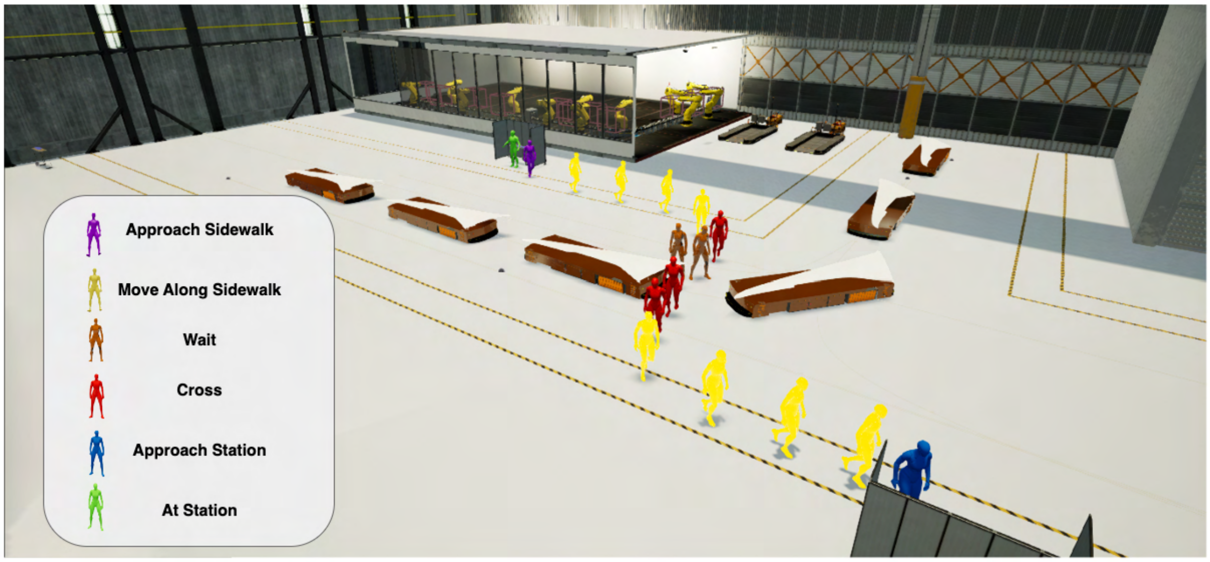


Fig. 1: Classification of walking behavior by our model in a virtual environment. States are differentiated by colors.

based on some known dynamical system of equations. Pattern-based approaches employ various functions (e.g. Gaussian Process, Neural Networks, etc.) to learn patterns from training data and predict human motion in the future. Planning-based approaches take into account human intent to form hypotheses on where the human would move next.

Under the planning-based approach, Jayaraman et al. [5], [15], [16] developed a finite automaton model to model the behavior of pedestrians near uncontrolled mid-block crosswalks and intersections. Yannis et al. [17] and Kadali et al. [18] developed gap acceptance models for predicting whether or not a pedestrian would cross at a well designated but uncontrolled crosswalk. The reasoning behind such models is that the pedestrians will only cross if the lateral distance between the crosswalk and the AGV is big enough. In our case, however, such models would not directly apply since the workers can cross the "road" on which the AGV drives at any point in their trajectory.

Our work uses a similar approach, but for modeling worker motion in manufacturing plants. The absence of well defined crosswalks within the manufacturing plant make this problem more unstructured in our case.

### B. Worker-AGV interaction

Research in worker-AGV interaction typically deals with modeling and designing safe interaction between workers and AGVs on a shop floor. Tubis et al. [19] provide a recent review of literature of this field. They categorize prior works into 5 different categories: 1) Review articles [20], [21], 2) Comparison of AGV and human work [22], [23], 3) Human-AGV cooperation [6], [7], 4) Designing a safe work environment [24], [25], and 5) Others. Within this categorization, our work falls into Human-AGV cooperation.

When humans and AGVs share the same workspace, it becomes essential to provide adequate safety for the human. The current industry standard is to use laser scanners mounted on the AGVs to detect whether an obstacle enters a bounding box around the AGV [24]. If an object is detected within this bounding box, the AGV's trajectory is modified to either slow down or stop depending on the distance of the object from the AGV. The effect of having an adaptive sizing of this bounding box is analyzed in Muhammad et al. [24]. Their results indicate that having an adaptive sizing of the bounding boxes results in a lower number of slowdowns for the AGV, thus increasing their efficiency.

Bergman et al. [7] provide design recommendations for AGV behavior to result in positive affect among the workers sharing the same workspace. They take inspiration from human social norms and analogies from nature to design two behaviors for AGVs when approaching a worker in a straight line. Their results indicate that the AGV following a curved path rather than coming to a stop results in more legible behavior by humans. However, there is no significant difference between the two behaviors in terms of predictability of the AGV behavior by humans.

Löcklin et al. [6], [26] consider the problem of trajectory prediction of moving workers for AGVs on the shop floor. They compare two methods for trajectory prediction of workers, namely, pattern based prediction via 2D Convolutional Neural Networks (CNN) and planning-based prediction using a pathfinding algorithm in conjunction with a semantically-extended map of the environment. Their results indicate that pattern-based trajectory prediction is faster but less reliable for long-horizon predictions. Planning-based trajectory prediction is slower, but more accurate for long-horizon predictions [26].

Prior works, however, have not leveraged the underlying structure of worker behavior when they are moving around the plant. Our work utilizes this structure to define intuitive states in a FAM to model worker behavior as they move between different work stations in the presence of AGVs sharing the same space.

## III. Finite Automaton Model

Workers in a manufacturing plant typically move between various stations. At the same time, AGVs transport heavy equipment from one location to another in the plant. Thus, the workers have to plan their trajectories around the trajectories of the AGVs. To allow for smooth and trustworthy interaction between the workers and the vehicles, it is essential for the vehicle to be able to estimate the worker's behavior and plan its own trajectory accordingly.

In this paper, we propose a Finite Automaton Model (FAM) to predict worker's behavior, as shown in Algorithm 1. Here we formally define $S, X, \mathscr{F}, \mathscr{C}, \mathscr{M}$, where:

1) $S$ is the set of states, as detailed in III-A. $S_{pred}$ denotes model's prediction.
2) $X := [F_1, F_2, ...F_q]$ is the feature and is detailed in III-B.
3) $\mathscr{F} : S \times X \rightarrow S$ is the state transition function, as detailed in III-A.
4) $\mathscr{C} : S \times X \rightarrow Boolean$ is the constraint evaluation function, as detailed in III-D.
5) $\mathscr{M}$ is the probability transition function. $\mathscr{M}_{s'}(s)$ returns the estimated probability from state $s'$ to state $s$, as detailed in III-D.

**Algorithm 1** Finite Automaton Model

**Initialize:** $S_{pred} \leftarrow$ "Error State", $X$
**Initialize:** *errorStack* with capacity limit
**while** system is operational **do**
  **if** $S_{pred} \neq$ "Error State" **then**
    *errorStack*.Update($\mathscr{C}(S_{pred}, X)$) ▷ Update replaces the oldest element in stack with new one.
    **if** *errorStack*.IsAllFalse() **then**
      **Reset** $\mathscr{M}_{S'}$ with $S' \leftarrow S_{pred}$
      $S_{pred} \leftarrow$ "Error State"
      *errorStack*.Clear()
    **else**
      $S_{pred} \leftarrow \mathscr{F}(S_{pred}, X)$
    **end if**
  **else**
    $Q \leftarrow \{s \in S : \mathscr{C}(s, X)\}$ ▷ Q is all states which satisfy the constraints
    **if** $Q \neq \varnothing$ **then**
      $S_{pred} \leftarrow \underset{s \in Q}{\operatorname{argmax}} \mathscr{M}_{S'}(s)$
      *errorStack*.Clear()
    **end if**
  **end if**
  **Update** $X$
**end while**

### A. States

We chose 6 states to represent typical worker trajectories while moving between stations: 1) At station, 2) Approach sidewalk, 3) Wait, 4) Move along sidewalk, 5) Approach Station and 6) Cross. Fig. 1 shows an ideal classification of worker's behavior from our model. We also defined transitions between these states grounded in the human-subjects data, as shown in Fig. 2. To avoid cluttering in the figure, we did not show transitions from states to themselves. However, it should be noted that each state can make a transition to itself. State transition criteria can be seen in Table I. In this table, whether the worker is moving or is stationary is determined by checking their speed against a threshold (0.3 $m/s$), whether the worker is looking at something is determined by whether their gaze direction is within a 40° cone of that object.

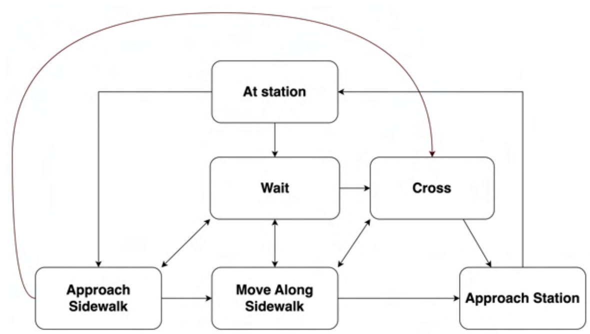


Fig. 2: Normal Operational Loop of Finite Automaton Model

### B. Features

The features utilized, denoted as $X$, are detailed in Table II. Data were collected directly from the virtual environment (see Sec. IV), including the positions of the AGV and the worker, along with the 3-D orientation of the worker's head. It is presumed that the model possesses comprehensive knowledge regarding the locations of stations and sidewalks, as well as the objects the worker is looking at.

### C. Error Feedback Loop

A significant limitation of the Finite Automaton Model (FAM) design is its vulnerability to prediction errors that can cause the system to become stuck. For instance, should the system erroneously identify a worker's state as being in the "Approach Station State" when they are actually in the "Cross State", the FAM is constrained to follow a predefined sequence of states—Approach Station, At Station, and then Cross—before it can align with the correct state. This rigid sequence increases the likelihood of the system remaining in an incorrect state. To address this issue, we have implemented an error feedback loop mechanism to enhance the system's ability to correct itself and return to its normal operational loop. As shown in Fig. 3, an additional "Error State" has been integrated alongside the original FAM model.

This setup involves a continuous monitoring process, evaluates if worker's status satisfy the constraint of the currently predicted state and maintains a record of the most recent results in a first-in-first-out *error stack*, which stores a series of Boolean values. Should this error stack contain exclusively error values, indicating persistent prediction inaccuracies, the system transitions into the "Error State." It remains in

TABLE I: State Transition Conditions and Empirically Observed Probabilities.
Note: The transition probabilities do not sum to 1 as we have not included the self-transition probabilities

| From state | To state | Condition | Probability |
|---|---|---|---|
| At Station | Approach Sidewalk | Worker is moving and is either near the sidewalk or facing the sidewalk | 0.024 |
| At Station | Wait | Worker is stationary and looking at the AGV | 0.022 |
| At Station | Move Along Sidewalk | **(Observed but not in FAM)** | 0.005 |
| Approach Sidewalk | Wait | Worker is stationary and looking at the AGV | 0.087 |
| Approach Sidewalk | Cross | Worker is moving on the road | 0.262 |
| Approach Sidewalk | Move Along Sidewalk | Worker is moving along the sidewalk | 0.087 |
| Wait | Cross | Worker is moving on the road and is facing the road | 0.041 |
| Wait | Move Along Sidewalk | Worker is moving along the sidewalk | 0.0075 |
| Wait | Approach Sidewalk | Worker is moving and is facing towards the sidewalk | 0.127 |
| Cross | Approach Station | Worker is moving and looking at the closest station | 0.131 |
| Cross | Wait | **(Observed but not in FAM)** | 0.015 |
| Cross | Move Along Sidewalk | Worker is moving along the sidewalk | 0.003 |
| Move Along Sidewalk | Wait | Worker is stationary and looking at the AGV | 0.0075 |
| Move Along Sidewalk | Approach Station | Worker is looking at the closest station and is near that station | 0.065 |
| Move Along Sidewalk | Cross | Worker is moving on the road | 0.003 |
| Move Along Sidewalk | Approach Sidewalk | **(Observed but not in FAM)** | 0.005 |
| Approach Station | At Station | Worker is stationary near a station and facing away from the road | 0.253 |
| Approach Station | Move Along Sidewalk | **(Observed but not in FAM)** | 0.0046 |

TABLE II: Input Features for Finite Automaton Model

| Parameter | Description |
|---|---|
| AGV_pos, W_pos | Position of AGV and worker |
| GazeDirection | 3-D direction of worker's gaze |
| AGV_speed, W_speed | Speed of AGV and worker |
| OnRoad / OnSidewalks | Boolean classifies worker's position |
| FacingRoad (alongSW) | Boolean classifies worker's gaze direction |
| ClosestStationDis | Distance to the closest station near worker |
| FacingClosestStation | If worker is looking at the closest station |
| GazeRatio | Worker's attention on AGV in previous second |
| IntentToCross | Boolean based on worker's acceleration and Gaze ratio |
| PossibleInteraction | Possible Interaction in 10s if velocity unchanged |

this state until some constraint in the set of all constraints in the model is satisfied, at which point the error stack is cleared, and the system resumes its normal operational loop in the state corresponding to the constraint that was satisfied. Consider the arbitrariness of worker's behaviors, the constraints are designed to be slack to avoid frequently entering the "Error State". At the initial detection of a worker, we set the model to the "Error State" to avoid restriction from any predefined transition sequence. The effectiveness validation and parameter selection, Section IV-D.2, further validates the effectiveness of this loop.

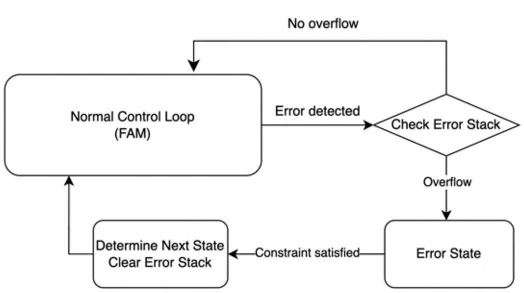


Fig. 3: Error Feedback Loop

### D. Constraints

The constraints $\mathscr{C}$ are series of checker functions that evaluate if current worker's status against the model's predictions. The constraints used in our model are not necessarily mutually exclusive. The constraints can, therefore, be more relaxed, making the model more robust to noise in the recorded data. This, however, means a worker's status could simultaneously satisfy the criteria for more than one state. Consequently, when the model is in the "Error State" and attempts to return to the standard operational loop, the sequence in which the constraints are checked can influence the determination of the next state, introducing a degree of uncertainty. To mitigate the impact of this uncertainty, we have incorporated an N-Grams model [27] to forecast the subsequent state based on the state preceding the transition into the "Error State". Our model has $m$ unique states $S_1, ...S_m$ and $S$ denotes the set of states Given a labeled sequence $L_1, ..., L_N$ (denoted by $L_{1:N}$), where $L_i \in S$ for $i \leq N$. Adapting the Markov Assumption for simplification, the probability of sequence can be expressed as:

$$P(L_1 L_2 ... L_N) = P(L_1) \prod_{k=2}^{N} P(L_k | L_{1:k-1}) \approx P(L_1) \prod_{k=2}^{N} P(L_k | L_{k-1}) \tag{1}$$

subject to the constraint

$$\sum_{i=1}^{m} P[(L_N = S_i) | L_{N-1}] = 1, \ \forall L_{N-1} \in S \tag{2}$$

We can then estimate the bi-gram probabilities $P(S_i|S_j)$ using maximum likelihood estimation

$$P(S_i|S_j) = \frac{C_L(S_j S_i)}{\sum_{k=1}^{m} C_L(S_j S_k)}, \ \forall i, j \in \{1, \cdots, m\} \tag{3}$$

where $C_L(S_j S_i)$ is the count of sub-sequence $S_j S_i$ in the labeled sequence. $\sum_{k=1}^{n} C_L(S_j S_k)$ is the count of all subsequences that start with $S_j$.

Therefore, to return from the error state, if the worker's status satisfies the constraints of both $S_i$ and $S_j$, we can compare between $P[S_i|S']$ and $P[S_j|S']$ where $S'$ is the last state before the "Error State", to determine the next state. The estimated transition probabilities from out labeled dataset are detailed in Table I. To address the issue of not having any prior state at the first timestep, we check the constraints in a pre-determined order based on the descending level of danger of that state, presented below:

$$D_{\text{Cross}} > D_{\text{Wait}} > D_{\text{Move Along Sidewalk}} > D_{\text{Approach Sidewalk}} > D_{\text{At Station}}$$

## IV. Experiments

### A. Testbed

We modeled a VR manufacturing plant environment in Unreal Engine 5.1.1. Participants used the Meta Quest Pro headset with the controllers to interact with the testbed. Additionally, the participants could walk around in the environment using the KAT Walk C2 omnidirectional treadmill. A screenshot of the AGV within the VR environment and the equipment can be seen in Fig. 4

(a) Apparatus (b) AGV

Fig. 4: a) A researcher in the omnidirectional treadmill wearing the headset and controllers. b) A screenshot of the AGV in the VR environment.

There are multiple workstations in the testbed between which the participants could move. At the same time, an AGV carried large components around the plant. The testbed records the participant's location, their gaze vector, and the AGV's location at 72 Hz. For our analysis, we downsampled by averaging the data over one second.

### B. Human Subject Study

We designed two conditions for the AGVs: slowing down when near the workers and not slowing down when near the workers. All AGVs would stop when a person was within a prescribed distance. Although not the focus of this study, the two conditions represent distinct vehicle behaviors with which workers could interact, and thus increase the generalizability of our results. We designed a within-subjects study; each participant completed 16 toolbox delivery tasks for each AGV behavior. We collected data from 19 participants (10 Male, 9 Female, Age = $23.3 \pm 3.6$ years), adding up to 24252 seconds. All participants were students from the University of Michigan. This study was regulated by the Institutional Review Board at the University of Michigan, study ID HUM00233665.

The study procedure was as follows: The participants signed an informed consent form. Then, the experimenter helped the participants into the equipment (wearing the footwear for the treadmill, getting on to the treadmill, and wearing the headset and controllers). After that the participants completed a training session that helped familiarize them with walking on the treadmill, viewing the environment through the headset, and using the controllers to interact with objects in the environment. Then, they performed two trials, one with the slowdown condition of the AGV and another with the no-slowdown condition of the AGV. The order of presentation of the condition was randomized to counter any learning effects. In the slowdown condition, the AGV reduced its speed in proximity to the participants. In the no-slowdown condition, the AGV did not slow down even when near the participants. In both conditions, however, the AGV came to a complete stop if the participant got too close (to prevent a collision). Each trial consisted of taking a toolbox from one station to another while remembering a 3-digit code representing the target location of the toolbox at the next station. Each participant completed 16 trials for each condition, taking approximately 25 minutes in total. The AGV paths in the 16 trials were designed to have variations of all 6 types of interactions [28] between workers and AGVs. After each trial, the participants were asked to answer questions to gauge their level of trust and comfort while sharing the workspace with the AGV. Each question was rated on a discrete scale from 1 to 10. Analysis of these surveys is not within the scope of this paper, which focuses on identifying worker motion.

### C. Inter-Rater Reliability

For the dataset, participant behavior was manually classified into states by three independent raters. We curated a dataset comprising 9 anticipated and 9 unanticipated instances of worker behavior, totaling 834 seconds, to calibrate the model's parameters and estimate the state transition probabilities (see Table I). Subsequently, we evaluated the model's performance on a test set encompassing 87 cases, totaling 3362 seconds.

Following Hallgren's methodology [29], we employed Light's Kappa to compute the average kappa coefficient across all pairs of raters, resulting in a value of 0.914. In general, a value of above 0.8 is considered to be indicative of good agreement between the raters. The pairwise Kappa Scores are detailed in Table III.

TABLE III: Inter-Rater Reliability For Rater A, B and C

| Rater 1 | Rater 2 | Kappa Score |
|---|---|---|
| A | B | 0.919 |
| A | C | 0.917 |
| B | C | 0.907 |

### D. Effectiveness Validation and Parameter Selection

To assess our method's efficacy, we performed several experiments, evaluating feature importance, as well as the effectiveness of the error feedback loop and probability estimation.

*1) Feature Importance:* In this analysis, we sequentially omitted each feature (set all values to 0), excluding raw features listed in Table II, to evaluate their impact on the FAM's performance. The significance of each parameter, based on its influence on the FAM's effectiveness, is summarized in Table IV. The features are ranked in ascending order of their importance, with the IntentToCross feature exerting the least influence on model performance.

TABLE IV: FAM Feature Importance Ranking

| Feature Omitted | Accuracy | Precision | Recall | F1 Score |
|---|---|---|---|---|
| N.A. (Total Model) | 0.8068 | 0.7030 | 0.6994 | 0.6950 |
| IntentToCross | 0.7856 | 0.6885 | 0.6784 | 0.6735 |
| PossibleInteraction | 0.7715 | 0.6915 | 0.6649 | 0.6609 |
| FacingRoad(alongSW) | 0.7668 | 0.7068 | 0.6469 | 0.6683 |
| GazeRatio | 0.7668 | 0.7025 | 0.6720 | 0.6784 |
| ClosestStationDis | 0.7067 | 0.5494 | 0.5606 | 0.5348 |
| FacingClosestStation | 0.6737 | 0.5806 | 0.5920 | 0.5654 |
| OnRoad/Sidewalk | 0.4994 | 0.4460 | 0.4543 | 0.4427 |
| AGV-User speed | 0.2497 | 0.0981 | 0.1790 | 0.1088 |

*2) Error Feedback Loop:* To assess the impact of the error feedback loop on our model's performance, we conducted a series of experiments where the model was tested with varying sizes of the error stack, as shown in Fig. 5. Model accuracy in total dataset was improved from **0.574** to **0.655** (without probability estimation referred in Sec. III-D), achieving peak performance with an error stack size of 3. For the rest of experiments, we kept this size fixed.

To statistically verify the effectiveness, we conducted a 10-fold cross validation test on the training data. Since we cannot guarantee the data is normally distributed, we applied the Wilcoxon Signed-Rank Test and the p-value was found to be 0.0019, meaning the improvement was significant.

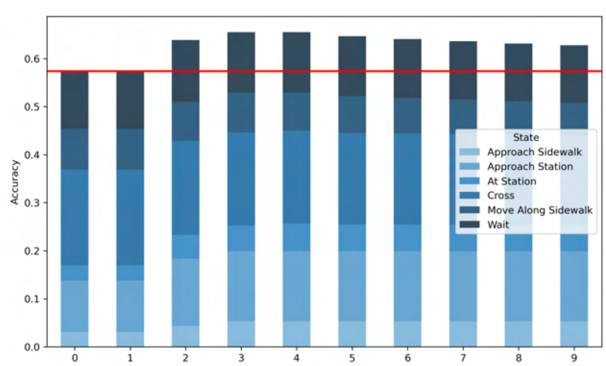


Fig. 5: Comparative accuracy of various models with different error stack sizes. The red line indicates the benchmark accuracy achieved by a model operating without error feedback loop.

*3) Probability Estimation:* In alignment with the methodology outlined in Section III-D, we incorporated a state transition probability estimation matrix to address uncertainties. Experimental results revealed an approximate **15%** increase in the model's overall accuracy. Similar to Sec. IV-D.2, we applied the Wilcoxon Signed-Rank Test on an 8-fold cross validation with 30% data in each sample, resulting in a p-value of 0.0078.

*4) Classification Methods Comparison:* To validate the effectiveness of our methodology, we compared our model's results with those from established classifiers available in scikit-learn [30], including Logistic Regression, Decision Tree, LightGBM, and Support Vector Machine (SVM). All these models were trained and tested with the same setup as mentioned in Sec. IV-C.

## V. Results

Our method of participant behavior classification has displayed superior performance. We reach the overall accuracy of **0.807** and the accuracy of different states are shown in the Table V. The confusion matrix further illustrating our model's predictive accuracy is presented in Fig. 6.

TABLE V: Model Performance of Different States

| State | Accuracy | Precision | Recall | F1 Score |
|---|---|---|---|---|
| Move Along Sidewalk | 0.9722 | 0.3333 | 0.9722 | 0.9859 |
| Cross | 0.9016 | 0.2000 | 0.9016 | 0.9483 |
| Approach Target Station | 0.8986 | 0.3333 | 0.8986 | 0.9466 |
| Approach Sidewalk | 0.7917 | 0.2000 | 0.7917 | 0.8837 |
| At Station | 0.7793 | 0.2500 | 0.7793 | 0.8759 |
| Wait | 0.7442 | 0.2000 | 0.7442 | 0.8533 |

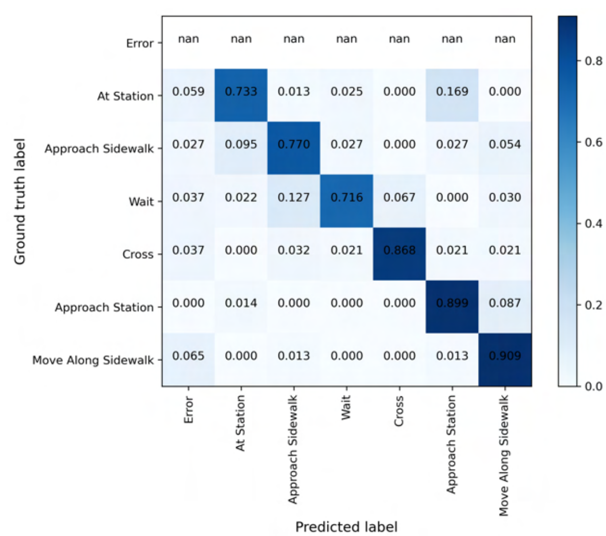


Fig. 6: Normalized confusion matrix of model's prediction. Dark grid in the diagonal indicates high accuracy across all classes. The off-diagonal lighter elements indicate fewer misclassifications between classes.

We compared our model with other classification models trained on the same extracted features and labeled data as described in IV-C. Our model outperformed in overall accuracy, Recall macro, and F1-score, as indicated in Table VI. We also compare model accuracy within each state, as shown in Fig. 7. While predictions from other models exhibited high variance across different states, our FAM model maintained relatively consistent performance. This consistency was particularly notable in more critical states such as "Cross State" and "Wait State", where our model alone provided stable and reliable predictions. This underscores the robustness and reliability of our approach in effectively identifying and classifying worker behaviors across a spectrum of scenarios, especially in critical safety-related contexts.

TABLE VI: Model Comparison

| Model | Accuracy | Precision | Recall | F1 Score |
|---|---|---|---|---|
| Neural Network | 0.2362 | 0.2653 | 0.2105 | 0.1915 |
| SVM | 0.3303 | 0.1486 | 0.2348 | 0.1780 |
| KNN | 0.3440 | 0.3368 | 0.3497 | 0.3396 |
| Logistic Regression | 0.4656 | 0.3917 | 0.4105 | 0.3825 |
| Decision Tree | 0.5872 | 0.5518 | 0.5982 | 0.5600 |
| LGBM | 0.6583 | 0.6062 | 0.5783 | 0.5634 |
| **Ours** | 0.8068 | 0.7029 | 0.6994 | 0.6950 |

## VI. CONCLUSION

In this paper, we presented a Finite Automaton Model (FAM) designed to model worker walking behavior in manufacturing environments where AGVs and humans share workspaces. The FAM not only offers an intuitive interpretation but also outperforms other classification techniques. The effectiveness validation and parameter selection evaluations corroborated the significance of feature engineering, the incorporation of an error feedback loop, and the state transition probability estimation in enhancing model performance. As a result, the model exhibits good accuracy in identifying critical and hazardous states, such as "Cross" and "Wait," prioritizing safety.

Although our proposed method achieves satisfying results in predicting worker behaviors, there are still limitations that need to be addressed. First, our model lacks predictive power. It only classifies worker behavior into different states depending on their location, velocity, head pose, etc. Future studies could look into predicting worker trajectories within the discrete states of our model and predicting future state transitions. The application of gap acceptance models [17], [18] and intent prediction models [5], [26] to a manufacturing plant setting could be a good direction for future research. Moreover, the Hidden Markov Model with labeled sequence [31] could be applied to predict future worker's behavior. Second, the constraints in our model are highly dependent on the layout of the manufacturing plant. The constraints are highly intuitive and easy to set up given the layout. Another direction of future work would be to look into modifying AGV behavior based on the predictions to ensure smooth and trustworthy worker-AGV interaction. Finally, we currently assume that the AGV knows the real-time coordinates of the human worker at all times. This is not typically available to the AGV. In general, AGVs can detect obstacles using mounted sensors and can have an idea of the relative position of that obstacle from itself. Our model can readily be altered to use this relative position instead of the global coordinates.

Finally, an AGV running our model to determine a worker's state can then communicate some information (e.g. the worker's state, the AGV's state, intent, etc.). Future work could look into best practices for such non-verbal communication [32]–[34].

## ACKNOWLEDGEMENT

This work was funded by The Boeing Company. The authors would like to thank Jundi Liu for collecting the dataset used in this paper and Jiahang (Jay) Mao and Ruifeng Xu for developing the VR environment.

## REFERENCES

[1] Mckinsey & Company, "What is industry 4.0 and the Fourth Industrial Revolution?" 2022. [Online]. Available: https://www.mckinsey.com/featured-insights/mckinsey-explainers/what-are-industry-4-0-the-fourth-industrial-revolution-and-4ir

[2] X. Xu, Y. Lu, B. Vogel-Heuser, and L. Wang, "Industry 4.0 and Industry 5.0—Inception, conception and perception," *Journal of Manufacturing Systems*, vol. 61, pp. 530–535, Oct. 2021.

[3] S. Bhat, J. B. Lyons, C. Shi, and X. J. Yang, "Evaluating the Impact of Personalized Value Alignment in Human-Robot Interaction: Insights into Trust and Team Performance Outcomes," in *Proceedings of the 2024 ACM/IEEE International Conference on Human-Robot Interaction*, ser. HRI '24. New York, NY, USA: Association for Computing Machinery, 2024, pp. 32–41.

[4] R. Luo, Y. Weng, Y. Wang, P. Jayakumar, M. J. Brudnak, V. Paul, V. R. Desaraju, J. L. Stein, T. Ersal, and X. J. Yang, "A workload adaptive haptic shared control scheme for semi-autonomous driving," *Accident Analysis & Prevention*, vol. 152, p. 105968, 2021.

[5] S. K. Jayaraman, L. P. Robert, X. J. Yang, and D. M. Tilbury, "Multimodal Hybrid Pedestrian: A Hybrid Automaton Model of Urban Pedestrian Behavior for Automated Driving Applications," *IEEE Access*, vol. 9, pp. 27 708–27 722, 2021.

[6] A. Löcklin, F. Dettinger, M. Artelt, N. Jazdi, and M. Weyrich, "Trajectory Prediction of Workers to Improve AGV and AMR Operation based on the Manufacturing Schedule," *Procedia CIRP*, vol. 107, pp. 283–288, 2022.

[7] M. Bergman, S. Bedaf, G. Van Heel, and J. Sturm, "Can I Just Pass by? Testing Design Principles for Industrial Transport Robots," in *Proceedings of the 4th International Conference on Computer-Human Interaction Research and Applications*. Budapest, Hungary: SCITEPRESS - Science and Technology Publications, 2020, pp. 178–187.

[8] M. Luber, J. A. Stork, G. D. Tipaldi, and K. O. Arras, "People tracking with human motion predictions from social forces," in *2010 IEEE International Conference on Robotics and Automation*, 2010, pp. 464–469.

[9] P. Coscia, F. Castaldo, F. A. N. Palmieri, A. Alahi, S. Savarese, and L. Ballan, "Long-term path prediction in urban scenarios using circular distributions," *Image and Vision Computing*, vol. 69, pp. 81–91, 2018.

[10] A. Alahi, K. Goel, V. Ramanathan, A. Robicquet, L. Fei-Fei, and S. Savarese, "Social LSTM: Human trajectory prediction in crowded spaces," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 961–971.

[11] A. Vemula, K. Muelling, and J. Oh, "Modeling cooperative navigation in dense human crowds," in *2017 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2017, pp. 1685–1692.

[12] V. Karasev, A. Ayvaci, B. Heisele, and S. Soatto, "Intent-aware long-term prediction of pedestrian motion," in *2016 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2016, pp. 2543–2549.

[13] E. Rehder, F. Wirth, M. Lauer, and C. Stiller, "Pedestrian prediction by planning using deep neural networks," in *2018 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2018, pp. 5903–5908.

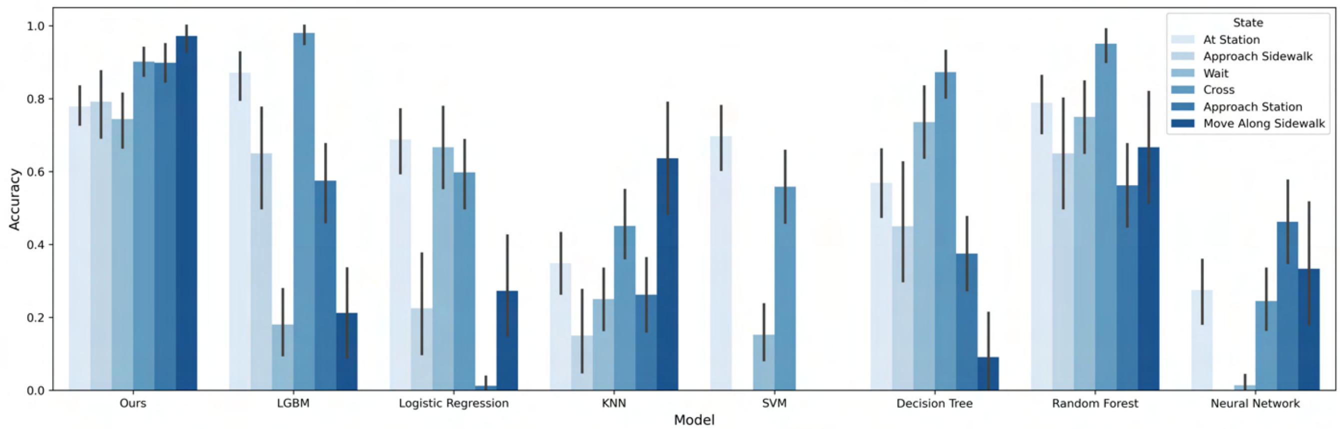


Fig. 7: Comparison of Model Accuracy by States.


[14] A. Rudenko, L. Palmieri, M. Herman, K. M. Kitani, D. M. Gavrila, and K. O. Arras, "Human Motion Trajectory Prediction: A Survey," *The International Journal of Robotics Research*, vol. 39, no. 8, pp. 895–935, Jul. 2020.

[15] S. K. Jayaraman, D. M. Tilbury, X. Jessie Yang, A. K. Pradhan, and L. P. Robert, "Analysis and Prediction of Pedestrian Crosswalk Behavior during Automated Vehicle Interactions," in *2020 IEEE International Conference on Robotics and Automation (ICRA)*. Paris, France: IEEE, May 2020, pp. 6426–6432.

[16] S. Jayaraman, L. Robert, J. Yang, and D. Tilbury, "Automated Vehicle Behavior Design for Pedestrian Interactions at Unsignalized Crosswalks," *SSRN Electronic Journal*, 2021.

[17] G. Yannis, E. Papadimitriou, and A. Theofilatos, "Pedestrian gap acceptance for mid-block street crossing," *Transportation Planning and Technology*, vol. 36, no. 5, pp. 450–462, Jul. 2013.

[18] B. R. Kadali, P. Vedagiri, and N. Rathi, "Models for pedestrian gap acceptance behaviour analysis at unprotected mid-block crosswalks under mixed traffic conditions," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 32, pp. 114–126, Jul. 2015.

[19] A. A. Tubis, H. Poturaj, and A. Smok, "Interaction between a Human and an AGV System in a Shared Workspace—A Literature Review Identifying Research Areas," *Sustainability*, vol. 16, no. 3, p. 974, Jan. 2024.

[20] C. Cronin, A. Conway, and J. Walsh, "State-of-the-art review of autonomous intelligent vehicles (AIV) technologies for the automotive and manufacturing industry," in *2019 30th Irish signals and systems conference (ISSC)*. IEEE, 2019, pp. 1–6.

[21] H. S. Hasan, M. Hussein, S. M. Saad, and M. A. M. Dzahir, "An overview of local positioning system: Technologies, techniques and applications," *International Journal of Engineering & Technology*, vol. 7, no. 3.25, pp. 1–5, 2018.

[22] W. Hellmann, D. Marino, M. Megahed, M. Suggs, J. Borowski, and A. Negahban, "Human, AGV or AIV? an integrated framework for material handling system selection with real-world application in an injection molding facility," *The International Journal of Advanced Manufacturing Technology*, vol. 101, pp. 815–824, 2019.

[23] T. Yokota, "Min-max-strategy-based optimum co-operative picking with AGVs in warehouse," in *2019 58th Annual Conference of the Society of Instrument and Control Engineers of Japan (SICE)*. IEEE, 2019, pp. 236–242.

[24] N. Muhammad, K. Hedenberg, and B. Astrand, "Adaptive warning fields for warehouse AGVs," in *2021 26th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA )*. Vasteras, Sweden: IEEE, Sep. 2021, pp. 1–8.

[25] S. Plosz and P. Varga, "Security and safety risk analysis of vision guided autonomous vehicles," in *2018 IEEE Industrial Cyber-Physical Systems (ICPS)*. IEEE, 2018, pp. 193–198.

[26] A. Locklin, M. Artelt, T. Ruppert, H. Vietz, N. Jazdi, and M. Weyrich, "Trajectory Prediction of Moving Workers for Autonomous Mobile Robots on the Shop Floor," in *2022 IEEE 27th International Conference on Emerging Technologies and Factory Automation (ETFA)*. Stuttgart, Germany: IEEE, Sep. 2022, pp. 1–8.

[27] S. Setia, V. Jyoti, N. Duhan, and D. Ruano-Ordás, "HPM: A hybrid model for user's behavior prediction based on n-gram parsing and access logs," *Sci. Program.*, vol. 2020, jan 2020.

[28] S. Molina, A. Mannucci, M. Magnusson, D. Adolfsson, H. Andreasson, M. Hamad, S. Abdolshah, R. T. Chadalavada, L. Palmieri, T. Linder, C. S. Swaminathan, T. P. Kucner, M. Hanheide, M. Fernandez-Carmona, G. Cielniak, T. Duckett, F. Pecora, S. Bokesand, K. O. Arras, S. Haddadin, and A. J. Lilienthal, "The ILIAD Safety Stack: Human-Aware Infrastructure-Free Navigation of Industrial Mobile Robots," *IEEE Robotics & Automation Magazine*, pp. 2–13, 2023.

[29] K. Hallgren, "Computing inter-rater reliability for observational data: An overview and tutorial," *Tutorials in Quantitative Methods for Psychology*, vol. 8, pp. 23–34, 07 2012.

[30] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[31] A. Krogh, "Hidden markov models for labeled sequences," in *Proceedings of the 12th IAPR International Conference on Pattern Recognition, Vol. 3 - Conference C: Signal Processing (Cat. No.94CH3440-5)*, vol. 2, 1994, pp. 140–144 vol.2.

[32] E. Cha, Y. Kim, T. Fong, and M. J. Mataric, "A survey of nonverbal signaling methods for non-humanoid robots," *Foundations and Trends® in Robotics*, vol. 6, no. 4, pp. 211–323, 2018.

[33] T. Schreiter, L. Morillo-Mendez, R. T. Chadalavada, A. Rudenko, E. Billing, M. Magnusson, K. O. Arras, and A. J. Lilienthal, "Advantages of multimodal versus verbal-only robot-to-human communication with an anthropomorphic robotic mock driver," in *2023 32nd IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, 2023, pp. 293–300.

[34] M. Pascher, U. Gruenefeld, S. Schneegass, and J. Gerken, "How to communicate robot motion intent: A scoping review," in *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, 2023, pp. 1–17.