Pilot Study Using Decision Trees to Diagnose the Efficacy of Virtual Offshore Egress Training

Jennifer Smith, Mashrura Musharraf, Brian Veitch, and Faisal Khan

Abstract—For the offshore energy industry, virtual environment technology can enhance conventional training by teaching basic offshore safety protocols such as onboard familiarization and emergency evacuation. Virtual environments have the added benefit of being used to investigate the impact of different training approaches on competence. This pilot study uses decision tree modeling to examine the efficacy of two pedagogical approaches, simulation-based mastery learning (SBML) and lecture-based training (LBT), in a virtual environment. Decision trees are an inductive reasoning approach that can be used to identify learners' egress strategies in offshore emergencies after training. The efficacy of the virtual training is evaluated in three ways: 1) analyzing participants' performance scores in test scenarios; 2) comparing the decision tree depiction of participant's understanding of emergency egress to the intended learning objectives; and 3) comparing the decision strategies developed under a different pedagogical approach. A comparison of the resulting decision trees from the SBML training with trees generated from the LBT showed that the different training methods influenced the participants' egress strategies. The SBML approach resulted in concise decision trees and better route selection strategies when compared to the LBT training. This pilot study demonstrates the diagnostic capabilities of decision trees as training assessment tools and recommends integrating decision trees into virtual training to better support the learning needs of individuals and deliver adaptive training scenarios.

Index Terms—Decision making in emergencies, decision trees, enter simulation-based mastery learning (SBML), offshore emergency egress, training efficacy, virtual environments.

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by Memorial University of Newfoundland's Interdisciplinary Committee on Ethics in Human Research under Application No. 20160794-EN and performed in line with the Tri-Council Policy Statement on Ethical Conduct for Research Involving Humans (TCPS2).

Jennifer Smith is with the Faculty of Engineering and Applied Science, Memorial University of Newfoundland, St. John's, NL A1B 3X5, Canada, and also with the School of Maritime Studies at the Fisheries and Marine Institute, Memorial University of Newfoundland and Labrador, St. John's, NL A1C 5R3, Canada (e-mail: jennifer.smith@mi.mun.ca).

Mashrura Musharraf is with the Faculty of Engineering and Applied Science, Memorial University of Newfoundland, St. John's, NL A1B 3X5, Canada, and also with Mechanical Engineering Department, Aalto University, 02150 Espoo, Finland (e-mail: mashrura.musharraf@aalto.fi).

Brian Veitch is with the Faculty of Engineering and Applied Science, Memorial University of Newfoundland, St. John's, NL A1B 3X5, Canada (e-mail: bveitch@mun.ca).

Faisal Khan is with the Centre for Risk, Integrity, and Safety Engineering and the Faculty of Engineering and Applied Science, Memorial University of Newfoundland, St. John's, NL A1B 3X5, Canada, and also with the Department of Chemical Engineering, Texas A&M University, College Station, TX 77843 USA (e-mail: fikhan@tamu.edu).

Digital Object Identifier 10.1109/TLT.2022.3207979

I. INTRODUCTION

EGRESS training is essential for coordinating the safe evacuation of offshore platforms and maritime vessels. For instance, in response to the January 2012 Costa Concordia accident that resulted in 32 fatalities [1], the International Maritime Organization enacted regulatory changes to the Safety of Life at Sea (SOLAS) Convention on Emergency Training and Drills [2]. As a result, to ensure crews and passengers are adequately prepared for emergencies at sea, the SOLAS Resolution MSC.350(92) requires that all passengers onboard are provided with muster drills and safety briefings prior to or immediately after departure [2].

Virtual environment technology can enhance conventional training for maritime and offshore energy industries by providing crews with worksite familiarization, practice with safety-critical operations, and experience in responding to emergencies. Virtual environment training, which will be referred to as virtual training throughout the remainder of this article, also has built-in capabilities to track and record the learner's performance during the training. The data are predominately used to assess the learner's competence, provide feedback, and deliver adaptive training scenarios. An additional advantage of the data recording capabilities of the virtual training is the added value of using data to evaluate the efficacy of the virtual training itself. Conventional ways of evaluating new training technologies involve experiments with volunteer participants to collect and report the statistical differences in human performance between multiple training interventions. However, these methods can be logistically challenging, time consuming, and costly.

This pilot study maximizes the use of data collected from empirical approaches and explores whether decision tree modeling can offer an alternative to evaluating training efficacy. This article uses the published human performance data from the two experiments [3], [4], to investigate the utility of decision tree modeling for evaluating different virtual training methods. The separate experiments were originally designed to test the application of two different pedagogical approaches in training naïve personnel for basic emergency duties, specifically, lecture-based training (LBT) and simulation-based mastery learning (SBML). The first experiment used conventional LBT training [3] to expose participants to emergency egress using video tutorials, platform walkthroughs, practice scenarios, and test scenarios in a first-person perspective virtual environment. The second experiment was conducted to assess

the relative merits of the SBML pedagogical approach [4]. Applying the SBML framework in this experiment, the training repeatedly exposed participants to emergency egress using platform walkthroughs, practice scenarios, and test scenarios. Further, the SBML training also included a minimum passing standard that required participants to master a topic before moving on to the next module.

Overall, the same first-person perspective virtual environment, learning objectives, test scenarios, and performance metrics were used in both studies. The key difference between the two experiments was the pedagogical approach, which included the delivery framework, formative assessment, minimum passing requirement, and feedback method. This was intentionally controlled so the results from both experiments could be compared in terms of pedagogy.

While the datasets collected from the pedagogical experiments were small, it is considered sufficient for a pilot study to demonstrate the value of using human performance data to model learning. As part of this pilot, data mining methods were applied to the datasets from the two experiments to further evaluate the efficacy of the different virtual delivery methods. Specifically, a decision tree algorithm was applied to the training datasets to create a collection of decision trees for each training approach. This information was used to determine how well the virtual training prepared participants for emergency scenarios. As an example, Musharraf et al. [5] applied decision trees to the LBT dataset to identify individuals' decision-making strategies in the context of selecting safe egress routes in virtual offshore emergencies. This article builds on the research from Musharraf et al. [5] by developing decision trees from the SBML dataset and comparing them to those developed from the LBT dataset.

The results of this pilot study provide a closer look at the diagnostic capabilities of decision tree modeling, as a complement to conventional performance metrics, to investigate the training efficacy of the pedagogical approaches. The analysis involves three methods to evaluate the virtual training:

1) an empirical analysis of the participants' performance scores in the test scenarios;
2) a comparison of the resulting decision trees against the intended learning objectives;
3) a comparison of the decision trees generated from two training studies as a way of assessing the relative merits of the two pedagogical approaches in terms of improving the performance of participants.

To orient the reader, the introduction will continue by outlining the two pedagogical frameworks employed in the research and explaining the decision tree theory used as a data-informed diagnostic lens to assess the virtual training.

A. Pedagogical Frameworks

LBT is a learning approach that is instructor-centered and in line with traditional lecture-style instruction [6]. This method is a passive form of training in which learners are exposed to the content through video instructions, demonstrations, and practice exercises. LBT represents a virtual facsimile of the orientation training that crews receive in the offshore energy and maritime industries. However, there are limitations to this form of training. Mainly the passive assessment often does not include a minimum passing component or feedback on how to improve performance, which can leave the learner with little means to assess their comprehension or gauge their progress.

SBML is designed to meet the needs and pace of the individual learner. SBML is a pedagogical approach developed in the medical education field [7], [8], [9], [10], [11], [12], [13], [14] and is based on Bloom's competence-based theory of learning for mastery. Bloom's mastery learning is an instructional strategy that ensures all learners achieve competence by providing them with formative assessment, opportunities for deliberate practice, individualized feedback, and corrective measures [15], [16]. The SBML protocol builds on Bloom's theory by using virtual environments and simulation to provide instruction and assessment. The SBML framework assesses the learner's entry-level knowledge, gradually walks the learners through the increasingly difficult content, and requires that the learners deliberately practice the exercises until they demonstrate competence. Learners are provided with formative assessments throughout the training, which includes constructive feedback for them to improve or correct their performance. The instructions, assessment, and feedback are automated in the AVERT simulator. Once the learners have demonstrated their understanding in test exercises (e.g., performing at or above the minimum passing requirement), they are able to move on to more advanced content. The SBML training allows learners to gauge their progress and requires that they demonstrate competence in completing the training.

B. Decision Tree Modeling

Decision trees are common classification approaches used in educational data-mining applications [17] and in artificially intelligent enabled adaptive instructional systems [18]. This pilot study uses decision trees to evaluate how well the virtual training prepared participants for emergency scenarios. Decision trees were selected for their visual simplicity, quick construction, and because they require no prior assumptions about the data, particularly when compared to other methods, such as artificial neural network or support vector machines [19]. Further, decision trees have behavioral pattern recognition capabilities that go beyond conventional methods of tracking trainee progress and performance outcomes and offer a diagnostic lens to assess training efficacy.

Decision tree modeling is one of several supervised learning techniques that are particularly well-suited for virtual training applications because they employ a repository of solved problems to draw inferences. For instance, virtual training can record each user's in-simulation performance data during practice exercises and store this data in a user-specific data repository. Decision tree modeling applies an algorithm to the observed performance data (i.e., collected during the virtual training) to develop generalized decision rules [20]. These generalized rules can be used for many applications, as illustrated by Musharraf et al. [5] application of decision trees for
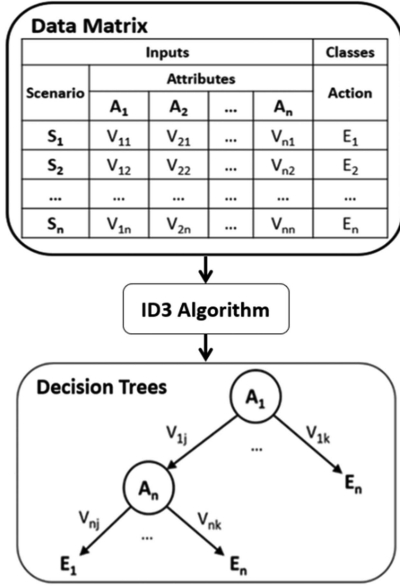
Fig. 1. Decision tree development. Attributes in the data matrix are classified using the ID3 algorithm so that the data in each group belongs to the same class. The class is the participant's action in the scenario (e.g., route choice).

identifying individual's egress route strategies in virtual training. The main benefits of decision trees for the offshore emergency egress application were that the decision trees were easy to interpret, useful in identifying patterns in participants' performance, and had diagnostic potential for determining the strengths and weaknesses of different decision-making strategies. Building on the research from [5], this article investigates decision tree modeling as a tool to assess different delivery approaches to virtual training.

C. Organization of the Paper

The rest of this article is organized as follows. Section II presents the theoretical background for the decision tree induction process. Section III explains the design of the virtual training experiments and the application of the decision tree modeling to the datasets. Section IV presents the performance results and subsequent decision trees from the two virtual training approaches. Finally, Section V concludes the strengths and weaknesses of the training and the utility of decision trees in assessing the virtual training.

II. THEORETICAL BACKGROUND

The decision tree algorithm employs an induction process whereby generalizations are made based on observed phenomena [20]. Following the rule-based methodology [21], a data matrix is created using human performance data from virtual training. In this pilot study, information from each participant's performance in virtual training scenarios is used to populate a data matrix consisting of scenarios (S_1–S_n), attributes (A_1–A_n), attribute values (V_{11}–V_{nn}), and actions (E_1–E_n). The scenarios and attributes are labeled inputs to the matrix and the participants' corresponding actions in the scenarios are known as classes. As depicted in Fig. 1, the induction process creates

generalized decision rules based on the content of the data matrix. The goal of the induction process is to classify the data in the matrix into groups such that the dataset in each group belongs to the same class. This article uses the ID3 decision tree algorithm, which uses *information gain* as an attribute selection measure, to classify the data into groups [22].

The ID3 decision tree algorithm takes two basic inputs: the performance data matrix from the virtual training scenarios and the list of attributes that were varied in each scenario. The output is a decision tree that describes a participant's decision preferences and can also be used to predict their future decisions based on the value of the attributes in a given scenario. During the decision tree induction, data are iteratively classified using the attribute that has the highest information gain. The ID3 algorithm calculates the highest information gain using three main calculations:

1) the entropy of the dataset;
2) the average information entropy of attributes;
3) the information gain for each attribute.

First, the entropy of the entire dataset is calculated as a measure of the uncertainty of the data [19]. This is achieved by defining the data matrix training set as S, where S contains m class labels and S_i is a subset of scenarios within the training set S. Then the entropy of S is calculated as

$$\text{Entropy}(S) = -\sum_{i=1}^{m} \frac{|S_i|}{|S|} \log_2 \frac{|S_i|}{|S|}. \tag{1}$$

Second, the training set S is partitioned using attribute A, where A has k distinct outcomes. This partition will result in subset S_j with $j = 1$ to k values. The average information entropy for all attributes (A_1–A_n) in S_j is calculated as

$$\text{Entropy}(A) = \sum_{j=1}^{k} \frac{|S_j|}{|S|} \text{Entropy}(S_j). \tag{2}$$

Finally, the *information gain*, which is the difference in entropy before and after splitting the dataset on the attribute A is calculated for each attribute in the data matrix as

$$\text{Gain}(A) = \text{Entropy}(S) - \text{Entropy}(A). \tag{3}$$

The attribute with the highest *information gain* is selected as the root node, which begins the partition of the dataset. The root node represents the attribute that minimizes the information needed and reduces the randomness of the partitions [22]. This process repeatedly splits the data subsets at each internal node until no attributes are left for classification, or the dataset is empty, or data in each group belong to the same class and no further classification is needed [5]. A complete tree has branches to leaf nodes (that represent the class label or final action of the participant). Algorithm 1 describes the iterative steps used to develop a decision tree.

III. METHODOLOGY

This section describes the design of the experiments and explains the decision tree modeling from the dataset.

Algorithm 1: Generate Decision Tree from Data Matrix [22]

Input: data matrix, attribute list, attribute selection method
Output: decision tree
1:   create a node, A_i
2:     **if** all scenarios S_n at current node are of same class **then**
3:       label the leaf nodes with the class labels
          (e.g., branch, V_n; leaf node, E_n);
4:     **end if**
5:     **if** data subset at the current node is empty **then**
6:       label the node with the majority class label in its parent
            dataset (e.g., branch, V_n; internal node, A_n);
7:     **end if**
8:     **if** no attributes are left for further classification **then**
9:       label leaf node with majority class label in current data
            subset (e.g., branch, V_n; leaf node, E_n);
10:   **end if**
11:   **for** each remaining attribute A_n
12:       compute $Gain(A_n)$ according to (1), (2), and (3);
13:       choose A_n with highest $Gain(A_n)$ to branch current node;
14:   **end for**
15:   **for** each branch node V_n go to step 2
16:   **end for**

A. Virtual Training Experiments

This pilot study used previously published datasets from two separate experiments to test the LBT and SBML pedagogical approaches to virtual training [3], [4]. The following will describe the participants, the AVERT simulator, and how the training approaches were applied to virtual training.

1) Participants: To measure learning from the virtual training, all participants from both experiments had no prior offshore experience and no exposure to the AVERT simulator prior to the study. Most participants for both experiments were undergraduate and graduate students. The LBT experiment had 36 naïve participants. For this experiment, the participants were divided into two treatment groups: one group received multiple training exposures, and the control group received only one training exposure. This pilot study includes the results of 17 participants from the multiple training exposure groups (represented by the label LBT, and of which 13 participants were male and four participants were female). The LBT participants' ages ranged from 19 to 39 years (mean = 27 years, standard deviation = ± 5.0 years). This pilot study also includes the results from the SBML training experiment which had 55 naïve participants (42 participants were male and 13 participants were female). The SBML participants' ages ranged from 18 to 54 years (mean = 27 years, standard deviation = ± 7.9 years).

2) AVERT Simulator: Both experiments trained participants in offshore emergency egress using the AVERT simulator. AVERT is a desktop virtual environment (e.g., computer-based simulator) that provides participants with a first-person perspective naturalistic representation of an offshore Floating Production Storage and Offloading vessel [23]. Participants use a gamepad controller (Xbox) to control their avatar of an offshore worker and interact with the virtual offshore platform. AVERT is configured to train general personnel in basic offshore emergency
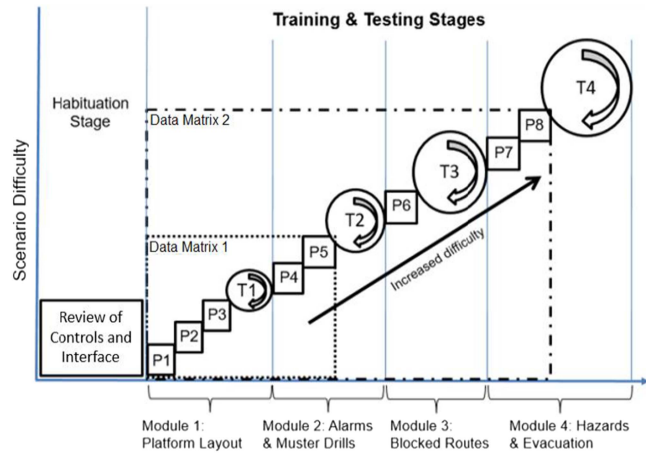

Fig. 2.   SBML training and testing stages. The squares represent practice scenarios and included in-simulation instruction, exercises, and feedback. The circles represent test scenarios that were performed until the criterion was reached.

egress duties. General personnel are individuals whose responsibility during an emergency is to muster at their designated muster stations. The AVERT learning objectives were developed with guidance from subject matter experts to address both spatial and procedural knowledge. The learning objectives included familiarity with the platform layout, emergency alarms, egress routes, safety protocols, and mustering procedures.

3) Applying the Pedagogical Frameworks to AVERT: Both implementations of the virtual training involved an initial habituation stage, followed by training and testing modules. The habituation stage familiarized participants with how to use the AVERT controls and introduced participants to the offshore platform. After the habituation stage, participants proceeded to the training and testing modules. The training and testing modules targeted the same learning objectives for both the LBT and SBML training. For the LBT training, participants were provided with tutorials, video instructions, and practice scenarios in the virtual environment before completing the virtual testing scenarios for each module. Details of the LBT training are provided in [3].

The SBML training involved four training and testing modules as depicted in Fig. 2. Each module was designed to train specific learning objectives and gradually taught participants the platform layout, how to recognize alarms, what to do in the event of blocked routes, and how to assess the situation and avoid hazards while evacuating the platform. Each training and testing module involved 1–3 practice scenarios and one test scenario. As shown in Fig. 2, the SBML training consisted of 12 scenarios in total (eight practice and four test scenarios). As part of the SBML training, participants were required to demonstrate competence in each scenario before they could advance to scenarios that were more complex.

Module 1 taught participants the platform layout and all the available egress routes from their cabin. Participants were tested on their spatial knowledge in scenario T1 by asking them to meet their supervisor at their designated lifeboat station. Module 2 taught participants the different alarm types on the platform: general platform alarm
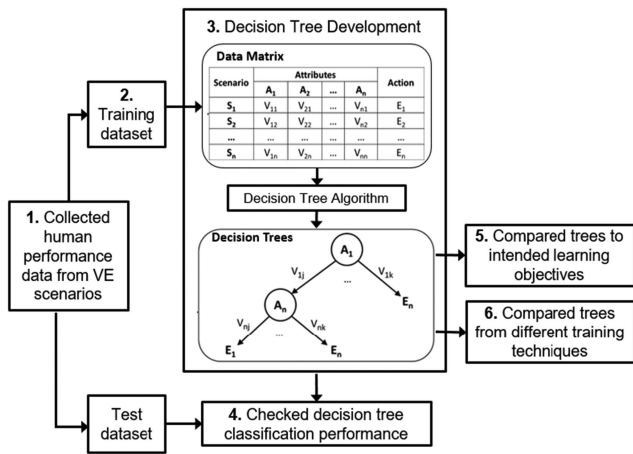
Fig. 3. Process used for each participant's data to develop decision trees and assess the training efficacy. After the decision trees were developed, the classification performance was evaluated. Then the decision trees were compared to the learning objectives and across training techniques.

TABLE I
DESCRIPTION OF THE TEST SCENARIOS [4]

Test Scenario	Scenario Description
T2 Muster Drill	This scenario assessed the participants' understanding of alarms and muster procedures. Participants responded to a muster drill (General Platform Alarm). During this alarm, all personnel were required to collect their safety equipment and muster at their primary muster station.
T4 Emergency Situation	This scenario assessed the participants' ability to avoid hazards and follow the safest route to their lifeboat station. Participants responded to an emergency involving a General Platform Alarm due to fire in the galley. The fire compromised the muster station with smoke and the situation escalated to a Prepare to Abandon Platform Alarm. Initially all personnel were required to go to the muster station but were forced to reroute to the lifeboat station because of the compromised muster station.

(GPA), prepare to abandon platform alarm (PAPA), and mustering procedures. Participants were tested on their spatial and procedural knowledge in scenario T2 by asking them to respond to a muster drill. Module 3 reminded participants of the alternative routes from the cabin to ensure they knew the route options in the event their egress route was obstructed. Module 4 taught participants the protocols necessary to respond to emergency scenarios with hazards such as smoke and fire. Test scenarios T3 and T4 tested the participants' ability to respond to emergency conditions and reroute if their planned route was blocked by hazards.

After each training module in AVERT, the participants' performance was assessed using test scenarios. In the scenarios, participants were tasked with responding to muster drills or emergency alarms and selecting the safest egress route from their cabin. There were two main routes for participants to choose from: primary or secondary. Each route had multiple decision points along the path. Participants were instructed to listen to the alarm, pay attention to the public address (PA) announcements, and follow the safest route to their muster or lifeboat station. Participants were assessed on their ability to recognize the alarm, take their safety equipment, follow the safest egress route, avoid exposure to hazards, reach the correct muster location, and register at the temporary safe refuge area. Participants received corrective feedback on their performance after each scenario attempt. To demonstrate competence, some participants required multiple attempts at the scenarios.

B. Decision Tree Modeling of the Virtual Training Data

The decision tree development and analysis framework are depicted in Fig. 3. This process involves six steps. First, each participant's performance data collected from the two virtual training experiments were separated into two datasets: a training and a testing dataset. Second, the training dataset (representing 2/3 of the participant's data) was used to develop a data matrix consisting of scenarios, attributes, values, and actions. The test scenarios (representing 1/3 of the participant's data)

were set aside to form the testing dataset. Third, the decision tree algorithm was applied to the data matrix to form decision trees, which represent each participant's behavioral pattern for route selection [5]. Once the decision trees were generated, the fourth step involved using the testing dataset to check the decision tree classification performance. The final steps involved using the decision trees to analyze the training efficacy. In step five, the resulting decision trees were used to compare the participants' understanding of the training with the learning objectives for each test scenario. Finally, in step six, the decision strategies from the SBML training were compared with the decision trees generated using data from the earlier LBT experiment [5].

1) Virtual Training Data Collection: As participants completed the scenarios, their performance data was collected in AVERT report files for each scenario. Observation logs were kept by the researcher to note any details that were not recorded in the automated report files. Each participant's data were organized into training and testing datasets. Of the 12 scenarios, 11 were used for the decision tree development. One training scenario was an orientation scenario and was not used in the analysis. Among the remaining 11 scenarios, 9 were used to populate the data matrix to train the decision tree algorithm and form the decision trees. These scenarios are referred to as the training dataset. Two test scenarios, T2 and T4, were set aside to form the testing dataset and are described in Table I. The testing dataset was used to evaluate the classification performance of the decision trees.

By T2, participants had familiarized themselves with the platform layout, the different alarm types, and the mustering procedures at the temporary safe refuge area. By T4, participants were able to assess the emergency, listen to cues in the PA announcement, recognize the tenability of the egress routes, and reroute if the primary or secondary egress route was obstructed due to poor lighting or other barriers.

2) Data Matrix: A two-dimensional data matrix was created using each participant's performance data collected from the AVERT report files and from observations logged *in situ*. As shown in Fig. 2, the data matrices were developed to

TABLE II
DESCRIPTION OF SCENARIO ATTRIBUTES

Attribute	Possible Values
End location	Muster, Lifeboat
Alarm type	None, General Platform Alarm (GPA),
	Prepare to Abandon Platform Alarm (PAPA)
Hazard presence	No, Yes
Route directed by PA	None, Primary route, Secondary route
Obstructed route	None, Primary route, Secondary route
Previous route taken	N/A, Primary route, Secondary route

The scenario attributes were varied for each practice and test scenario by changing the severity of the situation (e.g., adding or removing hazards, alarms, and PA information).

correspond with the training scenarios that were completed in two training modules: module 2 (denoted Data Matrix 1) and module 4 (denoted Data Matrix 2). The data matrix consisted of a combination of programmed attributes and the participants' actions. The programmed scenario attributes were varied for each scenario, such as the end location, alarm type, information presented in the PA announcements, presence of hazards, and location of obstructed routes. For each scenario, the data matrix included a record of the participant's actions, such as their route choices in the current and previous scenarios. Table II lists the attributes varied in the scenarios and their possible values.

Based on the value of the scenario attributes, the participant's goal was to select a safe egress route. Since the SBML training required participants to reattempt the scenarios until they demonstrated competence, only the participant's successful final attempt was stored in the data matrix. Table III shows the state of the data matrix for a sample participant after finishing all the training modules. Each row in the matrix contains different attribute values for the scenario and the corresponding route choice. As a basic example, scenario P4 from Table III is a situation in which participants practiced their egress routes and muster procedures. For a sample participant, the scenario attributes were recorded as End location = Muster; Alarm type = GPA; Route directed by PA = Primary, Hazard presence = No, Blocked route = none, and Previous route = Primary. In this case, the participant's choice of route was the primary route.

Complex emergency scenarios were dynamic in the sense that the value of some attributes changed during the scenarios. To capture the dynamic aspect, these scenarios were divided into multiple frames so that the attribute values in each frame remained static (e.g., the first frame F1 depicted the initial conditions and the second frame F2 depicted the changed conditions). Consequently, many of the emergency scenarios were multiframes scenarios. Fig. 4 shows an example of a two-frame training scenario (P8) and how the data matrix was updated to reflect the change in attributes. Scenario P8 is an emergency in which participants responded to changing conditions. For this sample participant, the scenario attributes in F1 were initially recorded as End location = Muster; Alarm type = GPA; Route directed by PA = None, Hazard presence = No, Blocked route = None, and Previous route = Secondary. However, the severity of the situation escalated in F2 and some attributes changed: End location = Lifeboat; Alarm type = PAPA, Route directed

by PA = Primary, Hazard presence = Yes, and Blocked route = Secondary. As such, the participant's choice was originally the primary route, but they rerouted to the secondary route when the attribute values changed.

3) Decision Trees: The data matrix generated in the previous step was used as an input for the decision tree algorithm. The resulting decision trees were used to visualize how participants formed emergency egress rules based on the content in the data matrix. They also provided insight as to which attributes had the biggest impact on participants' decision-making. Fig. 5 shows a decision tree based on the matrix in Table III for a sample participant.

The decision tree is based on evidence from the participant's performance in a series of virtual scenarios. The decision tree can be used to predict a participant's choice of route for a given future scenario. In this case, the participant's route selection was decided based on their understanding of the PA announcement. In future scenarios, if the PA directs the participant to a safest route, then the participant will likely take that route. If the PA does not provide any information regarding the safest route, then the participant's choice will likely default to their primary egress route.

4) Evaluating the Decision Tree Classification: There are limitations to the method used to split the participants' data to develop the decision tree and evaluate the classification, specifically when compared with other methods such as cross-validation. For example, cross validation involves dividing the dataset into mutually exclusive and equal-sized training sets to train the decision algorithm and test the resulting decision tree models on all the subsets [22], [24]. However, the cross-validation approach would not allow for researchers to observe the participants' learning as they progressed through the training which required the time-series formation of decision trees throughout training.

Using human performance data to develop the decision trees created an inherent class imbalance in the training dataset. For each participant, the training dataset was comprised of the participants' training exercises. These training exercises often contained more classes of a certain type over other classes (e.g., more examples of choosing the primary egress route). In this case, the route options were binary. The majority class label was used for the instance when the participant chooses the primary route and the minority class label was used for the instance when the participant choose the secondary route.

Common classification measures like classification accuracy or error are not appropriate when a class imbalance exists. Therefore, three threshold metrics specific for class imbalance were used to evaluate the decision trees. Methodology outlined in [22], [25] was used for defining the confusion matrix and calculating the classification performance. To test the decision tree's classifying performance, the predicted routes of the decision trees for the test scenarios (T2 and T4) were compared to the routes taken by the participants. For each comparison, the correct and incorrect classifications of the binary route choices were counted.

The terms used to evaluate the classification are outlined in the confusion matrix in Table IV. For the secondary route,

TABLE III
SAMPLE DATA MATRIX FOR TRAINING MODULE 2 (DATA MATRIX 1) AND TRAINING MODULE 4 (DATA MATRICES 1 AND 2)

| Category | Scenario | Attributes | | | | | | Route choice |
		End Location	Alarm	Route by PA	Hazard	Blocked Route	Previous Route	
Data Matrix 1	P1	Muster	None	Primary	No	None	N/A	Primary
	P3 (F1)	Lifeboat	None	Primary	No	None	Primary	Primary
	P3 (F2)	Muster	None	Secondary	No	None	Primary	Secondary
	T1	Lifeboat	None	None	No	None	Secondary	Primary
	P4	Muster	GPA	Primary	No	None	Primary	Primary
	P5	Muster	GPA	None	No	None	Primary	Primary
Test 1*	T2	Muster	GPA	None	No	None	Primary	Primary
Data Matrix 2	P6	Lifeboat	PAPA	Primary	No	Secondary	Primary	Primary
	T3	Muster	GPA	Secondary	No	Primary	Primary	Secondary
	P7	Lifeboat	PAPA	Secondary	Yes	Primary	Secondary	Secondary
	P8 (F1)	Muster	GPA	Primary	Yes	Secondary	Secondary	Primary
	P8 (F2)	Lifeboat	PAPA	Primary	Yes	Secondary	Secondary	Primary
Test 2*	T4 (F1)	Muster	GPA	Secondary	Yes	Primary	Primary	Secondary
	T4 (F2)	Lifeboat	GPA	Secondary	Yes	Primary	Primary	Secondary
	T4 (F3)	Lifeboat	PAPA	Secondary	Yes	Primary	Primary	Secondary

Highlighted rows represent the scenarios used to evaluate the decision trees classification performance.

Scenario P8 (Frame 1)

Attribute	Value
End location	Muster
Alarm Type	GPA
Route direction in PA	None
Presence of Hazard	No
Obstructed Route	None
Route taken in prev. scenario	Secondary
Action	
Choose egress route = Secondary	

Scenario P8 (Frame 2)

Attribute	Value
End location	Lifeboat
Alarm Type	PAPA
Route direction in PA	Primary
Presence of Hazard	Yes
Obstructed Route	Secondary
Route taken in prev. scenario	Secondary
Action	
Choose egress route = Primary	

t = 0   t1   t2   time

Fig. 4. Example of scenario frames 1 and 2 for scenario P8.

Route direction by PA

Primary   Secondary   None

Choice of route = Primary   Choice of route = Secondary   Choice of route = Primary
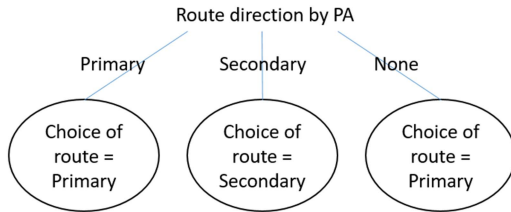
Fig. 5. Example decision tree developed after data matrices 1 and 2 in Table III. The decision rules for this tree were based on the participant's understanding of the PA announcement.

known as the minority class, the number of correct predictions is denoted as a true positive (TP) and the number of incorrectly predicted secondary route is denoted as a false positive (FP). For the primary route, known as the majority class, the number of correct predictions is denoted as a true negative (TN) and the incorrectly predicted primary route is denoted as a false negative (FN).

The confusion matrix was used to calculate the sensitivity, specificity, and geometric mean for all the SBML and LBT participants' decision trees

To evaluate the decision trees' classification performance, the following three evaluation measures that take class imbalance

TABLE IV
CONFUSION MATRIX

| Classes | Decision Tree Predicted Class | |
	Secondary Route	Primary Route
Actual Class: Secondary Route (participant choice)	True Positive (TP)	False Negative (FN)
Actual Class: Primary Route (participant choice)	False Positive (FP)	True Negative (TN)

The confusion matrix was used to calculate the sensitivity, specificity and geometric mean for all the SBML and LBT participants' decision trees

into consideration were used: 1) sensitivity, 2) specificity, and 3) geometric mean [20], [25]. Sensitivity is a measure of the proportion of the minority class that is correctly classified (e.g., choosing the secondary route). The sensitivity was calculated using (4) and represented the proportion of matches between the predicted secondary route and the observed secondary route outcome

$$\text{Sensitivity} = \frac{\text{TP}}{(\text{TP} + \text{FN})}. \quad (4)$$

Specificity is a measure of the portion of the majority class that is correctly identified (e.g., choosing the primary route). The specificity was calculated using (5) and represented the proportion of matches between the predicted primary route and the observed primary route outcome

$$\text{Specificity} = \frac{\text{TN}}{(\text{FP} + \text{TN})}. \quad (5)$$

Finally, the geometric mean is a combined score that measures the balance between sensitivity and specificity [25] and was calculated as

$$\text{GeometricMean} = \sqrt{\text{Sensitivity} \times \text{Specificity}}. \quad (6)$$
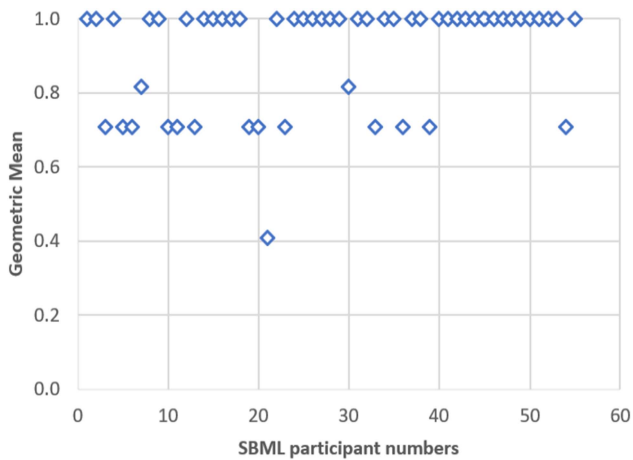
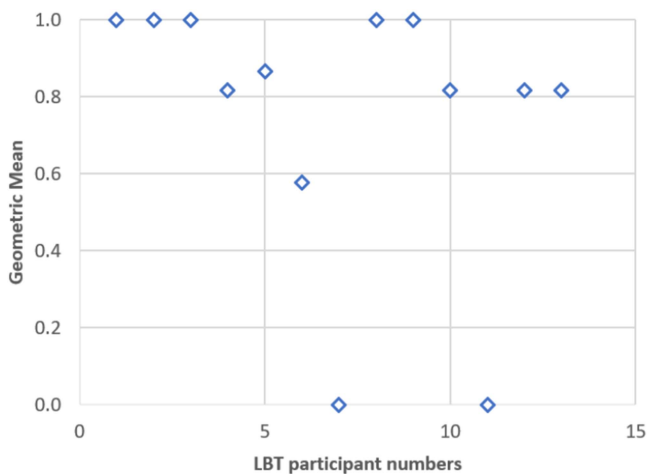Fig. 6. Geometric mean for all SBML trained participants.

Fig. 7. Geometric mean for all participants from the LBT training.

The resulting geometric mean scores for the SBML and LBT participants decision trees are represented in Figs. 6 and 7, respectively. The results indicate that the decision trees were suitable for classifying the decision strategies of participants. The results also show the decision trees' predictive potential.

IV. RESULTS AND DISCUSSION

The virtual training efficacy was assessed using three measures:

1) analyzing the trained participants' performance scores in the test scenarios;

2) comparing the participants' decision trees to the intended learning objectives; and

3) comparing the SBML and LBT-trained groups decision strategies.

The following subsections summarize the findings.

A. Empirical Results of SBML Training

Table V shows the percentage of SBML and LBT-trained participants who successfully completed each learning objective for scenarios T2 and T4. Differences in compliance were

TABLE V
PERCENTAGE OF SUCCESSFUL PARTICIPANTS BY LEARNING OBJECTIVE [4]

Performance Measures	T2 Muster Drill		T4 Emergency	
	SBML n = 55	LBT n = 17	SBML n = 55	LBT n = 17
Spatial Performance:				
1. Reached correct location	100%	94%	93%	94%
2. Correctly selected and followed safest egress route	98%	88%	55%	35%
3. Rerouted based on PA information	n/a	n/a	22%	0%
4. Rerouted if path blocked (encountered hazard)	n/a	n/a	16%	18%
5. Did not reroute (opened door to hazard and/or went through the hazard)	n/a	n/a	7%	47%
Procedural Performance:				
6. Recognized alarm and registered at TSR	98%	94%	100%	94%
7. Avoided hazard exposure	n/a	n/a	93%	53%
8. Avoided running	100%	18%	100%	24%
9. Closed all fire and watertight doors	96%	59%	93%	65%

n/a = not applicable. Some performance metrics did not apply to all scenarios.

observed between the groups for the spatial and procedural learning objectives.

Statistically comparing the results from two groups that were not randomly assigned in a controlled experiment is considered a quasi-experiment as it violates the assumptions of statistical independence [26]. This is the case for comparing the results of the SBML and LBT-trained participants in their ability to respond to virtual emergencies. To allow for comparisons, measures were taken to control aspects across the two experiments (e.g., recruiting from a similar population and testing participants using the same virtual environment and test scenarios). However, since the two training approaches were not tested in a controlled experiment, the observed differences between the groups may not be solely a result of the training but other confounding factors (e.g., the age ranges for the groups are different, with an older population in the SMBL group).

Fisher's Exact test [27] was used to determine if there was a relationship between the training received and the performance outcome for each learning objective. Contingency tables were created for each of the learning objectives and the number of participants from each training group who passed or failed the learning objectives were counted. The null hypothesis was that there would be no difference in the performance of the learning objectives between the two trained groups. The critical value for rejecting the null hypothesis was set to $\alpha = 0.05$.

From a spatial competence perspective, both the SBML and LBT-trained groups were able to locate the correct muster location and follow the egress routes in benign conditions. This is shown in the results from the muster drill scenario (T2 in Table V). In the emergency scenario (T4 in Table V), the main spatial competence differences observed between the SBML and LBT groups were related to route selection and rerouting when the egress path was blocked by hazards. This is supported by the Fisher's Exact tests, which resulted in no

statistical association between the training and the pass rate of the spatial tasks for the test scenarios, with the except for one learning objective for scenario T4. The Fisher's Exact test indicated that the proportion of SBML-trained participants who correctly rerouted to avoided hazard exposure in T4 was statistically different from the proportion of LBT-trained participants (p = 0.00059). Table V shows how differently participants in each group reacted to hazards blocking their egress routes, four participants from the SBML group (representing 7%) and eight participants from the LBT group (representing 47%) continued on the unsafe route and went directly through the smoke hazard.

While this work focused mainly on the spatial learning objectives 1–5 in Table V, it is worth noting that large differences were observed between the trained groups in the procedural performance. Specifically, learning objectives 7, 8, and 9 (i.e., avoiding hazards, refraining from running, and remembering to close the fire and watertight doors). The Fisher's Exact tests indicated statistical associations between the training and the pass rate of the procedural tasks for the test scenarios. The proportion of SBML-trained participants who correctly avoided hazards, refrained from running, and closed the fire and watertight doors were statistically different from the proportion of LBT-trained participants (e.g., avoiding hazard exposure, p = 0.00059; avoiding running, p = 3.36e-11; and closing the fire and watertight doors, p = 0.0089).

B. Decision Trees and Learning Objectives for SBML

Data collected from the LBT study was used by [5] to develop decision trees and identify general problem-solving strategies in emergency egress situations. The resulting decision trees showed that given the same training, people employed different learning strategies and developed their understanding of emergency protocols differently. Decision-making in high-stress emergencies varied from person to person. These results coincide with the empirical results [3], which found that the LBT did not provide an adequate assessment (i.e., practice and feedback) to ensure all participants gained competence.

The decision trees were also used to judge the efficacy of the SBML training by comparing the SBML-trained group's decision trees to the intended learning objectives at two stages of the training program. The different decision trees for the SBML training are summarized in Table VI. This table shows how the trees evolved as more training content was added to the participants' data repository.

1) Alarm Recognition Decision Tree (From Data Matrix 1): In the muster drill (T2) and the emergency (T4), the alarm type indicated the severity of the situation and dictated the final muster location (e.g., muster or lifeboat station). During the GPA alarm, personnel were required to gather at the muster station. During the PAPA alarm, personnel were required to muster at the lifeboat station. The main learning objective for module 2 was for participants to listen to the alarm and relevant instructions from the PA announcement and take the safest route available in response to the situation.

TABLE VI
TYPES OF DECISION TREES FORMED AFTER FINISHING THE SBML TRAINING

Decision Types	Percent of Participants		Compare Learning Objective
	DM1	DM2	
Type 1	73% (40)	64% (35)	Correct
Type 2	27% (15)	16% (9)	Correct
Type 3.1	0%	5% (3)	Correct
Type 3.2	0%	5% (3)	Correct
Type 4	0%	2% (1)	In-complete
Type 5	0%	5% (3)	Incorrect
Type 6	0%	2% (1)	Incorrect

The decision tree types were formed from data matrices DM1 and DM2. The numbers in brackets represent the participant count for each tree type.

Table VI, column data matrix 1 (DM1) shows the intended decision tree that was taught for a muster drill situation (denoted as Type 1). Seventy-three percent of participants achieved this type of decision tree before the test scenario (T2). The remaining 27% of participants also formed their route selection based on the PA announcement, but when the

PA announcement provided no route information, they relied on their intended end location, which was dictated by the alarm type. For these participants, the muster station meant taking the primary route and the lifeboat station meant taking their secondary route (denoted as Type 2). The formation of both decision trees (Type 1 and 2) after module 2 demonstrated that all participants achieved the intended learning objectives and were adequately prepared to respond to the muster drill (T2).

2) Assess Emergency Situation Decision Tree (From Data Matrices 1 and 2): Building on earlier learning objectives, module 4 trained participants in how to assess the emergency situation, avoid hazards, and follow the safest egress path to the designated muster or lifeboat station. In an emergency, if personnel encountered an obstructed route, they were required to reroute in response to the hazardous situation. A variety of decision trees were developed after module 4.

Table VI, column data matrix 2 (DM2) shows that there were six different strategies used by participants at the end of training. Sixty-four percent of participants continued to use the same decision tree in which they selected an egress route based on information from the PA announcement (Type 1). Sixteen percent of participants continued to use the strategy in which the end location (dictated by the alarm type) indicated the route choice in absence of a PA announcement (Type 2). Ten percent of participants followed the alarm type and PA announcement (Type 3). If no clear route direction was provided over the PA announcement, the participants would link the alarm type to an egress route. For example, if the GPA or PAPA alarm sounded, the participants would take the primary route. However, in the event of no alarm, they would take their secondary route. The remaining 10% of participants demonstrated more varied behaviors. In these cases, when the PA did not provide a route direction, some individuals put emphasis on different attributes to make their decision. One participant's data (representing 2%) formed a correct but incomplete decision tree where the route decision was based solely on whether the route was obstructed or not (Type 4). The remaining four participants' data (representing 8%) formed incorrect decision trees that wrongly considered the presence of hazards (Type 5) and the previous route taken (Type 6).

When comparing the decision tree variations with the learning objectives, some weaknesses in the training and the participants were identified. The formation of an incomplete decision tree (e.g., Type 4) suggests that this participant required more targeted scenarios to focus on the missing decision rules (e.g., additional practice for situations to create the intended PA decision rules). The incorrect decision trees (e.g., Types 5 and 6) show that some participants (7%) require additional practice opportunities and feedback to ensure they reach the intended competence. If incorrect trees persist, then it is possible the participants are not suitable for virtual training or are not taking the training seriously (e.g., Type 6 where the participant's decision involved their previous route taken).

3) In-Depth Decision Tree Analysis of SBML Training: The decision tree analysis revealed information about the participants' performance that would otherwise not be apparent when looking solely at the performance metrics of the learning objectives. Specifically, the diagnostic capabilities of decision trees were used to identify the strengths and weaknesses of participants' decision-making strategies. Looking at the benefits of the SBML training, the majority of participants' decision trees matched the intended learning objectives (100% for data matrix 1 and 90% for data matrix 2). These participants, whose data formed decision tree types 1, 2, and 3, demonstrated the decision-making skills taught by the SBML program. They were able to identify attributes that were critical to success and come up with strategies that led to safe egress.

The decision trees also provided indications of deficiencies in the SBML training. Such as, the need to provide participants with sufficient practice to establish robust rerouting strategies and the participants over reliance on PA announcements during emergencies. For example, three decision trees (Types 1, 2, and 3) revealed that the participants' egress strategies centered on the PA announcement. In the absence of an announcement, some participants focused their attention on a variety of different attributes (e.g., presence of hazards), which were useful in terms of making effective egress decisions. However, the formation of decision trees due to missing or unclear PA announcements provides valuable information on whether the decision-making skills taught during the training were sufficient for all emergencies. These are areas that could be improved in future iterations of the training. Adaptive training could recognize these deficiencies in real time and provide additional training scenarios that focus on teaching participants what to do if there is no PA announcement or instructions on what is happening during the emergency.

C. Comparison of SBML and LBT Trees

For this pilot study, the decision tree results from both training experiments were compared directly as another lens to observe the overall training efficacy. The decision trees modeled from the SBML training data are summarized in Table VII. The decision trees modeled from the LBT data are summarized in Table VIII.

Comparing the resulting decision trees generated from the SBML and LBT data showed that the different training methods influenced the participants' egress strategies. Over the course of the SBML training, the SBML-trained participants' behaviors in responding to emergencies gradually converged to a few expected decision trees (except for a few participants). Ninety percent of SBML-trained participants achieved the intended learning objectives as demonstrated by the decision trees (Types 1, 2, and 3). Only 10% of SBML-trained participants displayed varied behaviors that could be addressed with targeted training.

Conversely, the decision trees of the LBT participants' behaviors for the emergency response scenarios diverged. Only 29% of LBT-trained participants achieved the intended learning objectives as demonstrated by the decision trees (Type 1). Many of the remaining LBT participants had a poor understanding of the egress procedures and were not compliant. Thirty-five percent of the LBT participants' data presented

TABLE VII
RESULTING SBML DECISION TREES FOR ALL 55 PARTICIPANTS AFTER FINISHING TRAINING MODULES 2 AND 4
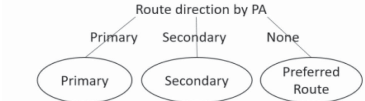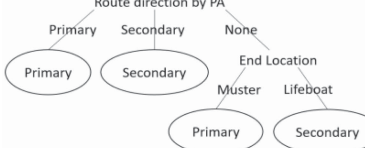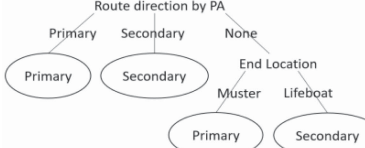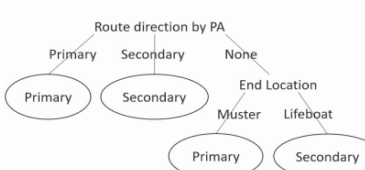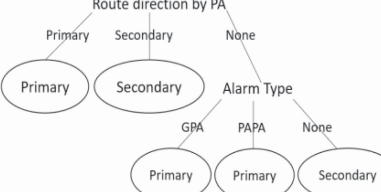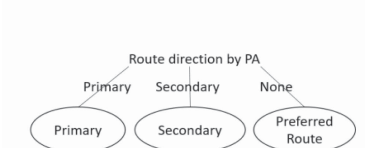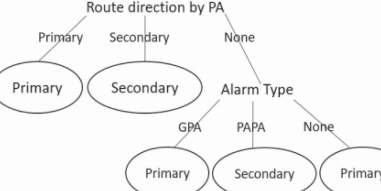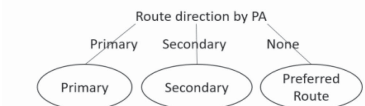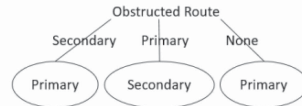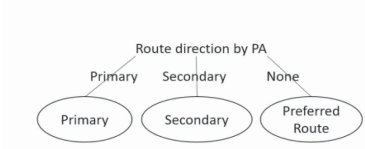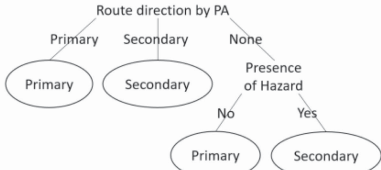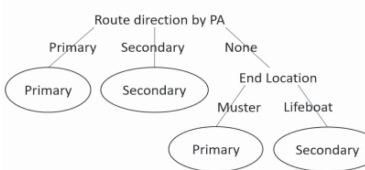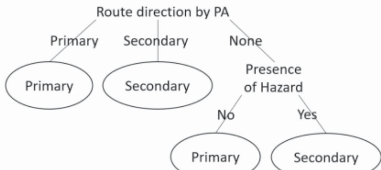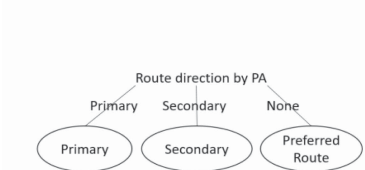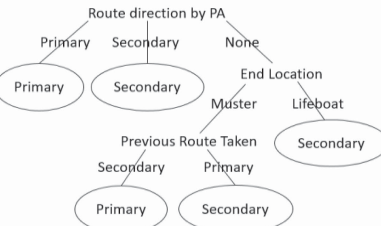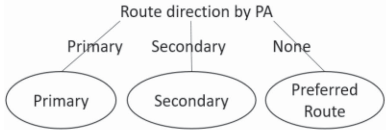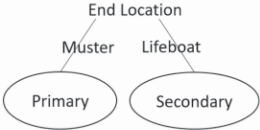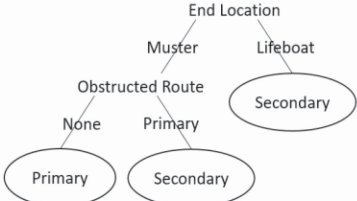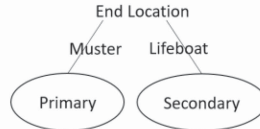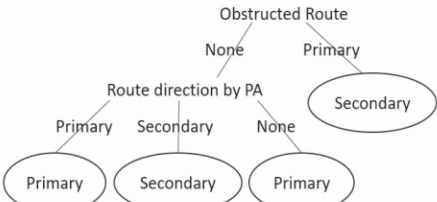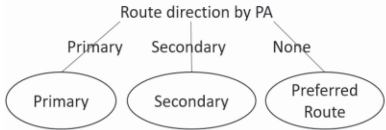
% Participants	Decision rules from Data Matrix 1 (until test scenario T2)	Decision rules from Data Matrix 2 (until test scenario T4)
64% (35 participants)		Remained the same.
16% (9 participants)		Remained the same.
5% (3 participants)		
4% (2 participants)		
2% (1 participants)		
2% (1 participants)		
4% (2 participants)		
2% (1 participants)		
2% (1 participants)		

TABLE VIII
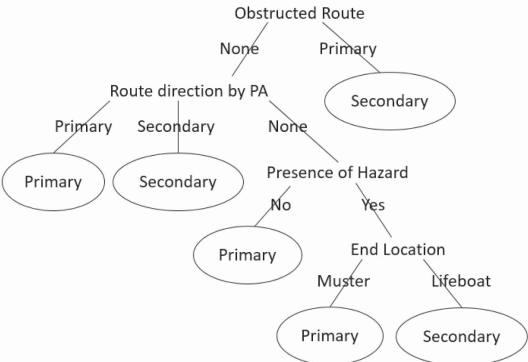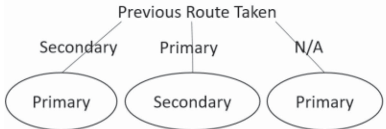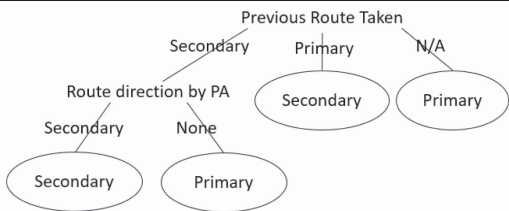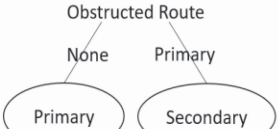RESULTING LBT DECISION TREES FOR ALL 17 PARTICIPANTS FOR TEST SCENARIOS T2 AND T4 [5]

% Participants	Decision rules from Data Matrix 1 (until test scenario T2)	Decision rules from Data Matrix 2 (until test scenario T4)
*29% (5 participants)	**Route direction by PA** — Primary → Primary; Secondary → Secondary; None → Preferred Route	Remained the same. *One participant had a slightly different tree for part 2. Depending on whether the participant understood the PA, they either followed the same tree or made a choice based on the obstructed route. If no route was obstructed, they would take the primary route. If the primary route was obstructed then they would take the secondary route.
17% (3 participants)	**End Location** — Muster → Primary; Lifeboat → Secondary	**End Location** — Muster → [Obstructed Route — None → Primary; Primary → Secondary]; Lifeboat → Secondary
6% (1 participant)	**End Location** — Muster → Primary; Lifeboat → Secondary	**Obstructed Route** — None → [Route direction by PA — Primary → Primary; Secondary → Secondary; None → Primary]; Primary → Secondary
6% (1 participant)	**Route direction by PA** — Primary → Primary; Secondary → Secondary; None → Preferred Route	**Obstructed Route** — None → [Route direction by PA — Primary → Primary; Secondary → Secondary; None → [Presence of Hazard — No → Primary; Yes → [End Location — Muster → Primary; Lifeboat → Secondary]]]; Primary → Secondary
6% (1 participant)	**Previous Route Taken** — Secondary → Primary; Primary → Secondary; N/A → Primary	**Previous Route Taken** — Secondary → [Route direction by PA — Secondary → Secondary; None → Primary]; Primary → Secondary; N/A → Primary
6% (1 participant)	At any condition, the participant's choice of route was the primary route.	**Obstructed Route** — None → Primary; Primary → Secondary
6% (1 participant)	At any condition, the participant's choice of route was the secondary route.	Remained the same.
*24% (4 participants)	No behavioral pattern or strategy was identified. The participants' choice of route was random and as a result the decision tree could not provide any more generalization than the data matrix. *One participant had trouble with the controls to open/close doors and chose a route with fewer doors. This participant was excluded because the behaviour specific to AVERT and not realistic to real evacuations.	

decision tree strategies that included special conditions for PA announcements, alarm type, obstructed routes, and hazards. The decision trees for two participants (representing 12% of LBT-trained participants) showed how inflexible they were in their route choice. For example, their decision tree represented behaviors of taking the same route regardless of the emergency condition.

For 24% of LBT-trained participants, the choice of route was random, and the decision trees could not provide any more generalization than the data matrix. Overall, the LBT participants' decision trees weighted attributes of the scenario that were not useful for making effective egress decisions [5]. The variability and incorrect behaviors observed in the LBT decision trees show that this method of training was inadequate for preparing participants for emergency conditions.

The SBML approach resulted in better route selection strategies compared to the LBT approach. As shown in the previous section, the majority of the observed route strategies for the SBML-trained participants (representing 90%) led to the successful completion of the test emergency scenario. Conversely, the majority of LBT-trained participants (representing 71%), displayed incomplete or incorrect decision trees. Therefore, the SBML training resulted in higher safety compliance and more concise decision trees than the LBT training. This indicates that participants from SBML training were generally better equipped for managing emergency scenarios.

V. CONCLUSION

In this pilot study, decision tree modeling was used to evaluate the training efficacy of two pedagogical approaches, SBML and LBT, which were assessed in the context of training naïve personnel for basic emergency duties. The article presents the proof-of-concept of using decision tree modeling as an alternative evaluation method to compare with conventional experimental performance outcomes.

In terms of measured performance, the SBML pedagogical approach was clearly superior to the alternative LBT approach. The comparison of performance metrics in both training experiments indicated that the SBML-trained participants performed better than the LBT-trained participants did; however, the performance metrics did not offer information as to why one group outperformed the other.

As a complement to conventional performance outcomes, the decision tree modeling provided a comprehensive analysis of the participants' route performance. The decision trees generated by the participants' data in both training experiments provided an explanation as to how the route selection performance in the SBML and LBT trained groups differed. The decision trees showed that when selecting egress routes in virtual emergencies, the decision-making strategies of the SBML-trained participants were more consistent with the intended learning objectives and represented safer behaviors than the decision tree strategies of LBT-trained participants.

This pilot study demonstrated the diagnostic capabilities of decision trees as training assessment tools. In both training cases, the decision trees provided a convenient visual representation of the individual strategies employed by participants. As illustrated in this work, the visual simplicity of the decision trees can be useful for identifying systemic deficiencies in training (and even in how procedures are designed). This is a useful feature for instructional designers.

Decision trees can also be used to diagnose the strengths and weaknesses of individual trainees, a capability that has additional value in terms of adapting the virtual training to meet the needs of individuals. This adaptive training potential could be realized by coupling the SBML approach with a built-in decision tree diagnostic tool in the virtual training such that each learner's performance can be automatically tracked and assessed in real-time, thereby providing the data required by a built-in decision tree diagnostic tool. The diagnostic tool would compare the individual learner's data-informed decision trees at various stages in the training with the intended learning objectives and suggest adaptive training scenarios for the learner to perform to assist them in achieving the intended decision strategies. For this to work in practice, the training scenarios must be carefully designed, as they are, in effect, experiment conditions for the diagnostic decision trees. Additional training scenarios would also be required to provide sufficiently specific pathways for adaptive training.

Finally, the decision trees were shown to have the considerable predictive capability. This feature could also be useful to instructional designers in terms of evaluating pedagogical strategies, such as determining when trainees are likely to be sufficiently capable of responding to a wide variety of emergencies, without necessarily training them for all potential eventualities.

VI. LIMITATIONS AND RECOMMENDATIONS

The pilot study had limitations that impacted the results. Specifically, the pilot employed a small sample size that was not representative of the offshore and maritime population. The choice of student participants with no prior offshore experience and no exposure to the simulator was intentional to measure the learning of basic emergency duties of naïve personnel. However, the small dataset collected from mostly student participants, rather than members from the offshore and maritime workforce limits the generalizability of the findings.

Further, the pilot study used small datasets to generate individual decision trees for each participant. Human performance data can be logistically difficult to collect and often results in small individual datasets. As such, this article applied a decision tree algorithm to an available sample of human performance data. While the datasets were small, there was valuable information gained from using human performance data in this proof-of-concept pilot to investigate the utility of decision trees for modeling learning.

To address these limitations, future full-scale empirical studies should increase the sample size and recruit personnel working in the offshore and maritime domains. Additionally, future full-scale modeling and testing of decision trees should involve a larger source of human performance data. Both limitations could be achieved by testing the virtual training technology with a large group of offshore personnel receiving basic safety training at an onshore facility prior to the individuals boarding offshore platforms and maritime vessels.

ACKNOWLEDGMENTS

REFERENCES

[1] Marine Casualties Investigative Body, "Cruise ship Costa Concordia marine casualty on January 13, 2012: Report on the safety technical investigation," Italian Ministry of Infrastructures and Transports, Italy, May 2013.

[2] IMO, "Resolution MSC.350(92) amendments to the international convention for the safety of life at sea (SOLAS), chapter III, part b, regulation 19 emergency training and drills," International Maritime Organization, London, U.K., Jun. 2013.

[3] J. Smith, B. Veitch, and S. MacKinnon, "Achieving competence in offshore emergency egress using virtual environment training," in *Proc. ASME 34th Int. Conf. Ocean Offshore Arctic Eng. Struct., Saf. Rel.*, 2015, pp. 1–8, doi: 10.1115/OMAE2015-41132.

[4] J. Smith and B. Veitch, "A better way to train personnel to be safe in emergencies," *ASCE-ASME J. Risk Uncertainty Eng. Syst. Part B, Mech. Eng.*, vol. 5, no. 1, Mar. 2019, Art. no. 011003, doi: 10.1115/1.4040660.

[5] M. Musharraf, J. Smith, F. Khan, and B. Veitch, "Identifying route selection strategies in offshore emergency situations using decision trees," *Rel. Eng. Syst. Saf.*, vol. 194, Feb. 2020, Art. no. 106179, doi: 10.1016/j.ress.2018.06.007.

[6] S. S. Wingfield and G. S. Black, "Active versus passive course designs: The impact on student outcomes," *J. Educ. Bus.*, vol. 81, no. 2, pp. 119–126, Nov./Dec. 2005, doi: 10.3200/JOEB.81.2.119-128.

[7] J. H. Barsuk, E. R. Cohen, W. C. McGaghie, and D. B. Wayne, "Long-term retention of central venous catheter insertion skills after simulation-based mastery learning," *Acad. Med.*, vol. 85, no. 10, pp. S9–S12, Oct. 2010, doi: 10.1097/ACM.0b013e3181ed436c.

[8] E. R. Cohen et al., "Making july safer: Simulation-based mastery learning during intern boot camp," *Acad. Med.*, vol. 88, no. 2, pp. 233–239, Feb. 2013, doi: 10.1097/ACM.0b013e31827bfc0a.

[9] F. Moazed et al., "Retention of critical care skills after simulation-based mastery learning," *J. Graduate Med. Educ.*, vol. 5, no. 3, pp. 458–463, Sep. 2013, doi: 10.4300/JGME-D-13-00033.1.

[10] W. C. McGaghie, S. B. Issenberg, J. H. Barsuk, and D. B. Wayne, "A critical review of simulation-based mastery learning with translational outcomes," *Med. Educ.*, vol. 48, no. 4, pp. 375–385, Apr. 2014, doi: 10.1111/medu.12391.

[11] J. H. Barsuk, E. R. Cohen, D. B. Wayne, V. J. Siddall, and W. C. McGaghie, "Developing a simulation-based mastery learning curriculum: Learning from 11 years of advanced cardiac life support," *Simul. Healthcare*, vol. 11, no. 1, pp. 52–59, Feb. 2016, doi: 10.1097/SIH.0000000000000120.

[12] W. C. McGaghie, S. B. Issenberg, E. R. Petrusa, and R. J. Scalese, "Effect of practice on standardised learning outcomes in simulation-based medical education," *Med. Educ.*, vol. 40, no. 8, pp. 792–797, Aug. 2006, doi: 10.1111/j.1365-2929.2006.02528.x.

[13] S. Griswold-Theodorson, S. Ponnuru, C. Dong, D. Szyld, T. Reed, and W. C. McGaghie, "Beyond the simulation laboratory: A realist synthesis review of clinical outcomes of simulation-based mastery learning," *Acad. Med.*, vol. 90, no. 11, pp. 1553–1560, Nov. 2015, doi: 10.1097/ACM.0000000000000938.

[14] D. A. Cook, R. Brydges, B. Zendejas, S. J. Hamstra, and R. Hatala, "Mastery learning for health professionals using technology-enhanced simulation: A systematic review and meta-analysis," *Acad. Med.*, vol. 88, no. 8, pp. 1178–1186, Aug. 2013, doi: 10.1097/ACM.0b013e31829a365d.

[15] B. S. Bloom, "Mastery learning," in *Mastery Learning: Theory and Practice*, J. H. Block, Ed. New York, NY, USA: Holt, Rinehart and Winston, 1971, pp. 47–63.

[16] T. R. Gusky, "Closing achievement gaps: Revisiting Benjamin S. Bloom's 'learning for mastery'," *J. Adv. Acad.*, vol. 19, no. 1, pp. 8–31, Nov. 2007, doi: 10.4219/jaa-2007-704.

[17] H. Aldowah, H. Al-Samarraie, and W. Fauzy, "Educational data mining and learning analytics for 21st century higher education: A review and synthesis," *Telematics Inform.*, vol. 37, pp. 13–49, Apr. 2019, doi: 10.1016/j.tele.2019.01.007.

[18] T. Kabudi, I. Pappas, and D. H. Olsen, "AI-enabled adaptive learning systems: A systematic mapping of the literature," *Comput. Educ. Artif. Intel.*, vol. 2, Mar. 2021, Art. no. 100017, doi: 10.1016/j.caeai.2021.100017.

[19] Y. Liu, "Data mining and its applications in digital human modeling," in *Handbook of Digital Human Modeling: Research for Applied Ergonomics and Human Factors Engineering*, V. G. Duffy, Ed. Boca Raton, FL, USA: CRC Press, 2009.

[20] M. Badino, "An application of information theory to the problem of the scientific experiment," *Synthese*, vol. 140, no. 3, pp. 355–389, Jun. 2004.

[21] P. C. Cacciabue, F. Decortis, B. Drozdowicz, M. Masson, and J. P. Nordvik, "COSIMO: A cognitive simulation model of human decision making and behavior in accident management of complex plants," *IEEE Trans. Syst. Man, Cybern.*, vol. 22, no. 5, pp. 1058–1074, Sep./Oct. 1992, doi: 10.1109/21.179844.

[22] J. Han, M. Kamber, and J. Pei, "Classification: Basic concepts," in *Data Mining: Concepts and Techniques*, 3rd ed. Waltham, MA, USA: Morgan Kaufmann, 2012, pp. 327–391, doi: 10.1016/B978-0-12-381479-1.00008-3.

[23] A. W. H. House, J. Smith, S. MacKinnon, and B. Veitch, "Interactive simulation for training offshore workers," in *Proc. MTS/IEEE OCEANS Conf. Exp.*, 2014, pp. 1–6, doi: 10.1109/OCEANS.2014.7003211.

[24] S. B. Kotsiantis, "Supervised machine learning: A review of classification techniques," *Informatica*, vol. 31, pp. 249–263, Jul. 2007.

[25] J. Brownlee, "Model evaluation," in *Imbalanced Classification with Python: Choose Better Metrics, Balance Skewed Classes, and Apply Cost-Sensitive Learning*, Self-published, 2021, pp. 38–46.

[26] W. R. Shadish, T. D. Cook, and D. T. Campbell, "Experiments and generalized causal inference," in *Experimental and Quasi-Experimental Designs For Generalized Causal Inference*. Boston, MA, USA: Houghton Mifflin, 2002.

[27] G. Corder and D. Foreman, "Tests for nominal scale data: Chi-square and fisher exact tests," in *Nonparametric Statistics: A Step-By-Step Approach*, 2nd ed. New Jersey, NY, USA: Wiley, 2014.

Jennifer Smith received the B.S. degree in applied science from the Acadia University, Wolfville, NS, Canada, in 2002, the B.Eng. degree in biological engineering from the Dalhousie University, Halifax, NS, Canada, in 2008, and the M.Eng. and Ph.D degrees in ocean and naval architectural engineering from the Memorial University of Newfoundland (MUN), St. John's, NL, Canada, in 2015 and 2020, respectively.

From 2008 to 2020, she was a research coordinator on several projects with MUN's Faculty of Engineering and Applied Science. She is currently working as an instructor (research) with the School of Maritime Studies, MUN's Marine Institute. Her research interests include pedagogical and human factors issues in marine and offshore simulation training.

Mashrura Musharraf received the B.Eng. degree in computer science and engineering from the Bangladesh University of Engineering and Technology, Dhaka, Bangladesh, in 2009, and the M.Eng. and Ph.D. degrees in computer engineering from the Memorial University of Newfoundland (MUN), St. John's, NL, Canada, in 2013 and 2018, respectively.

From 2018 to 2020, she was a postdoctoral research fellow with the Safety at Sea project and a per-course instructor with the Faculty of Engineering and Applied Science, MUN. Since 2020, she has been an Assistant Professor of marine technology with Aalto University, Espoo, Finland. Her research interests include data mining, machine learning, and artificial intelligence for building and deploying human-centered systems and solutions to create a safer marine industry.

Brian Veitch received the B.Eng. degree in naval architectural engineering and the M.Eng. degree in ocean engineering from the Memorial University of Newfoundland (MUN), St. John's, NL, Canada, in 1988 and 1990, respectively, and the Ph.D. degree in mechanical engineering from the Helsinki University of Technology (now part of Aalto University), Espoo, Finland, in 1995.

Since 1998, he has been teaching ocean and naval architectural engineering with the Faculty of Engineering and Applied Science, MUN.

Dr. Veitch is also the Natural Sciences and Engineering Research Council of Canada's Husky Energy Industrial Research Chair in Safety at Sea.

Faisal Khan received the B.Eng. degree in chemical engineering from the Aligarh Muslim University, Aligarh, India, in 1992, the M.Eng. degree in computer-aided process plant design from the Indian Institute of Technology, Roorkee, India, in 1994, and the Ph.D. degree in computer-aided risk analysis from the Pondicherry University, Pondicherry, India, in 1998.

From 2019 to 2021, he was a Professor and the Associate Dean (graduate studies) with the Faculty of Engineering and Applied Science, MUN and the Canada Research Chair (Tier I) of Offshore Safety and Risk Engineering. From 2008 to 2019, he was the Discipline Chair and Head of process engineering and founder of the Centre for Risk Integrity and Safety Engineering (C-RISE), MUN. Since 2021, he has been a Professor of chemical engineering with Texas A&M University, College Station, TX, USA. He is also the Mike O'Connor II Chair and Director with the Mary Kay O'Connor Process Safety Center, Texas A&M University.