Journal of *Marine Science and Engineering*

MDPI

Article

Multimodal Data-Driven Intelligent Assessment of Man-Overboard Rescue Operations in a Navigation Simulator

Huibin Wang [1], Helong Shen [1,*], Yiqing Gu [2], Xiang Fan [2] and Xiangtao Ding [3]

[1] Navigation College, Dalian Maritime University, Dalian 116026, China; whb_1068@dlmu.edu.cn
[2] Shanghai Merchant Ship Design & Research Institute, Shanghai 201203, China
[3] Rizhao Pilot Station, Rizhao 276827, China
* Correspondence: shenhelong@dlmu.edu.cn

Abstract

Conventional assessment of man-overboard rescue training in navigation simulators relies heavily on manual observation, which introduces subjectivity and delays feedback. We developed a multimodal data-driven intelligent assessment system to address these limitations. The system acquires structured simulator outputs and scores rescue operations in real time through data cleaning, speech recognition, video-based behaviour detection, time synchronisation, feature extraction, and multimodal data fusion. We evaluated the system using simulator data from different recovery manoeuvres and modality-availability conditions. The results showed that the system identified critical actions and task-completion states, while distinguishing standard operations, manoeuvring deviations, and missing modalities. Experimental validation and agreement analysis against expert ratings supported the effectiveness and reliability of the proposed method. The system provides objective, quantitative, and traceable evidence for man-overboard rescue training, BRM debriefing, and seafarer competence assessment.

Keywords: man overboard; navigation simulator; multimodal data; intelligent assessment; speech recognition; video-based behaviour detection; bridge resource management training

1. Introduction

Man-overboard (MOB) rescue requires the bridge team to raise the alarm, issue orders, manoeuvre the vessel and coordinate personnel under time pressure [1,2]. Bridge resource management (BRM) training relies largely on instructor observation and debriefing, and grades can vary with the assessor and the criteria applied [3,4]. Behavioural observation, navigational data analysis and automated scoring have been used across simulator tasks [5–11].

For collision avoidance, Juszkiewicz and Żukowska [10] weighted deviations in passing distance and radar use. Their system broadly agreed with 13 experts on low-scoring solutions, but the rankings of the best solutions differed. For MOB recovery, Wang and Yin [12] used simulator-derived return distance, minimum vessel–casualty separation and speed at closest approach to grade manoeuvres using a backpropagation neural network. Audio-visual event localisation uses temporal and semantic correspondence [13], while uncertainty-aware fusion adjusts modality contributions according to data and model uncertainty [14]. Maritime simulator research has examined team interaction and feedback [15,16], synchronised eye movements with control records [17], and assessed large-vessel handling and port entry and departure [18,19]. In real-time MOB assessment, an additional challenge is

Check for updates

Academic Editors: Zhengrong Xiang and Xiaofei Yang

Received: 27 October 2026
Revised: 20 November 2026
Accepted: 24 November 2026
Published: 29 November 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

to link a spoken order to the corresponding control action. Device-specific delays and clock drift can shift the apparent timing of events across data streams [20]. For a hard-over order, such misalignment can affect both the identification of the corresponding rudder command and the estimated response time.

Conventional simulator assessment relies primarily on instructor observation and post-exercise debriefing, making it difficult to capture concurrent vessel manoeuvres, bridge communications and crew behaviours comprehensively. In this study, we developed a multimodal framework for assessing man-overboard rescue operations that aligns navigation, speech and video data using timestamps. The framework links vessel states, spoken orders and behavioural observations to provide complementary assessment evidence. It applies hierarchical scoring to 11 indicators across three stages: immediate action, emergency procedure initiation and subsequent rescue action. Assessment scores are traceable to specific events, supporting instructor review, targeted feedback and debriefing. We examined scoring performance across cases involving different recovery manoeuvres and missing modalities and compared the system-assigned grades with expert assessments.

2. Man-Overboard Rescue Manoeuvres

Method selection and execution are scored separately. The scenario defaults to an Anderson turn at visibility of at least 2 NM and a Williamson turn below 2 NM. An instructor-specified task takes precedence. The Williamson assessment examines counter-rudder, return to amidships, reciprocal heading and recovery of the original track. The Anderson assessment examines sustained turning in one direction, rudder easing, deceleration and the approach path. Both include final approach speed.

2.1. Williamson Turn

The standard Williamson turn begins with an immediate hard-over rudder command towards the side of the casualty. When the heading has changed by approximately 60° from the initial course, the rudder is put hard over to the opposite side. The rudder is returned to amidships when the vessel is approximately 20° short of the reciprocal course. After stabilising on the reciprocal heading, the vessel follows the original track in the opposite direction and gradually slows when approaching the casualty [21]. The workflow is shown in Figure 1.

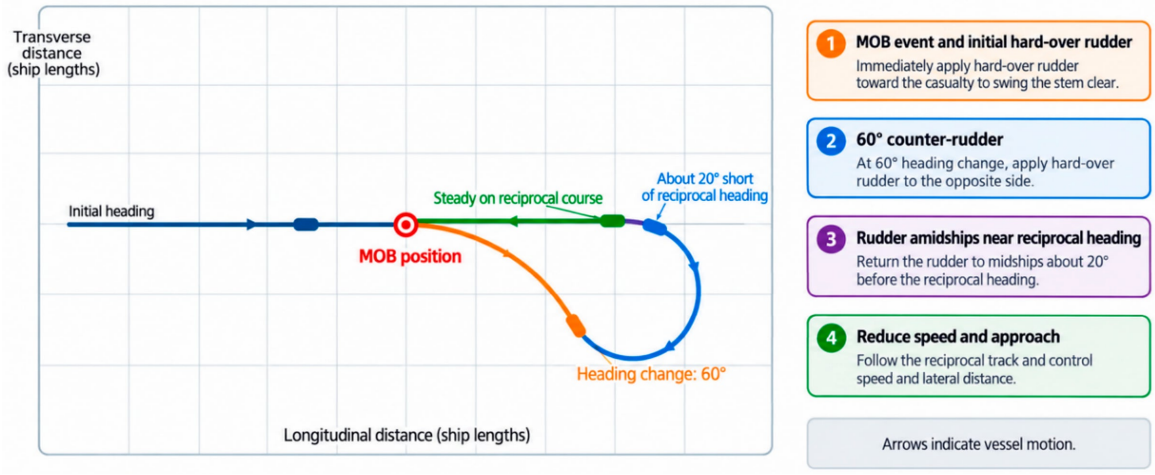


Figure 1. Schematic of the Williamson turn manoeuvre.

2.2. Anderson Turn

The Anderson turn is suitable when the casualty remains continuously visible and a rapid return is required [21]. The rudder is immediately put hard over towards the casualty and held to maintain a one-sided turn. After the vessel has turned approximately 270° and its bow approaches the casualty's bearing, the rudder is returned to amidships. The return heading is then corrected according to the casualty's relative bearing. During the final approach, speed is reduced and vessel position is adjusted for wind and current to support safe recovery. The workflow is shown in Figure 2.

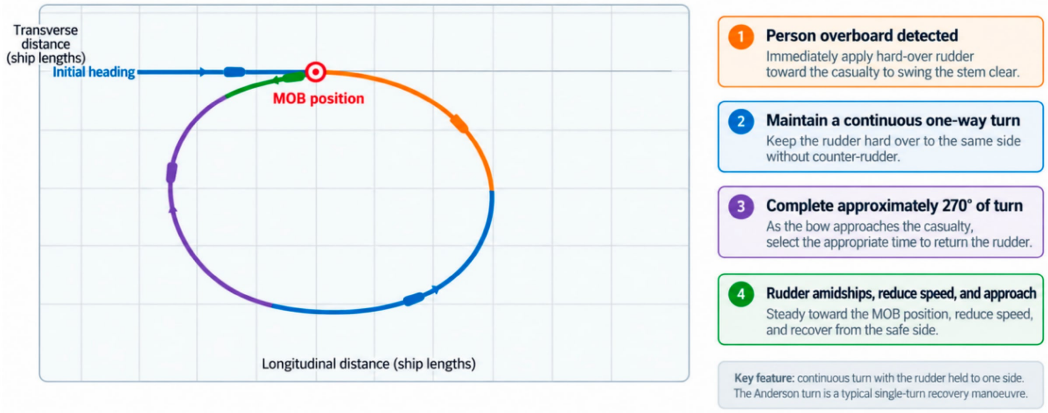


Figure 2. Schematic of the Anderson turn manoeuvre.

3. Multimodal Data-Driven Intelligent Assessment Model

The system uses a modular architecture that coordinates back-end and front-end functions across three modules: multimodal data processing, BRM assessment, and result interaction. The back end receives data from the navigation simulator, electronic charts, bridge audio, video, and operational logs. It performs quality control, time synchronisation, feature extraction, and event fusion to generate standardised event sequences. The assessment module uses event-triggered stage assessment to organise MOB evidence and calculates indicator-level and overall scores. These calculations apply assessment rules, indicator weights, and membership functions. The front end supports status monitoring, result visualisation, and report generation, while linking results to source data for traceability. The workflow is shown in Figure 3.

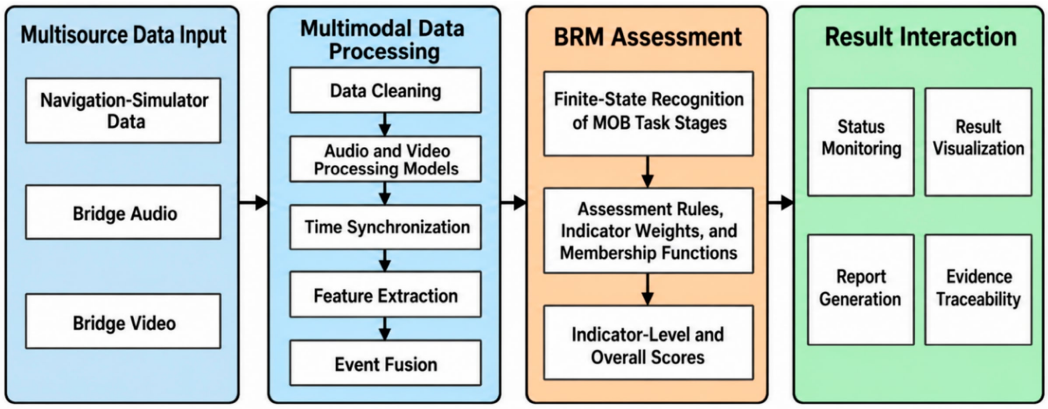


Figure 3. Overall architecture of the multimodal data-driven intelligent assessment system for man-overboard rescue.

3.1. Multimodal Data-Processing Model

3.1.1. Data Cleaning

Data cleaning was performed between multisource data input and the audio-video processing modules. It removed duplicate, missing, anomalous, and invalid records generated during data acquisition, network transmission, and interface parsing. The inputs included structured navigation-simulator data, bridge audio, and bridge video. Processing comprised format validation, duplicate removal, anomaly detection, missing-value handling, and quality labelling. Table 1 summarises the checks and the proportion of records affected.

Table 1. Multimodal data quality checks and affected records.

Stream	Check	Criterion	Handling	Affected
Audio	Decoding	Decode failure, truncated segment or stream interruption	Flag interval and exclude from recognition	1.50%
Audio	Continuity	Compare duration with sample rate, sample count and timestamps	Log gaps and retain silence	0.91%
Video	Frame validity	Undecodable or corrupt frame	Exclude from recognition	3.51%
Video	Duplicates and gaps	Check frame ID, timestamp and frame rate	Retain one copy and omit evidence in gaps	1.31%
Simulator	Validity and duplicates	Missing, mistyped or nonfinite fields, or identical timestamp and data from the same source	Exclude invalid fields and retain one copy of each duplicate	0.90%
Simulator	Range	Speed −10 to 30 kn, rudder angle ±35.5°, heading mapped to [0°, 360°)	Flag excursions for review	1.01%
Simulator	Rate of change	Over approximately 1 s, speed changes by >1 kn/s or rudder angle by >10°/s	Review adjacent records and control logs	1.01%

3.1.2. Speech-Recognition Model

Whisper [22] converts 16 kHz mono audio to log-Mel spectra and produces timestamped transcripts with confidence estimates. Faster-Whisper v1.1.0 runs large-v3 with automatic Chinese/English language detection, beam size 3 and a no-speech threshold of 0.4. Log-probability thresholds are −0.8 without context and −1.0 with context. A minimum segment-average log probability below −0.5 triggers additional decoding and semantic reranking. WebRTC provides echo cancellation, noise suppression, high-pass filtering and automatic gain control. A bilingual vocabulary, scenario prompts and contextual and phonetic correction standardise maritime terms. Offline NeMo diarisation [23] uses MarbleNet voice activity detection and TitaNet-Large embeddings, with clustering and sigmoid thresholds of 0.25 and 0.7. Rules applied to aligned speaker labels and transcripts generate events with speaker, content, time, type, confidence and closed-loop confirmation. The workflow is shown in Figure 4.

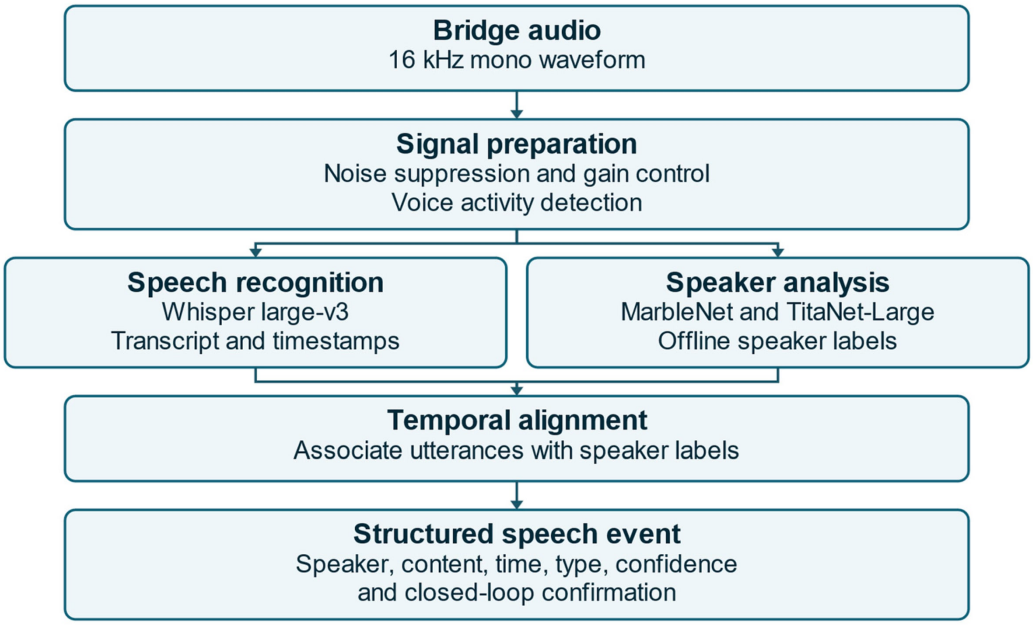


Figure 4. Workflow for Whisper and NeMo speech-data processing.

Across six manually annotated bridge recordings, 428 errors occurred in 2429 reference characters (character error rate, 17.62%). Speaker labels matched the reference in 152 of 160 segments (95.00%), and 110 of 123 command events were identified (89.55%).

3.1.3. Video-Based Behaviour-Recognition Model

YOLOv8 detects bridge personnel [24]. ByteTrack associates detections across frames and uses low-confidence detections to recover interrupted tracks [25]. MediaPipe estimates head, shoulder, elbow, wrist and trunk landmarks [26]. For each track, a sliding window combines duration, valid-frame count and proportion, track stability and pose confidence with the position of landmarks relative to equipment. Outputs include the person ID, behaviour, start and end times, duration, confidence and supporting frame indices. The workflow is shown in Figure 5.

Predictions agreed with manual image review at 121 of 170 observations (71.18% accuracy, macro F1 66.74%; Table A5). ECDIS observation had the highest recall (89.23%) and console operation the highest precision (82.35%). Steering F1 was 66.67%; five steering observations were classified as console operation. Ten of 29 observations with an indeterminate screen target were classified as ECDIS observation. The timestamped detections locate segments for verification against bridge orders and equipment records.

3.1.4. Time-Synchronisation Model

Navigation frames provide the time base. Online, audio and video observations are assigned to the first navigation frame processed after receipt. Its timestamp marks the speech onset or video observation for fusion, while original speech boundaries are retained. Observations queued for more than 15 s on the host monotonic clock are discarded. Offline, video segment onset is corrected using Equation (1) and matched to the nearest navigation frame using Equation (2).

$$t'_i = t_i + \Delta t \quad (1)$$

Here, t_i and t'_i are the original and corrected segment onset, respectively, and Δt is the configured fixed offset (default 0 s).

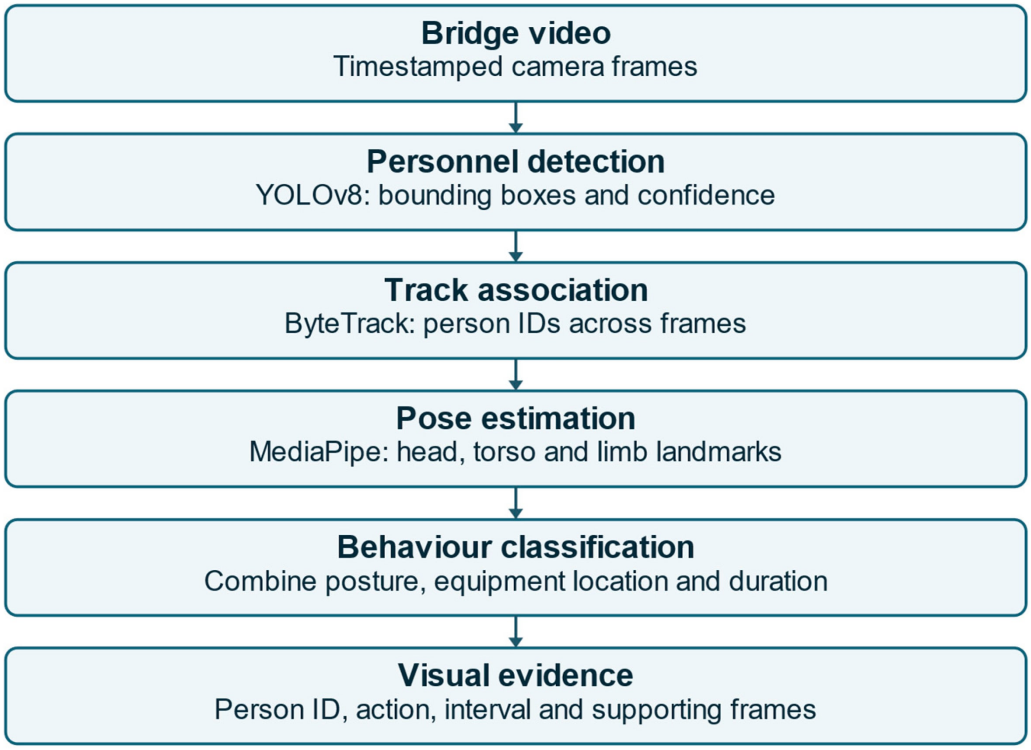


Figure 5. Workflow for video-based behaviour recognition.

The matched frame index is k^*, where t_k denotes navigation-frame time.

$$k^* = argmin_k |t'_i - t_k| \quad (2)$$

Offline matching imposes no maximum time difference. Device offsets and drift were not independently estimated, and residual alignment errors remain unmeasured.

For samples at t_1 and t_2, $\alpha = (t - t_1)/(t_2 - t_1)$. Scalars are interpolated linearly and heading and course over ground along the shortest arc.

$$x(t) = x_1 + \alpha(x_2 - x_1), \quad \theta(t) = (\theta_1 + \alpha\Delta\theta) \ mod \ 360° \quad (3)$$

The signed angular increment is $\Delta\theta = ((\theta_2 - \theta_1 + 180°) \bmod 360°) - 180°$. Interpolation from 359° to 1° therefore passes through 0°.

Helm, engine and voice commands remain discrete events. Between 10 and 90 s, a 1 s overestimate of response time reduces the stopping score by 1.25 points on a 100-point scale and the hard-over score by 0.75 points at fixed rudder angle. These are analytical sensitivities of the scoring functions.

3.1.5. Feature Extraction and Multimodal Data Fusion

Figure 6 shows the three fusion levels. At the feature level, frame k contains navigation and control records N_k, speech events A_k and visual observations V_k:

$$X_k = (N_k, A_k, V_k) \quad (4)$$

No common modality weights are assigned, and speech and video confidence values are not averaged. Visual events are deduplicated by evidence ID and type. For hard-over action, the system seeks the first absolute rudder order of at least 30°, then checks the vessel rudder field as a fallback. An order heard on audio is retained as evidence even when the threshold is unmet; the indicator then remains zero until qualifying telemetry arrives. The orderRudder field verifies command magnitude. It does not establish physical rudder

travel or directional agreement. Detected speech and behaviour events are linked to their source recordings and timestamps to support verification against simulator records and the operational context. Indicators supported solely by recognition outputs should be reviewed when evidence conflicts or when their scores could change the final competence grade. With response delay Δt in seconds, maximum absolute ordered angle δ_{max} in degrees and $C(x) = \min(1, \max(0, x))$, the score is

$$s_H = \begin{cases} 0.60\, C((90 - \Delta t)/80) + 0.40\, C((\delta_{\max} - 10)/20), & \textit{hard over detected} \\ 0, & \textit{otherwise} \end{cases} \quad (5)$$

At the decision level, the AHP stage weights are $W = (0.3579, 0.1951, 0.4470)$ [27]. For stage d and indicator j, w and s denote local weight and normalised score. The overall score is

$$S = 100(1 - B) \sum_d W_d \sum_j w_{dj} s_{dj} \quad (6)$$

B equals one when a critical veto is triggered and zero otherwise. For example, failure to detect either a stop or a hard-over response within 120 s triggers the immediate-action veto; an earlier exercise termination uses the evidence available at termination.

The MOB event establishes a common time origin for assessment. Immediate actions and emergency communications are evaluated independently from this point and may overlap. Table 2 summarises the activation and evidence rules for each stage.

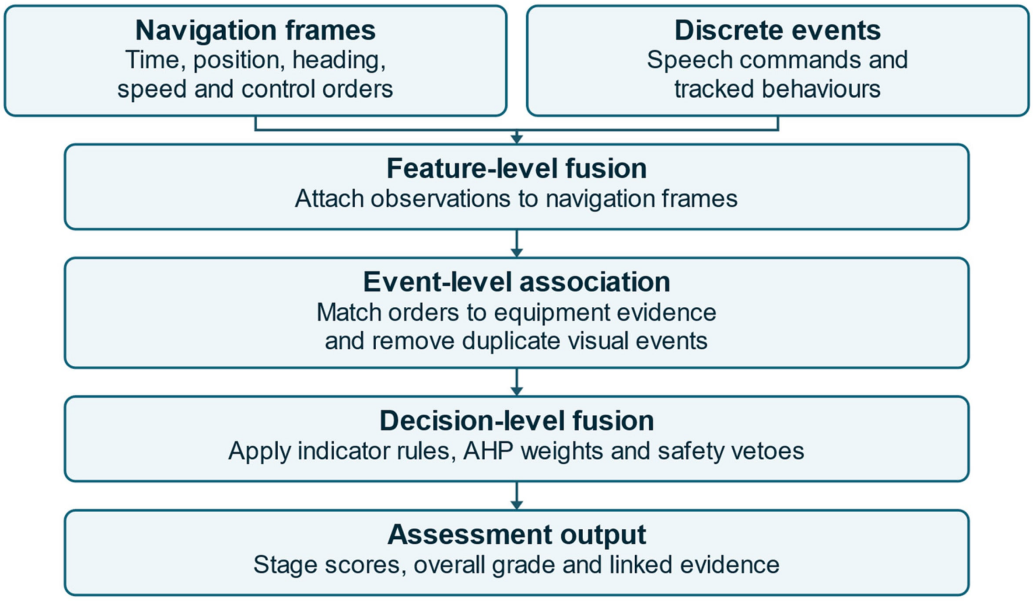


Figure 6. Workflow for feature extraction and hierarchical multimodal data fusion.

3.2. Assessment Model

The assessment followed the Assessment Standards for Seafarer Competence of the People's Republic of China (2024 edition) and MOB emergency-response requirements. The indicators were organised into three stages: immediate action, emergency procedure initiation, and subsequent rescue action. Navigation, speech, and video data characterised vessel manoeuvring, communication commands, and personnel behaviour, respectively. Each indicator was normalised to a dimensionless score in [0,1] and aggregated using hierarchical weights. Table 3 lists the indicators and their data sources.

Table 2. Activation and evidence rules for the three MOB assessment stages.

Assessment Stage	Activation and Evidence Rules
Immediate action	Assessment begins at the recorded MOB event time. If an explicit event is unavailable, the first appearance of the simulator MOB marker provides the time origin. The system checks for engine stop, hard-over rudder, lifebuoy deployment and casualty-position recording.
Emergency procedure initiation	From the same time origin, the system checks event and speech records for the MOB alarm, notification of the competent authority and a navigational warning. Each indicator is scored independently of completion of immediate actions.
Subsequent rescue action	Assessment begins when any one of the following is detected: a recovery manoeuvre, a return manoeuvre, a command to launch the rescue boat, or a record that the casualty has been recovered.

Table 3. Assessment indicators and data sources for man-overboard rescue.

Assessment Stage	Indicator	Primary Data Source
Immediate action	Stopping timeliness; hard-over rudder response and magnitude; lifebuoy deployment timeliness; MOB position recording	Simulator; event logs; ASR
Emergency procedure initiation	MOB alarm; report to the competent authority/VTS; navigational warning	ASR; simulator events
Subsequent rescue action	Recovery-manoeuvre quality; low-speed approach distance; rescue-boat launch order issued; monitoring of surrounding vessel traffic	Simulator; video; event logs; bridge speech (launch order)

3.2.1. Weight Determination

The questionnaire elicited pairwise comparisons among indicators at the same hierarchical level. The panel included experienced chief officers and masters, as well as maritime educators with at least eight years of teaching experience.

The analytic hierarchy process [27] was used to determine the weights of the MOB rescue assessment indicators. Experts made pairwise comparisons among indicators at the same level using Saaty's 1–9 scale, producing the following judgement matrix:

$$Q = (q_{ij})_{n\times n}, \quad q_{ii} = 1, \quad q_{ji} = \frac{1}{q_{ij}} \tag{7}$$

where Q is the judgement matrix; q_{ij} is the importance ratio of indicator i to indicator j; and n is the number of indicators at the same level.

Weights were calculated using the geometric-mean method. The row products, geometric means, and normalised weights of the judgement matrix were calculated as follows:

$$M_i = \prod_{j=1}^{n} q_{ij}, \quad i = 1, 2, \ldots, n \tag{8}$$

$$\overline{w_i} = \sqrt[n]{M_i}, \quad i = 1, 2, \ldots, n \tag{9}$$

$$w_i = \frac{\overline{w_i}}{\sum_{k=1}^{n} \overline{w_k}}, \quad i = 1, 2, \ldots, n \tag{10}$$

where M_i is the product of the elements in row i of the judgement matrix; $\overline{w_i}$ is the geometric mean for indicator i; and w_i is the normalised local weight. The maximum eigenvalue and consistency indices were calculated as follows:

$$\lambda_{max} = \frac{\sum_{i=1}^{n} \frac{(Qw)_i}{w_i}}{n} \quad (11)$$

$$CI = \frac{\lambda_{max} - n}{n-1}, \; CR = \frac{CI}{RI} \quad (12)$$

where λ_{max} is the maximum eigenvalue of the judgement matrix; CI is the consistency index; RI is the random consistency index; and CR is the consistency ratio. When $CR < 0.10$, the judgement matrix passes the consistency test; otherwise, the expert judgements must be revised.

Of 35 questionnaires distributed, 26 were valid (74.29%). Of the 26 respondents whose questionnaires were included in the AHP analysis, 10 were masters (38.5%), 12 were chief officers (46.2%) and four were maritime educators (15.4%). All four educators had at least eight years of teaching experience. All judgement matrices had CR < 0.10. Individual judgements were aggregated using the geometric mean. Table 4 gives the weights and the between-expert standard deviation of each local indicator weight. Appendix A gives the four aggregate matrices and their consistency ratios (Tables A1–A4).

Table 4. Weights of assessment indicators for the man-overboard rescue scenario.

Primary Indicator	Weight	Secondary Indicator	Weight	Between-Expert SD of Local Weight
Immediate action	0.3579	Stopping-response timeliness	0.3579 × 0.2500	0.0767
		Hard-over rudder response timeliness and magnitude	0.3579 × 0.2508	0.0560
		Lifebuoy-deployment timeliness	0.3579 × 0.2435	0.0647
		MOB position recording	0.3579 × 0.2556	0.0635
Emergency procedure initiation	0.1951	MOB-alarm timeliness	0.1951 × 0.4715	0.0569
		Timeliness of reporting to the competent authority	0.1951 × 0.2800	0.0803
		Issuance of a navigational warning	0.1951 × 0.2485	0.0761
Subsequent rescue action	0.4470	Recovery-manoeuvre quality	0.4470 × 0.3249	0.0588
		Minimum distance to the casualty	0.4470 × 0.3356	0.0761
		Rescue-boat launch order issued	0.4470 × 0.1738	0.0473
		Monitoring of surrounding vessel traffic	0.4470 × 0.1658	0.0404

In a leave-one-expert-out sensitivity analysis, the three stage weights were re-estimated 26 times, each time using 25 of the 26 valid questionnaires. Stage scores, grade boundaries and veto rules were held fixed. All nine case grades were unchanged, and the largest absolute difference from the full-panel overall score was 0.45 points.

3.2.2. Assessment Indicators and Scoring Functions

The 11 indicators in Table 5 assess immediate action, initiation of emergency procedures and subsequent rescue. Scores range from 0 to 1. The actions were drawn from the 2024 Chinese competency assessment specification for seafarers. Thresholds were initially based on BRM simulator experience and refined by experts for this exercise. IMO guidance recommends selecting the approach side in light of wind, sea and swell and providing shelter during final approach [28]. In this wind-dominated scenario, placing the casualty on the windward side during final approach triggers a veto and sets the overall score to zero.

Table 5. Assessment indicators and scoring functions for man-overboard rescue.

No.	Assessment Indicator	Scoring Function
1	Stopping-response timeliness	$S_s = \begin{cases} 1, & t_s \le 10 \\ \frac{90 - t_s}{80}, & 10 < t_s < 90 \\ 0, & t_s \ge 90 \end{cases}$ Here, t_s is the time from the MOB trigger until a stop command is registered at the engine telegraph. Without valid stopping evidence, the score is zero.
2	Hard-over rudder response timeliness and magnitude	$S_{rt} = \begin{cases} 1, & t_r \le 10 \\ \frac{90 - t_r}{80}, & 10 < t_r < 90 \\ 0, & t_r \ge 90 \end{cases}$ $S_{ra} = \begin{cases} 0, & \delta_{max} \le 10^\circ \\ \frac{\delta_{max} - 10}{20}, & 10^\circ < \delta_{max} < 30^\circ \\ 1, & \delta_{max} \ge 30^\circ \end{cases}$ Here, t_r is the time to the first hard-over response, and δ_{max} is the maximum absolute rudder angle in degrees. Without hard-over evidence, the indicator scores zero.
3	Lifebuoy-deployment timeliness	$S_{lb} = \begin{cases} 1, & t_{lb} \le 15 \\ \frac{120 - t_{lb}}{105}, & 15 < t_{lb} < 120 \\ 0, & t_{lb} \ge 120 \end{cases}$ Here, t_{lb} is the interval from the MOB trigger to the first valid evidence of lifebuoy deployment.
4	MOB position recording	$S_p = \begin{cases} 1, & E_p = 1 \\ 0, & E_p = 0 \end{cases}$ E_p equals one when valid MOB coordinates or an explicit position-recording event are detected and zero otherwise.
5	MOB-alarm timeliness	$S_a = \begin{cases} 1, & t_a \le 30 \\ \frac{300 - t_a}{270}, & 30 < t_a < 300 \\ 0, & t_a \ge 300 \end{cases}$ Here, t_a is the time from the MOB trigger to the first valid MOB alarm.

Table 5. *Cont.*

No.	Assessment Indicator	Scoring Function
6	Timeliness of reporting to the competent authority	$S_r = \begin{cases} 1, & t_{report} \le 300 \\ \frac{1800 - t_{report}}{1500}, & 300 < t_{report} < 1800 \\ 0, & t_{report} \ge 1800 \end{cases}$ Here, t_{report} is the interval from the MOB trigger until valid notification of VTS or the competent authority.
7	Issuance of a navigational warning	$S_w = \begin{cases} 1, & E_w = 1 \\ 0, & E_w = 0 \end{cases}$ E_w indicates whether a navigational warning or a request for other vessels to keep clear is detected.
8	Recovery-manoeuvre quality	Williamson turn $S_m = S_b \times P_m \times S_W$ $P_m = \min(1, \frac{\Theta_{Hdg}}{180^\circ})$ $s_1 = \begin{cases} 1, & t_r \le 10 \\ \frac{90 - t_r}{80}, & 10 < t_r < 90 \\ 0, & t_r \ge 90 \end{cases}$ $s_2 = \begin{cases} 0, & \delta_{max} \le 10^\circ \\ \frac{\delta_{max} - 10}{20}, & 10^\circ < \delta_{max} < 30^\circ \\ 1, & \delta_{max} \ge 30^\circ \end{cases}$ $s_3 = \max(0, 1 - \frac{\lvert\theta_{rev} - 60^\circ\rvert}{15^\circ})$ $s_4 = \begin{cases} 1, & e_3 \le 5^\circ \\ \frac{30^\circ - e_3}{25^\circ}, & 5^\circ < e_3 < 30^\circ \\ 0, & e_3 \ge 30^\circ \end{cases}$ $s_5 = \begin{cases} 1, & e_4 \le 5^\circ \\ \frac{30^\circ - e_4}{25^\circ}, & 5^\circ < e_4 < 30^\circ \\ 0, & e_4 \ge 30^\circ \end{cases}$ $s_6 = \begin{cases} 1, & v_{app} \le 3\ kn \\ \frac{8 - v_{app}}{5}, & 3 < v_{app} < 8\ kn \\ 0, & v_{app} \ge 8\ kn \end{cases}$ $S_W = 0.15s_1 + 0.10s_2 + 0.20s_3 + 0.20s_4 + 0.20s_5 + 0.15s_6$ Here, t_r is the interval to the first hard-over response, δ_{max} is the initial maximum absolute rudder angle, S_b measures method suitability for the visibility and P_m is the completion factor based on cumulative heading change Θ_{Hdg}. The execution score S_W combines s_1 through s_6: hard-over timing, rudder magnitude, counter-rudder timing, return to amidships, alignment with the casualty and approach speed. The heading change at first counter-rudder is θ_{rev}. The smallest angular separation between heading at return to amidships and the reciprocal of the initial heading is e_3. The minimum angular separation after return to amidships between heading and the true bearing to the casualty is e_4; bearing is calculated from simultaneous vessel and casualty positions. Angles are in degrees. Mean approach speed over ground, v_{app}, is in knots. A windward-side casualty at final approach triggers the scenario veto.

Table 5. *Cont.*

No.	Assessment Indicator	Scoring Function
8	Recovery-manoeuvre quality	Anderson turn At the first return of the rudder to amidships after the turn begins, the accumulated change in heading in the initial turn direction (Θmid) is measured. Successive headings across 0°/360° are unwrapped before calculating the deviation from 270°. $e_A = \lvert\Theta_{mid} - 270°\rvert$ $S_A = \begin{cases} 1 & e_A \le 5° \\ \frac{30° - e_A}{25°} & 5° < e_A < 30° \\ 0 & e_A \ge 30° \end{cases}$ $S_m = S_b S_A$ The indicator score is the product of the method-suitability factor and the execution score. If no return to amidships is detected by the end of the manoeuvre, the indicator scores zero. A return at 268° earns full execution credit; a return at 250° earns 0.40.
9	Low-speed approach distance	$S_d = \begin{cases} 1, & d \le 0.05 \text{ NM} \\ \frac{0.50 - d}{0.45}, & 0.05 < d < 0.50 \text{ NM} \\ 0, & d \ge 0.50 \text{ NM} \end{cases}$ Here, d is the vessel–casualty distance in the most recent valid rescue-phase record with vessel speed below 5 kn, in nautical miles.
10	Rescue-boat launch order issued	$S_{lt} = \begin{cases} 1, & E_{lt} = 1 \\ 0, & E_{lt} = 0 \end{cases}$ The binary indicator E_{lt} equals 1 when an explicit order to launch the rescue boat is detected in the bridge speech record and 0 otherwise. The simulator does not support physical deployment of the rescue boat.
11	Monitoring of surrounding vessel traffic	$S_t = max(min(1, \frac{N_r + N_e + N_v}{10}), R_l)$ For each window, the three activity counts N_r, N_e and N_v are rudder-order changes, stop operations deduplicated by control value, and recorded nonvisual events. Ten activities earn full activity credit. The lookout ratio R_l is the proportion of video samples with detected lookout. The greater of activity and lookout scores is retained for each window, and the maximum across windows is reported.

3.2.3. Scoring Model

Scoring combined rule-based decisions, AHP-derived or custom weights, membership functions, and event-triggered stage assessment. The overall score was calculated as follows:

$$S = 100 \times \Sigma(w_i \times s_i) \quad (13)$$

In Equation (13), w_i is the indicator weight and s_i its membership score. A windward-side veto at final approach sets the overall score to zero and the grade to fail. The underlying stage scores and evidence remain available for review.

4. Experimental Validation and Results

4.1. Experimental Platform and Data Acquisition

The experiments used the V.Dragon 5000 navigation simulator (Dalian Maritime University, Dalian, China) at Dalian Maritime University. The modelled vessel was the training ship Yu Peng (approximately 30,000 t deadweight, length 199.75 m, breadth 27.8 m and

moulded depth 15.5 m). Heading, speed, rudder and position data were aligned with bridge audio and video relative to the MOB trigger.

BRM course records were used for model development and calibration. Independent simulator assessments used nine separate teams of three maritime practitioners. Casualty side, wind direction, other wind and current conditions, and audio/video availability were pre-set. Case 3 had no video, and Case 6 had no audio. Table 6 distinguishes pre-set conditions from the manoeuvres observed and post-exercise classifications.

Table 6. Simulator scenario settings and observed case information.

No.	Simulator Pre-set	Manoeuvre Performed	Data Availability	Post Hoc Classification
1	Port side/easterly wind	Williamson	Complete	Standard operation
2	Port side/easterly wind	Anderson	Complete	Manoeuvring deviation
3	Port side/easterly wind	Williamson	Video unavailable	Missing modality
4	Port side/westerly wind	Williamson	Complete	Manoeuvring deviation
5	Port side/westerly wind	Anderson	Complete	Standard operation
6	Port side/westerly wind	Williamson	Speech unavailable	Missing modality
7	Starboard side/easterly wind	Anderson	Complete	Standard operation
8	Starboard side/easterly wind	Williamson	Complete	Compound deviation
9	Starboard side/easterly wind	Anderson	Complete	Veto-condition test

4.2. Experimental Data and Scoring Results

Figure 7a–c show manoeuvring tracks for a port-side casualty under an easterly wind. Case 1 completed the recovery track. Case 2 deviated during rudder-amidships and stopping-distance control. Case 3 showed lower turning quality, and the acquisition log indicated missing video evidence. Figure 7d–f show tracks for a port-side casualty under a westerly wind. Case 4 deviated from the original track during the reciprocal return. Case 5 satisfied the assessment rules for action timing, turning quality, and approach control. Case 6 lacked speech evidence, preventing valid scoring of the emergency-procedure initiation stage. Figure 7g–i show tracks for a starboard-side casualty under an easterly wind. Case 7 satisfied all operational assessment rules. Case 8 had delays in hard-over rudder application, lifebuoy deployment, and VTS reporting, together with a return-track deviation. Case 9 completed an Anderson turn, but the casualty was on the windward side during the final approach, triggering the veto condition. The corresponding event data and overall scores are presented in Tables 7 and 8.

Missing speech or video data do not necessarily mean that the corresponding actions were not performed. The assessment model retains the predefined indicator weights and assigns zero to indicators lacking valid evidence. Indicators remain assessable when sufficient evidence is available from other data sources. Scores obtained under missing-modality conditions therefore reflect both recorded performance and evidence availability. These scores should be interpreted alongside records of missing data, and affected indicators should be reviewed before a final judgement of seafarer competence is made.

Table 7 gives the operational measurements, Table 8 the resulting scores and Table 9 the system and expert grades. Grades were excellent (score ≥ 90), good (70 to <90), pass (60 to <70) and fail (<60).

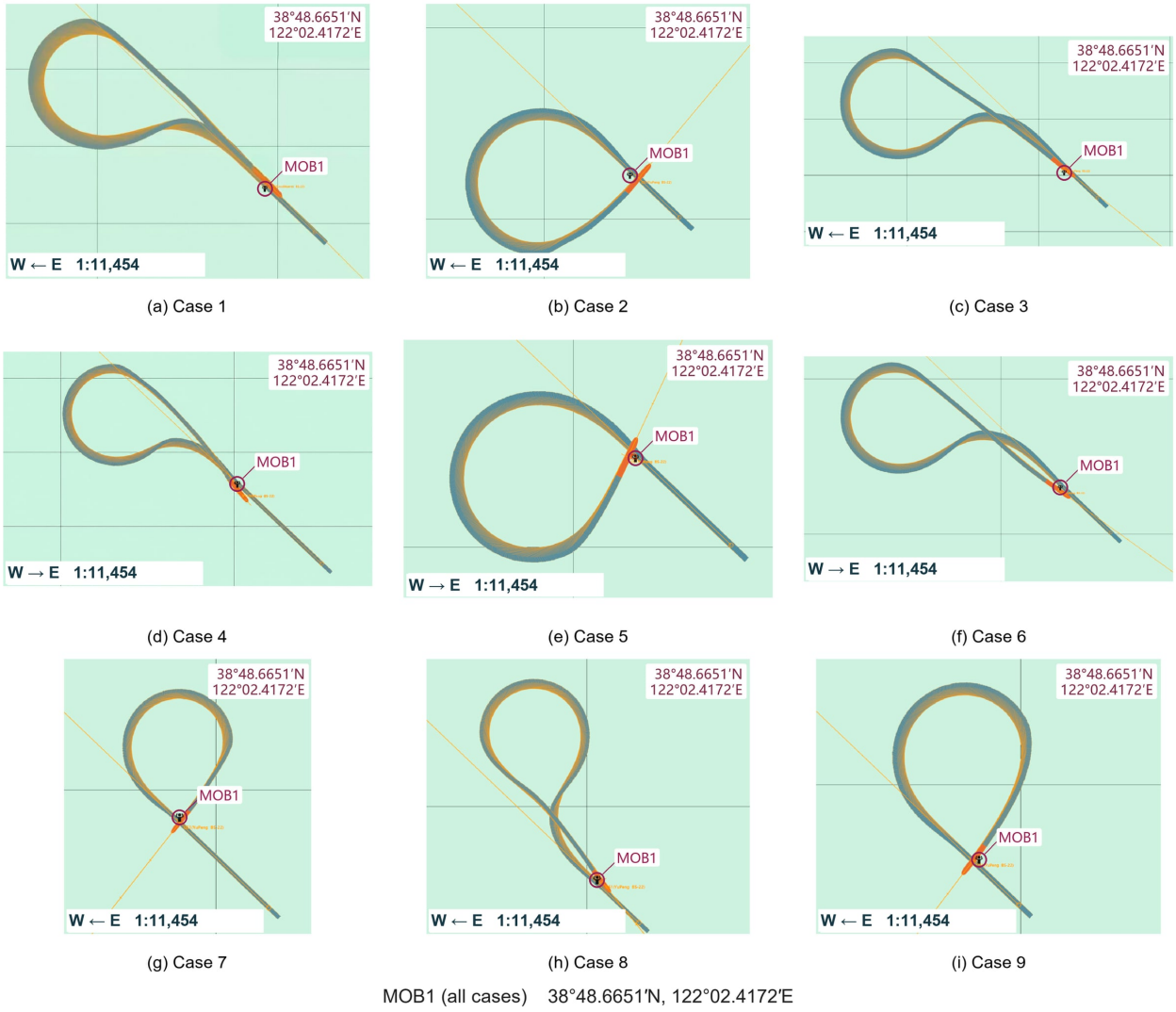


Figure 7. Recovery tracks for nine simulator assessments. MOB1 denotes the casualty. The arrow indicates the direction of wind travel between west (W) and east (E). The scale is the simulator display scale at capture.

Two professional assessors from the Dalian Maritime Safety Administration jointly reviewed each of the seven cases with a reference grade. They discussed the performance without access to the automated score and agreed on a reference grade in each case. No third assessor was required. Table 9 compares the system grade with each panel's joint judgement.

Table 7. Emergency-response and event data for the nine experimental cases.

No.	Stopping Response (s)	Hard-Over Response (s)	Lifebuoy Deployment (s)	Position Recorded	Alarm (s)	VTS Report (s)	Navigational Warning	Key Rudder-Angle Deviation (°)	Approach Speed (kn)	Low-Speed Approach Distance (NM)	Rescue-Boat Launch Order Issued	Vessel Traffic Monitored
1	4.327	6.914	9.583	1	12.436	183.647	1	0.5	4.61	0.034	1	1
2	6.184	8.127	11.742	1	17.816	226.438	1	1.2	5.80	0.047	1	1
3	5.731	9.268	13.417	1	22.463	274.629	1	6.0	6.32	0.041	1	—
4	7.946	9.784	14.623	1	27.218	292.847	1	5.1	5.65	0.049	1	1
5	3.764	5.932	8.517	1	9.681	141.236	1	1.0	4.40	0.028	1	1
6	6.628	8.443	12.874	—	—	—	—	2.6	4.83	0.036	—	1
7	4.872	6.536	10.419	1	15.647	198.734	1	0.3	5.70	0.031	1	1
8	14.238	16.472	20.726	1	25.381	500.024	1	12.0	6.00	0.045	1	1
9	4.216	7.043	9.174	1	13.963	181.147	1	1.6	5.96	0.031	1	1

Note: All time measurements are referenced to the MOB-event trigger. A value of 1 in the position-recording and navigational-warning columns indicates event completion. In the launch-order column, 1 denotes a detected spoken order and a dash denotes unavailable speech evidence.

Table 8. Stage scores and overall assessment results for the experimental cases.

No.	Immediate Action S_I	Emergency Procedure Initiation S_E	Subsequent Rescue Action S_R	Overall Score	Grade/Outcome
1	1.000	1.000	1.000	100.0	A/Excellent
2	1.000	1.000	0.940	97.3	A/Excellent
3	1.000	1.000	0.682	85.8	B/Good
4	1.000	1.000	0.909	95.9	A/Excellent
5	1.000	1.000	1.000	100.0	A/Excellent
6	0.744	0.000	0.761	60.7	C/Pass
7	1.000	1.000	1.000	100.0	A/Excellent
8	0.884	0.963	0.821	87.1	B/Good
9	1.000	1.000	0.000 (veto)	0.0	E/Fail

Note. The overall score S determines the grade as follows: Excellent ($S \geq 90$), Good ($70 \leq S < 90$), Pass ($60 \leq S < 70$), and Fail ($S < 60$).

Table 9. Comparison of intelligent assessment scores and expert evaluations.

No.	Intelligent Assessment	Expert Evaluation
1	100.0 (Excellent)	Excellent
2	97.3 (Excellent)	Excellent
3	85.8 (Good; video unavailable)	—
4	95.9 (Excellent)	Excellent
5	100.0 (Excellent)	Excellent
6	60.7 (Pass; speech unavailable)	—
7	100.0 (Excellent)	Excellent
8	87.1 (Good)	Pass
9	0.0 (Fail)	Fail

4.3. Agreement with Expert Evaluations

All standard cases with complete modalities and no triggered veto met the pass criterion. Cases 1, 5, and 7 satisfied all assessment rules and achieved the maximum overall score. The deviations in Cases 2, 4, and 8 corresponded to the tracks in Figure 7b,d,h and the quantitative measurements in Table 7. In Case 9, the casualty was on the windward side during the final approach, triggering the predefined veto condition. Both the system and the experts classified this case as fail. This finding shows that the system identified the critical safety error and applied the predefined rule.

The missing-modality cases showed that each data stream contributed distinct and complementary evidence within the assessment framework. Case 3 lacked video evidence, preventing verification of crew behaviour and some operational states. The system assessed the available indicators using navigation and speech data and produced an overall score of 85.8. Case 6 lacked speech evidence, so emergency-procedure indicators for alarms, reports, and command execution could not be evaluated. Its overall score consequently decreased to 60.7. Video supported behavioural verification, whereas speech supported

BRM communication and emergency-procedure assessment. Together with navigation data, both modalities maintained the completeness of the assessment evidence chain.

System and expert grades agreed in Cases 1, 2, 4, 5, 7 and 9 (6/7, 85.7%). In Case 8, the system assigned good and the experts assigned pass. Both identified track deviation and delayed return. The experts considered the combined effect of delayed emergency actions and manoeuvring deviations, while the weighted system score remained in the good interval. Across the four ordered grades, linearly weighted Cohen's κ was 0.857 [29]. The agreement observed in the seven expert-rated cases provides preliminary support for the model's assessment consistency within the tested scenarios.

5. Conclusions

We developed a multimodal intelligent BRM assessment system to address subjective manual assessment, delayed feedback, and fragmented multisource evidence in MOB rescue training. The system integrates navigation data, bridge speech, and video-based behaviour evidence. Time synchronisation, hierarchical fusion, and rule-based decisions enable stage-specific scoring of the rescue process. A veto mechanism also enforces critical safety conditions. Case-based validation showed that the system distinguished standard operations, manoeuvring deviations, and missing modalities. Its ratings agreed closely with expert judgements. The system can therefore provide objective and quantitative evidence for training diagnosis, post-task review, and seafarer competence assessment within the tested simulator setting.

This study was limited by the small number of simulator cases and expert assessments, and no direct comparison with other assessment methods was conducted. The findings support the feasibility of the proposed framework within the tested scenarios but do not establish its general applicability across vessels and training environments. Future work will evaluate the framework across vessel types, environmental conditions, trainee groups and simulator platforms.

6. Patents

The following Chinese patent application resulted from this work and has been accepted for processing but has not yet been granted: Dalian Maritime University; Dalian Haida Zhilong Technology Co., Ltd. A BRM Assessment Method Based on Multimodal Data from a Navigation Simulator. Chinese Patent Application No. CN202610241710.6, 12 June 2026.

Author Contributions: Conceptualization, H.S.; methodology, H.W.; software, H.S.; investigation, H.W.; data curation, H.W.; writing—original draft preparation, H.W.; supervision, Y.G., X.F. and X.D.; project administration, Y.G., X.F. and X.D. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Liaoning Provincial Key R&D Program, grant number 2025JH2/102700009, and the Shandong Provincial Key R&D Program, grant number 2025CXGC010710.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Informed consent was obtained from all participants for participation and for the collection and research use of identifiable audio and video data.

Data Availability Statement: The data are not publicly available because the identifiable audio and video records are subject to participant-privacy and institutional restrictions.

Acknowledgments: During the preparation of this manuscript, the authors used OpenAI Codex (OpenAI, San Francisco, CA, USA) for English translation and the translation and processing of text in figures. The authors reviewed and edited the output and take full responsibility for the content of

this publication. The authors would like to express their sincere gratitude to Shen for his valuable academic guidance. We also gratefully acknowledge the financial support provided by the Liaoning Provincial Key R&D Program and the Shandong Provincial Key R&D Program.

Conflicts of Interest: Author Huibin Wang was employed by the company Zhilong (Dalian) Marine Technology Co., Ltd. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Appendix A

Table A1. Stage-level pairwise comparisons (CR = 0.002602).

Indicator	Immediate Action	Emergency Procedure Initiation	Subsequent Rescue
Immediate action	1.0000	1.9382	0.7579
Emergency procedure initiation	0.5159	1.0000	0.4611
Subsequent rescue	1.3195	2.1689	1.0000

Table A2. Immediate-action indicators (CR = 0.000265).

Indicator	Stopping Response	Hard-Over Response	Lifebuoy Response	Position Recorded
Stopping response	1.0000	0.9857	1.0000	1.0145
Hard-over response	1.0145	1.0000	1.0353	0.9659
Lifebuoy response	1.0000	0.9659	1.0000	0.9330
Position recorded	0.9857	1.0353	1.0718	1.0000

Table A3. Emergency procedure indicators (CR = 0.000769).

Indicator	MOB Alarm	Authority/VTS Report	Navigation Warning
MOB alarm	1.0000	1.6341	1.9548
Authority/VTS report	0.6120	1.0000	1.0937
Navigation warning	0.5116	0.9143	1.0000

Table A4. Subsequent rescue indicators (CR = 0.000302).

Indicator	Manoeuvre Quality	Low-Speed Approach Distance	Rescue-Boat Launch Order	Traffic Monitoring
Manoeuvre quality	1.0000	0.9330	1.8771	2.0237
Low-speed approach distance	1.0718	1.0000	1.8993	1.9831
Rescue-boat launch order	0.5327	0.5265	1.0000	1.0353
Traffic monitoring	0.4941	0.5043	0.9659	1.0000

Entries are geometric means of 26 expert judgements. Values above one favour the row indicator over the column indicator. Consistency ratios were calculated before rounding.

Table A5. Confusion matrix for video behaviour recognition. Rows give manual reference labels, and columns give model predictions.

Reference/Prediction	Steering	Console Operation	ECDIS Observation	Lookout	Screen Target Unclear	Total
Steering	10	5	2	0	0	17
Console operation	1	28	7	2	8	46
ECDIS observation	1	1	58	3	2	65
Lookout	0	0	1	8	4	13
Screen target unclear	1	0	10	1	17	29
Total	13	34	78	14	31	170

References

1. Case, S.L.; Lincoln, J.M.; Lucas, D.L. Fatal falls overboard in commercial fishing—United States, 2000–2016. *MMWR Morb. Mortal. Wkly. Rep.* **2018**, *67*, 465–469. Correction in *MMWR Morb. Mortal. Wkly. Rep.* **2018**, *67*, 613. [CrossRef] [PubMed]
2. Marine Accident Investigation Branch. Planning and Preparation Vital to Reduce Man Overboard Fatalities. 2023. Available online: https://www.gov.uk/government/news/planning-and-preparation-vital-to-reduce-man-overboard-fatalities (accessed on 28 October 2026).
3. Karahalil, M.; Lützhöft, M.; Scanlan, J. Formative assessment in maritime simulator-based higher education. *WMU J. Marit. Aff.* **2023**, *22*, 181–207. [CrossRef]
4. Tusher, H.M.; Nazir, S.; Ghosh, S.; Rusli, R. Seeking the best practices of assessment in maritime simulator training. *TransNav Int. J. Mar. Navig. Saf. Sea Transp.* **2023**, *17*, 105–114. [CrossRef]
5. Atik, O. Eye tracking for assessment of situational awareness in bridge resource management training. *J. Eye Mov. Res.* **2020**, *12*, 7. [CrossRef] [PubMed]
6. Röttger, S.; Krey, H. Experimental study on the effects of a single simulator-based bridge resource management unit on attitudes, behaviour and performance. *J. Navig.* **2021**, *74*, 1127–1141. [CrossRef]
7. Ernstsen, J.; Nazir, S. Performance assessment in full-scale simulators—A case of maritime pilotage operations. *Saf. Sci.* **2020**, *129*, 104775. [CrossRef]
8. Zhang, X.K.; Yang, Z.Z.; Hao, H.Q. Evaluation system of LNG liquid cargo operations simulator training. *J. Dalian Marit. Univ.* **2013**, *39*, 13–16. (In Chinese)
9. Cao, H.; Zhang, J. Cloud model-based intelligent evaluation method in marine engine room simulator. *IEEE Access* **2020**, *8*, 168502–168515. [CrossRef]
10. Juszkiewicz, W.; Żukowska, A. The use of the K-Sim Polaris simulator in the process of automatic assessment of navigator competence in the aspect of anticollision activities. *Appl. Sci.* **2023**, *13*, 915. [CrossRef]
11. Shen, H.L.; Zhao, H.; Cao, H.; Liang, S.Y. Intelligent assessment for marine engine simulator based on membership cloud gravity center evaluation. *Navig. China* **2022**, *45*, 7–12+20. (In Chinese)
12. Wang, H.C.; Yin, Y. Intelligent evaluation of rescuing persons from water in navigation simulator. *J. Syst. Simul.* **2021**, *33*, 2647–2655. (In Chinese)
13. Zhang, P.; Wang, J.; Wan, M.; Zhang, S.; Jing, J.; Ding, L.; Shi, P. Audio-visual event localization with dual temporal-aware scene understanding and image-text knowledge bridging. *Complex Intell. Syst.* **2025**, *11*, 15. [CrossRef]
14. Guo, J.; Liu, Y.; Ma, Y.; Liu, M.; Qiu, L.; Zhang, Y.; Nie, X.; Guo, Y.; Yin, Y. Data-and-Model Uncertainty-Aware Multimodal Dynamic Fusion. *IEEE Trans. Circuits Syst. Video Technol.* **2026**, *early access*. [CrossRef]
15. Sellberg, C.; Sharma, A. Toward multimodal learning analytics in simulation-based collaborative learning: A design ethnography of maritime training. *Int. J. Comput.-Support. Collab. Learn.* **2025**, *20*, 201–221. [CrossRef]
16. Lall, V.; Liu, Y. AI meets maritime training: Precision analytics for enhanced safety and performance. *arXiv* **2025**, arXiv:2507.01274.
17. Bhagat, S.; Kim, T.E.; Munim, Z.H. Effective use of visual data for performance assessment in maritime simulator training. *Ocean Coast. Manag.* **2026**, *282*, 108351. [CrossRef]
18. Chen, J.B.; Wu, G.Y.; Ying, S.J. Evaluation of large-vessel handling simulation based on synthetic grey-fuzzy method. *J. Shanghai Marit. Univ.* **2008**, *29*, 1–5. (In Chinese)

19. Wang, D.L.; Ren, H.X.; Zhao, Y.L.; Fang, X.W. Research on automatic assessment system for the maneuver of entering or leaving port based on ship handling simulator. *J. Dalian Marit. Univ.* **2014**, *40*, 1–4. (In Chinese)
20. Kothe, C.; Shirazi, S.Y.; Stenner, T.; Medine, D.; Boulay, C.; Grivich, M.I.; Artoni, F.; Mullen, T.; Delorme, A.; Makeig, S. The lab streaming layer for synchronized multimodal recording. *Imaging Neurosci.* **2025**, *3*, IMAG.a.136. [CrossRef] [PubMed]
21. Kim, I.; Chae, C.; Lee, S. Simulation study of the IAMSAR standard recovery maneuvers for the improvement of serviceability. *J. Mar. Sci. Eng.* **2020**, *8*, 445. [CrossRef]
22. Radford, A.; Kim, J.W.; Xu, T.; Brockman, G.; McLeavey, C.; Sutskever, I. Robust speech recognition via large-scale weak supervision. In Proceedings of the 40th International Conference on Machine Learning, Honolulu, HI, USA, 23–29 July 2023; Proceedings of Machine Learning Research, Volume 202, pp. 28492–28518.
23. Kuchaiev, O.; Li, J.; Nguyen, H.; Hrinchuk, O.; Leary, R.; Ginsburg, B.; Kriman, S.; Beliaev, S.; Lavrukhin, V.; Cook, J.; et al. NeMo: A toolkit for building AI applications using neural modules. *arXiv* **2019**, arXiv:1909.09577.
24. Ultralytics. Ultralytics YOLOv8. 2023. Available online: https://docs.ultralytics.com/models/yolov8/ (accessed on 15 November 2026).
25. Zhang, Y.; Sun, P.; Jiang, Y.; Yu, D.; Weng, F.; Yuan, Z.; Luo, P.; Liu, W.; Wang, X. ByteTrack: Multi-object tracking by associating every detection box. In *Computer Vision—ECCV 2022*; Avidan, S., Brostow, G.J., Cissé, M., Farinella, G.M., Hassner, T., Eds.; Springer Nature Switzerland: Cham, Switzerland, 2022; Volume 13682, pp. 1–21.
26. Lugaresi, C.; Tang, J.; Nash, H.; McClanahan, C.; Uboweja, E.; Hays, M.; Zhang, F.; Chang, C.L.; Yong, M.G.; Lee, J.; et al. MediaPipe: A framework for building perception pipelines. *arXiv* **2019**, arXiv:1906.08172.
27. Saaty, T.L. A scaling method for priorities in hierarchical structures. *J. Math. Psychol.* **1977**, *15*, 234–281. [CrossRef]
28. International Maritime Organization. Guide to Recovery Techniques. MSC.1/Circ.1182/Rev.1, 21 November 2014. Available online: https://wwwcdn.imo.org/localresources/en/OurWork/Safety/Documents/MSC.1-Circ.1182-Rev.1%20-%20Guide%20To%20Recovery%20Techniques.pdf (accessed on 16 November 2026).
29. Cohen, J. Weighted kappa: Nominal scale agreement with provision for scaled disagreement or partial credit. *Psychol. Bull.* **1968**, *70*, 213–220. [CrossRef] [PubMed]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.