Experiment Design and Implementation for Human-in-the-Loop Study Towards Maritime Autonomous Surface Ships

Baiheng Wu, Martin Lied Sæter, Hans Petter Hildre, Houxiang Zhang, and Guoyuan Li†

Abstract—**The development of maritime autonomous surface ships (MASS) has triggered interest from both academia and industry in recent years. Nevertheless, human operators will continue to perform dominant roles onboard for the next decades. There are several critical phases where human operators are in the navigation loop (MASS-I, II, and III) before full ship autonomy (MASS-IV) is achieved. The authors conceive a cyber-physical human framework for the experiment design and implementation in the maritime domain: based on multiple experimental platforms with data exchange ports, apply monitoring on and learning from navigators' behaviors, and adapt the ship-bridge system to provide decision support in terms of guidance, navigation, and control. The platforms include a compact simulator from Kongsberg® and an immersive simulator for preliminary research design, a group of standard maritime training simulators, and a research vessel. These platforms are utilized for data collection, scenario design, testbeds to demonstrate and verify new techniques and algorithms, etc. The authors illustrate how the framework aids the MASS research and benefits the development process.**

Index Terms—**Maritime autonomous surface ships (MASS), human-in-the-loop (HITL), behavioral analysis.**

I. Introduction

The maritime transportation sector contributes dominantly (around 95 % of total volumes) to the global trade [1]; thus, sailing security is crucial in many respects, including onboard crew's life safety and the world economy. It is reported by Lloyd's that during the 2021 Suez Canal obstruction, the value of goods delayed per day was about 10 billion dollars [2]. The statistics from different resources show that human factor-related failure and accidents take a main proportion in marine perils [3]. From another respect, human resources cost is one of the highest expenditures onboard [4]. In this regard, the development of ship autonomy can be beneficial to both enhancing maritime transportation security and promoting economic shipping efficiency. The International Maritime Organization (IMO) has conceptualized and formulated maritime autonomous surface ships (MASS) at four levels (I-IV) in terms of human crew presence and ship intelligence degrees [5]. Except for the level MASS-IV where ships are fully automated, the human crew are considered to be in the loop either onboard or remotely, either as the decision-maker or monitor [6]. Autonomous ships and intelligent maritime transportation are promising and desired, but not until issues in technology, ethics, and sociology, are addressed [7] [8].

TABLE I: Specification of MASS I-IV

Level	Crew	Human Control	Autonomy
I	Onboard	Onboard	Decision support
II	Onboard	Remote	Crew ready to takeover
III	Off board	Remote	No onboard intervention
IV	Off board	No intervention	Full

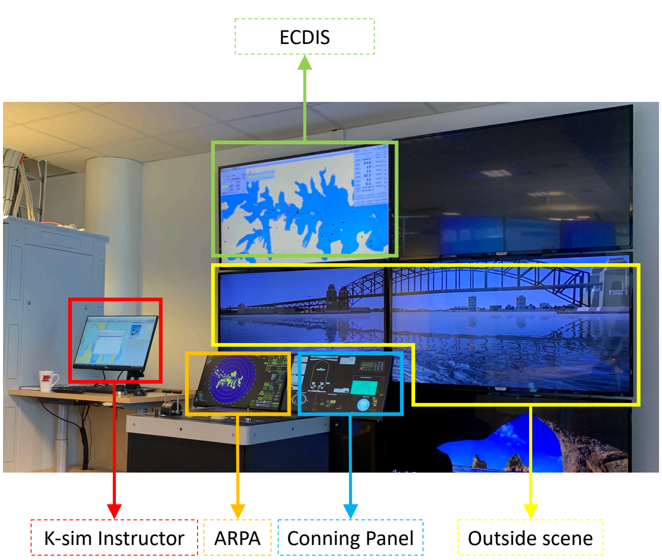
Table I illustrates the details of human presence, intervention, and ship autonomy of MASS I-IV. It plans how MASS gradually gains independence from human crews level by level, from onboard operation to remote control, from human-centered maneuvering to human as monitor, etc. There are essential factors to consider in each step of progress to secure safety and add redundancy. As three out of four MASS levels keep the human crew in the loop to different extents, this also suggests that human-in-the-loop (HITL) maritime research should still be a core issue for the development of MASS [9] [10]. In another respect, though human-factor-related failures are dominant in statistics, human navigators are still the most eligible players. They are well-educated and trained professionals, and they have recorded countless safe-sailing hours. Therefore, taking advantage of their occupational and expertise knowledge is also beneficial to the development of MASS [11]. From the whole maritime transportation environment, a transition period is to be expected where manned, semi-autonomous, fully-automated ships may encounter each other in the traffic, and the unpredictability and difficulty of intercommunication between human and ship intelligence may be a problem. So, learning human navigators' operational and navigational logic and mechanisms and feeding the ship intelligence with the expertise and knowledge may improve the coexistence of traffic when there are different types of ship governance.

Learning and accumulating knowledge from the practice is a tradition in this industry; the precedent practitioners have recorded and concreted their experience as routines, regulations, practice recommendations, etc. Their efforts have significantly improved the marine traffic safety, and efficiency [12]. While keeping this conventional way, the emerging technologies, including machine learning, data mining, and sensor fusion, enable us to objectively model human navigators' behavior with data-based methods. Moreover, simulators designed in accordance with the onboard ship-bridge environment and facilitated with actual navigational

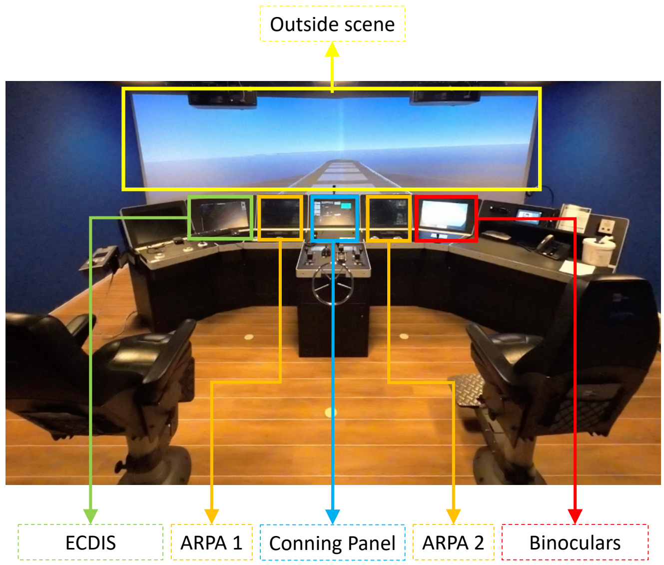
*The research is supported in part by the MAROFF KPN project "Digital Twins for Vessel Life Cycle Service" (Project No.: 280703) and in part by the Startplugg 2021 (NTNU i Ålesund) project "Intelligent Ship Bridge System for Learning Navigators' Operational Behaviors" (Project No.: 70444453) in Norway.

*B. Wu is with both the Department of Ocean Operations and Civil Engineering (IHB) and the Department of ICT and Natural Sciences (IIR), NTNU, Ålesund Campus, Larsgårdsvegen 2, 6009 Ålesund, Norway; M.L. Sæter, H.P. Hildre, H. Zhang, and G. Li are with the Department of Ocean Operations and Civil Engineering (IHB).

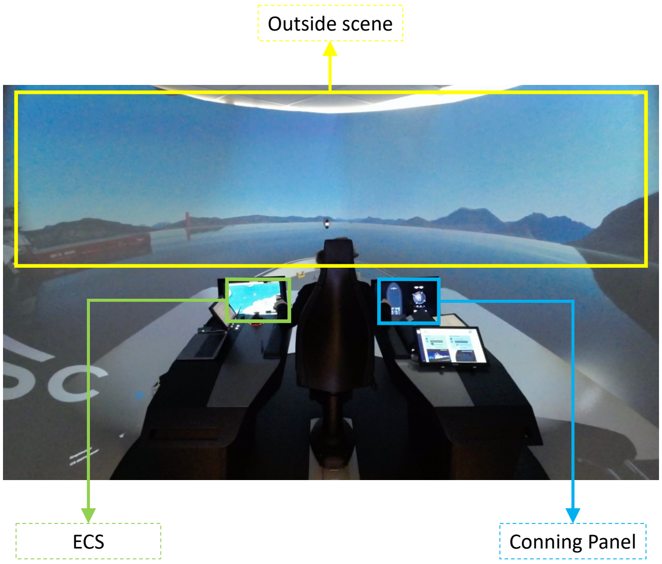
†Correspondance Email: guoyuan.li@ntnu.no



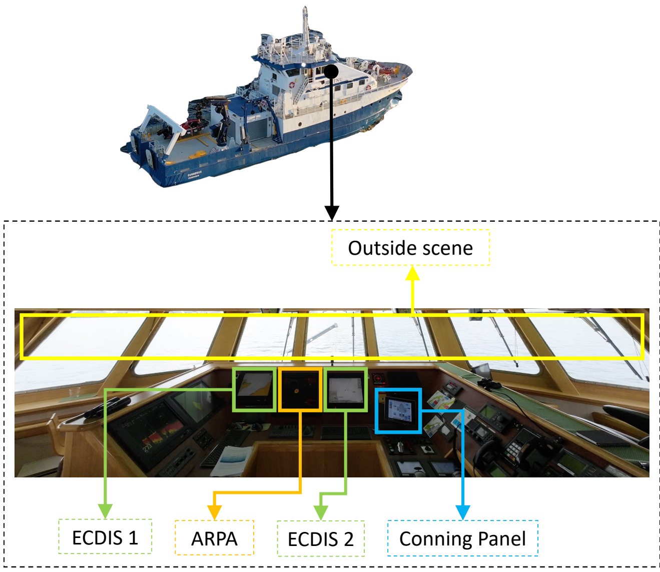
(a) Compact simulator from Kongsberg® for preliminary research design.



(b) Kongsberg® maritime training simulator.



(c) Immersive ship-bridge simulator powered by Offshore Simulator Centre AS (OSC).



(d) NTNU's research vessel R/V Gunnerus and its ship bridge.

Fig. 1: Experimental platforms. (Terminology: Electronic Chart Display and Information System (ECDIS); Electronic Chart Systems (ECS)); Automatic Radar Plotting Aids(ARPA).)

operations' functionalities expand the flexibility and accessibility of experiment design, and implementation process [13] [14]. We illustrate all these platforms including different simulators and research vessels (as shown in Fig. 1) in Section II.

With the substantial experiment facilities, we have been designing and implementing research in the scope of HITL onboard operations. The wearable electroencephalography (EEG) equipment is used to record the navigators' brain active level to detect the signals of fatigue and assess the fatigue levels [15] [16] [17]. The navigators wear the wearable eye tracker to trace their eye movement to reflect their visual interests and transition between different areas of interest (AOIs), and it has been used in operational analysis, including crane lifting [18] [19], high-speed cruising [20], and ship maneuvering for collision avoidance [21]. Both massive simulator-generated and real-ship data help to conclude and summarize the navigators' generality and similarity in navigational operations wu [22] [23]. In addition to the traditional analytical method and expert interpretation, algorithms, for example, support vector machine (SVM), are also utilized to realize classification and recognition tasks in HITL maritime domain [24].

In this paper, the authors illustrate the laboratory environment of maritime simulators and research vessels owned by NTNU and how they are used for HITL experiment design and implementation to support comprehensive research on human navigational operations modeling and onboard decision support contribute to the development of MASS.

The paper is organized as follows: Section II introduces the experimental platforms we have been using at NTNU i Ålesund for HITL research in detail, including their specifications and difference; Section III describes how the platforms are used to aide the experimental design and the implementation, and how different parties of interests are involved in the process; Section IV demonstrates how the platforms are played in practice to support HITL research in the course of different applied aims; At last, a conclusion is made in Section V to summarize the content in the paper.

II. Platforms

In this section, we introduce the platforms used at different experimental stages and the aims of the HITL research. The simulators in Fig. 1a and 1b are from Kongsberg®, and the former is for research aim while the latter is for navigator training; The ship-bridge system with fixed seat shown in Fig. 1c is often seen on compact vessels and catamarans, and this simulator projects the outside scene on a dome screen which provides an immersive navigational experience; Fig. 1d is the research vessel R/V Gunnerus which can perform most maneuvering operations and tasks (for example, dynamic positioning, zig-zag path following).

A. for Preliminary Study

The preliminary study applies to the problems which are never being focused on and concerned by existing research and to questions that have been well studied to some extent. However, some hypotheses are made to dig deeper. Such a study exists whenever a new research project is launched. To verify that the problem is concerned does exist and is worth a look into, a preliminary study should be carried out on a small scale in both experimental design and the number of participants. This procedure may increase the experimental efficiency and avoid unnecessary failures.

The platforms in Fig. 1a and 1c are often used for preliminary study. Main differences between simulators in Fig. 1a and 1b are:

- K-sim Instructor which coordinates the experiment implementation is in the control (briefing) room for the standard maritime training simulator in Fig. 1b. At the same time, the K-sim Instructor computer is put just aside from the ship-bridge system in Fig. 1a. This configuration makes the tune-test process much easier and more efficient for researchers to design and modify the scenario;
- as the name suggests, the simulator in Fig. 1a is compact. It is only facilitated with basic functions (no binoculars screen, back view, and only one ARPA system).

Thus the compact simulator is not capable of being used as a training platform but is quite eligible for research aim.

The simulator in Fig. 1c is flexible for research because in contrast to Kongsberg® products, it is not yet commercialized. As NTNU holds a share of the intelligence property to the hardware, we have the opportunity to shape the simulator and its capability in accordance to exact our demands.

Sometimes the research vessel is also used for preliminary study. Though the simulator platforms are made closer and closer to the reality, two issues made preliminary studies on real vessels non-substitutable:

- the experimental scenarios on the simulators are set up by known physical laws and based on accepted deduced principles/inference. However, there are unknown natural mechanisms that result in the difference between the simulator setup and what it simulates; this applies especially to new problems;
- as simulator setup can be made arbitrarily sometimes the experimental designer might deviate too far from the actual situations, and it makes the efforts meaningless and costs in time and money in vain.

In this regard, a research vessel is the best platform to discover any uncharted problems and phenomena. It also applies to humans in a laboratory simulator environment, and actual open-sea conditions are still different after all.

B. for Quantitative Study

A quantitative study takes place when the problem is preliminarily verified as worth a comprehensive study. Simulators in Fig. 1b are used for quantitative study the most. There are five standard maritime training simulators at NTNU i Ålesund, four of them are the same as in Fig. 1b and another one is equipped with a dome screen wall and made on an even larger scale (equal to a regular ferry). The experimental setup and scenario are designed and defined in the central control room by K-sim Instructor, and all participants are operating in the same environments. These facilities are initially for maritime navigators' training, but this process can be deemed as a typical way of massive data collection. In this group of simulators, two types of experiments are carried out:

- individual scenario: all participants are operating individually, and their own ships are invisible to each other, which means they are carrying the same task out in different copies. Using this group of simulators to implement such an experiment drastically saves time and expense as we can collect five sets of data from five simulators, and simulator-based experiments do not burn even a drop of diesel.
- cooperative scenario: since there are five simulators, their own ships can be configured as visible to each other to become target ships to other navigators. Taking advantage of this, experimental scenarios which need cooperation and intercommunication (in a traffic system) can also be implemented.

Simulator in Fig. 1c is also used for a quantitative study. Since the standard maritime training simulators' schedules are tightly occupied almost every day for teaching, training, and research tasks, and this non-standard self-developed platform cannot be used for teaching and training aims, it has plenty of time intervals for research tasks. A booking system is developed for the simulator so that once a research scenario is defined on it, navigators with different levels of experience are invited to book timeslots to contribute their expertise.

C. as Testbed for Verification and Demonstration

As new decision support systems and other automatic functions are developed, they need to be tested, verified, and demonstrated. Platforms in Fig. 1b, 1c have been used as

TABLE II: Usage of different platforms in Fig. 1

Platform	Preliminary	Quantitative	Testbed
Fig. 1a	✓	-	-
Fig. 1b	-	✓	✓
Fig. 1c	✓	✓	✓
Fig. 1d	✓	-	✓

testbeds for these aims in our research. These two platforms have different data exchange interfaces:

- data output: both platforms are capable of exporting real-time data, including ship maneuvering figures, geographical information, and environmental loads. These data export enables us to develop a decision support system for navigational performance enhancement;
- data input: only platform in Fig. 1c has the potential to integrate algorithms in the loop. This feature outperforms K-sim platforms, making the test of automated ship guidance, navigation, and control possible.

The research vessel is definitely also a good testbed, and there have been relevant research items carried out based on data onboard Gunnerus [25] [26]. Moreover, it may have a similar problem as the K-sim platform in Fig. 1b, where besides the technical problem of data input, onboard safety and ethical issues might be concerned.

The usage of each platform is summarized as in Table II.

III. Design & Implementation

As the experimental platforms have been introduced in Section II, we continue with how to design and implement experiments on these platforms.

As shown in Fig. 2, there are mainly three groups of experimenters:

- researchers: who are in charge of and responsible for the whole experiment throughout all procedures, including design, execution, analysis, and demonstration;
- consultants: who provide expertise and occupational suggestions to improve the experiment process;
- participants: who carry out the specific experiment tasks on experimental platforms.

Besides, there are five steps in general when designing and implementing an experiment:

- Preliminary design: the researchers come up with novel research ideas and problems to explore. They should do a preliminary literature review to concrete the problem based on the current state-of-the-art and propose a preliminary route map to improve the method or explore the problem;
- Validation & modification: as the problem to be explored might deviate from the researchers' exact interests and professions, consultants are invited to provide expertise. In HITL maritime domain, captains and university lecturers often play these roles by discussion and workshop to improve the experimental design to trim the irrational parts and make it more practical;
- Massive implementation: researchers are responsible for organizing massive data collection and supervising the experiment implementation (to secure the process is according to the plan), but the participants who carry out the tasks are usually skilled navigators and navigation students (in the bachelor's program);
- Analysis: researchers are in charge of the data interpretation and analysis and developing new decision support/automatic functions and algorithms; the supervisor in this step can also perform supervision over students;
- Verification & demonstration: as researchers develop the new functions and algorithms, they are obligated to organize new experimental sessions to verify the results and performance of refined products; as the main difference between consultants and participants is their experience, in the demonstration phase, practitioners with different levels of experience are invited to provide their opinions for any further improvement.

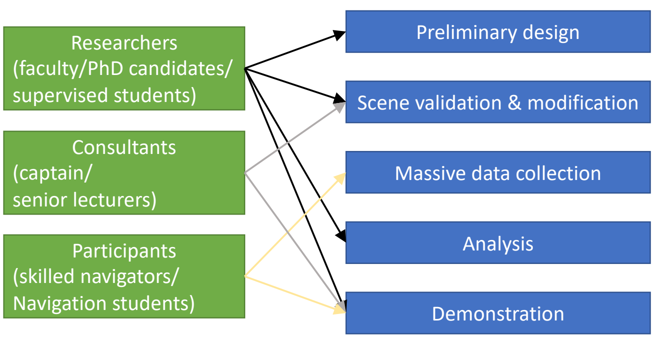


Fig. 2: Participation in experiment design process.

IV. Utilization

As the experimental platform and process of design and implementation have been introduced in previous sections, we illustrate how we use the facilities to carry out HITL research in the onboard navigation domain. Fig. 3 depicts the workflow of an intelligent ship-bridge system in general. In this framework, there are following several parts:

- executives: the human navigators and are also the bearers to be monitored and recorded by the multi-sensor system;
- multi-sensor monitoring: more channels of sensors are expected in the monitoring systems to provide cross-explain to the navigators' behaviors; However, the more sensors equipped, the more influence might be caused on the navigators' performance. In this context, non-intrusive solutions are preferable. If not available, the number of sensor channels should be controlled; the applied examples are shown in Fig. 4;
- behavior modeling: data interpretation and data mining by machine learning algorithms to explore the potential navigational mechanisms;
- derived functions: some passive functions without intervention to the control loop are developed based on the modeled navigational behaviors.

The workflow of behavior modeling and derived functions parts are shown in Fig. 5.

A. Monitoring & Data Collection

Tools for monitoring can be sorted as intrusive and non-intrusive. To record EEG data, a wearable brain data hardware is inevitable; When it comes to visual attention tracking, the solutions are various: wearable eye trackers are still

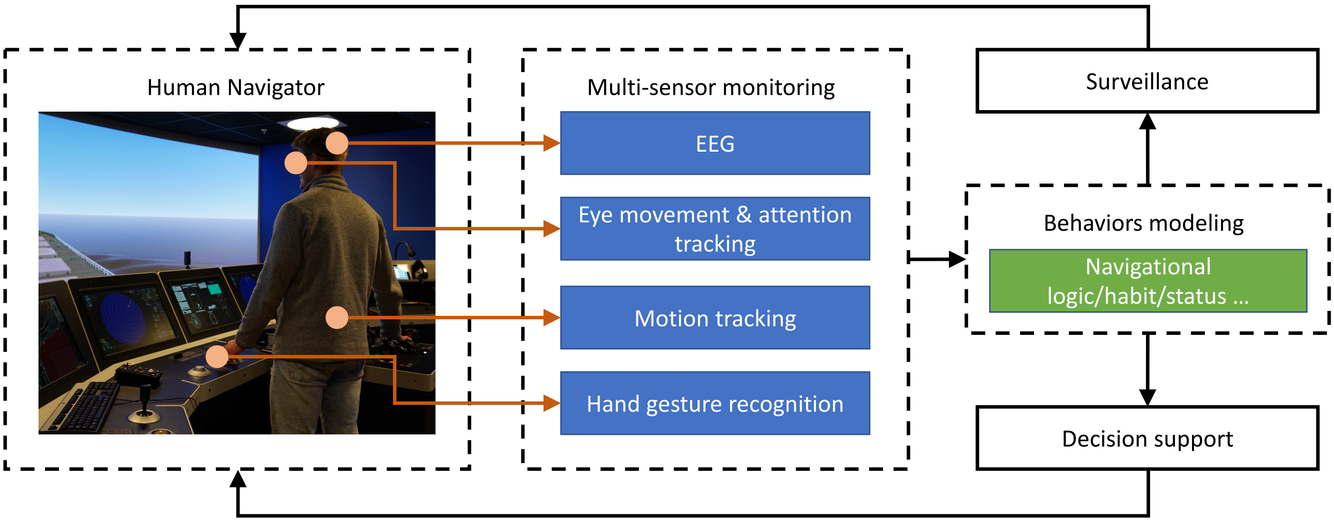


Fig. 3: Intelligent ship-bridge system for research on HITL navigational operations.

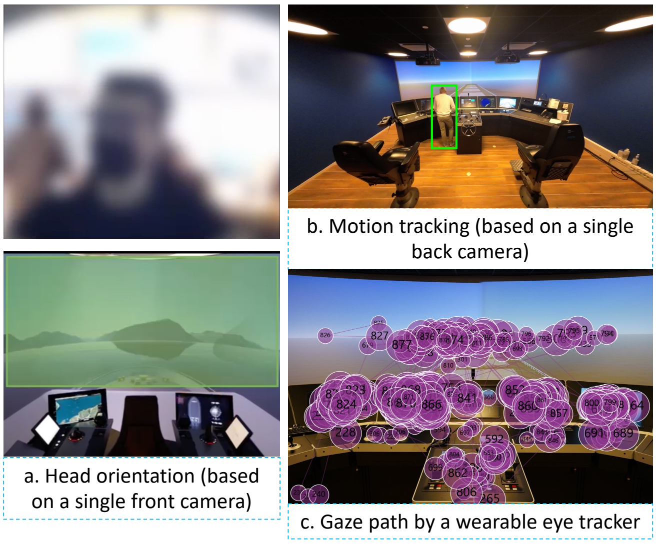


Fig. 4: Applied multi-sensor monitoring scenes.

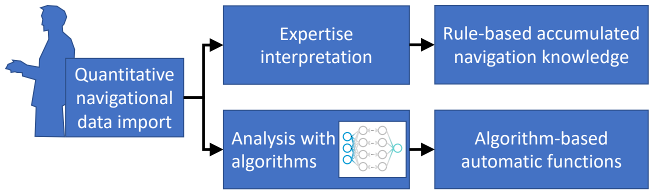


Fig. 5: Analysis procedures.

popular and versatile in conducting various research items (as in Fig. 4c, wearable eye tracker is capable of presenting precise visually transitting path with glance durations), but non-intrusive camera-based (RGB/RGBD) algorithms have been developed based on collected videos as shown in Fig. 4a. In this application, by feeding a deep learning neural network with images with facial and head orientational information, the trained algorithm can tell the navigator's visual attention zones (the navigator is looking at the green zone in Fig. 4a).

The configuration of cameras for tracking and tracing varies. We have put single/dual-camera systems into practice in our experiment environment. We deal with head orientation detection with a single camera system and visual attention and motion of unseating navigators with a dual-camera system. The more data we collect in this stage, the easier the analysis process can be in the following step.

B. *Operations Analysis and Learning*

To deal with collected data, there are two different ways. One is data interpretation based on expertise, and another is the algorithm-based data-driven method. The two methods have their virtues:

- Expertise interpretation: first, expertise may bring a qualitative assessment of the data and rationalize the logic of navigational operations; it is not handful for them to discover existed problems; a step further, by using analytical methods to calculate correlations between different variables, and conducting interviews with experimenters, expertise interpretation may give an intuitionistic and comprehensible result. This may prove a boon to navigational training and teaching.
- Data-driven algorithms: mature machine learning algorithms such as SVM and neural network/deep learning have a satisfying performance in classification and pattern recognition tasks. They are qualified to track navigators' visual attention, body motion, and gestures. It is a good choice for developing automatic functions and automated products, but as the calculation is done within a black box, the processing is difficult to clarify and explain. However, this also triggers the research interest in explainable artificial intelligence.

C. *Online Surveillance & Decision Support*

Fig. 4a shows how the deep learning algorithm classifies visual attention zones, and Fig. 4b illustrates how the trained convolutional-neural-network algorithm traces the navigator's motion (body movement). These algorithms are developed as passive, but it has the potential to detect anomalies that may occur and give warnings and signals to avoidance failures.

V. Conclusion

This paper introduces the experimental platforms to carry out HITL research in the maritime navigation domain. The difference and applicable scenes to each platform are discussed, and simulator and research vessel-based HITL experimental design and implementation routines are consolidated by clarifying the roles of each involved party of interests. This framework has brought great benefits to HITL research. While a primary concern of this framework is: that there are four introduced platforms, and each of them has different data exchange interface. Promoting the data flow in the framework and increasing the generalization between different platforms can vastly improve the efficiency and effectiveness of the framework.

Acknowledgment

Acknowledgement is extended to Girtz Strazdins, Øivind Kjerstad, and Ottar Osen for providing suggestions; to Arnt Håkon Barmen for organizing tests and experiments; to Terje Ole Slinning for his technological support.

References

[1] "Review of maritime transport 2021." UNCTD (United Nations Conference on Trade and Development), United Nations Publications, New York, USA, 2021.

[2] L. A. LaRocco, "Suez canal blockage is delaying an estimated $400 million an hour in goods," *CNBC*, 2021.

[3] B. Wu, T. L. Yip, X. Yan, and C. G. Soares, "Review of techniques and challenges of human and organizational factors analysis in maritime transportation," *Reliability Engineering & System Safety*, vol. 219, p. 108249, 2022.

[4] S. C. Mallam, S. Nazir, and A. Sharma, "The human element in future maritime operations–perceived impact of autonomous shipping," *Ergonomics*, vol. 63, no. 3, pp. 334–345, 2020.

[5] "Regulatory scoping exercise on maritime autonomous surface ships," in *Maritime Safety Committee, 100th session*. International Maritime Organization, 2018.

[6] C. Liu, X. Chu, W. Wu, S. Li, Z. He, M. Zheng, H. Zhou, and Z. Li, "Human–machine cooperation research for navigation of maritime autonomous surface ships: A review and consideration," *Ocean Engineering*, vol. 246, p. 110555, 2022.

[7] A. Vagale, R. Oucheikh, R. T. Bye, O. L. Osen, and T. I. Fossen, "Path planning and collision avoidance for autonomous surface vehicles i: a review," *Journal of Marine Science and Technology*, pp. 1–15, 2021.

[8] I. Ashraf, Y. Park, S. Hur, S. W. Kim, R. Alroobaea, Y. B. Zikria, and S. Nosheen, "A survey on cyber security threats in iot-enabled maritime industry," *IEEE Transactions on Intelligent Transportation Systems*, 2022.

[9] M. Shahbakhsh, G. R. Emad, and S. Cahoon, "Industrial revolutions and transition of the maritime industry: The case of seafarer's role in autonomous shipping," *The Asian Journal of Shipping and Logistics*, vol. 38, no. 1, pp. 10–18, 2022.

[10] T. Miyoshi, S. Fujimoto, M. Rooks, T. Konishi, and R. Suzuki, "Rules required for operating maritime autonomous surface ships from the viewpoint of seafarers," *The Journal of Navigation*, vol. 75, no. 2, pp. 384–399, 2022.

[11] H. Karahalios, "Evaluating the knowledge of experts in the maritime regulatory field," *Maritime Policy & Management*, vol. 44, no. 4, pp. 426–441, 2017.

[12] Y. Zhou, W. Daamen, T. Vellinga, and S. Hoogendoorn, "Review of maritime traffic models from vessel behavior modeling perspective," *Transportation Research Part C: Emerging Technologies*, vol. 105, pp. 323–345, 2019.

[13] J. Ernstsen and S. Nazir, "Performance assessment in full-scale simulators–a case of maritime pilotage operations," *Safety Science*, vol. 129, p. 104775, 2020.

[14] L. Zhao, G. Li, K. Remøy, B. Wu, and H. Zhang, "Development of onboard decision supporting system for ship docking operations," in *2020 15th IEEE Conference on Industrial Electronics and Applications (ICIEA)*. IEEE, 2020, pp. 1456–1462.

[15] T. G. Monteiro, C. Skourup, and H. Zhang, "Using eeg for mental fatigue assessment: A comprehensive look into the current state of the art," *IEEE Transactions on Human-Machine Systems*, vol. 49, no. 6, pp. 599–610, 2019.

[16] T. G. Monteiro, H. Zhang, C. Skourup, and E. A. Tannuri, "Detecting mental fatigue in vessel pilots using deep learning and physiological sensors," in *2019 ieee 15th international conference on control and automation (icca)*. IEEE, 2019, pp. 1511–1516.

[17] T. G. Monteiro, C. Skourup, and H. Zhang, "A task agnostic mental fatigue assessment approach based on eeg frequency bands for demanding maritime operation," *IEEE Instrumentation & Measurement Magazine*, vol. 24, no. 4, pp. 82–88, 2021.

[18] G. Li, R. Mao, H. P. Hildre, and H. Zhang, "Visual attention assessment for expert-in-the-loop training in a maritime operation simulator," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 1, pp. 522–531, 2019.

[19] R. Mao, G. Li, H. P. Hildre, and H. Zhang, "A survey of eye tracking in automobile and aviation studies: Implications for eye-tracking studies in marine operations," *IEEE Transactions on Human-Machine Systems*, vol. 51, no. 2, pp. 87–98, 2021.

[20] O. S. Hareide and R. Ostnes, "Scan pattern for the maritime navigator," *TransNav: International Journal on Marine Navigation and Safety of Sea Transportation*, vol. 11, no. 1, 2017.

[21] B. Wu, L. Zhao, S. R. Thattavelil Sunilkumar, H. P. Hildre, H. Zhang, and G. Li, "Visual attention analysis for critical operations in maritime collision avoidance," in *2022 ieee 17th international conference on control and automation (icca)*. IEEE, 2022.

[22] B. Wu, G. Li, L. Zhao, H. P. Hildre, and H. Zhang, "A human-expertise based statistical method for analysis of log data from a commuter ferry," in *2020 15th IEEE Conference on Industrial Electronics and Applications (ICIEA)*. IEEE, 2020, pp. 1471–1477.

[23] B. Wu, G. Li, L. Zhao, H.-I. J. Aandahl, H. P. Hildre, and H. Zhang, "Navigating patterns analysis for onboard guidance support in crossing collision-avoidance operations," *IEEE Intelligent Transportation Systems Magazine*, 2021.

[24] B. Wu, G. Li, T. Wang, H. P. Hildre, and H. Zhang, "Sailing status recognition to enhance safety awareness and path routing for a commuter ferry," *Ships and Offshore Structures*, vol. 16, no. sup1, pp. 1–12, 2021.

[25] P. Han, G. Li, R. Skulstad, S. Skjong, and H. Zhang, "A deep learning approach to detect and isolate thruster failures for dynamically positioned vessels using motion data," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–11, 2020.

[26] T. Wang, G. Li, L. I. Hatledal, R. Skulstad, V. Æsøy, and H. Zhang, "Incorporating approximate dynamics into data-driven calibrator: A representative model for ship maneuvering prediction," *IEEE Transactions on Industrial Informatics*, vol. 18, no. 3, pp. 1781–1789, 2021.