POAC'25
St. John's, Newfoundland and Labrador, Canada

Proceedings of the 28th International Conference on Port and Ocean Engineering under Arctic Conditions
Jul 13-17, 2025
St. John's, Newfoundland and Labrador
Canada

Multimodal Deep Learning Framework for Estimating Local Ice Loads on Polar Vessels

Eun-Jin Oh[1], Jung-Seok Ha[1], Seong-Yeop Jeong[1]
[1] Korea Research Institute of Ships and Ocean Engineering, KRISO, Ice Model Basin, Daejeon, Republic of Korea

ABSTRACT

The safe navigation and structural resilience of ice-capable vessels in polar regions depend critically on accurate estimation of local ice loads acting on the hull. Traditional approaches rely on strain gauge measurements and inverse finite element method (FEM) modeling to reconstruct the external ice loads based on internal structural responses. While effective under controlled conditions, these methods are inherently time-intensive, highly dependent on prior modeling, and fundamentally unsuited for real-time applications where immediate decision-making is essential. Recent advancements in machine learning offer a promising alternative through data-driven surrogate models capable of learning the complex nonlinear relationships between measured strain signals and corresponding ice loads. Additionally, high-resolution video recordings—commonly available onboard icebreakers—offer rich contextual information about the ice–structure interaction environment, including contact geometry, ice floe morphology, and fracturing dynamics. However, such visual data has rarely been incorporated into load estimation frameworks, leaving untapped potential for enhanced performance and situational awareness. This paper explores the feasibility of integrating strain gauge time-series data and synchronized visual inputs through a multimodal deep learning framework to achieve real-time local ice load estimation. We review recent developments in AI-based inverse load modeling using radial basis function neural networks and support vector machines, and we extend this paradigm by introducing multimodal architectures such as CNN–LSTM hybrids and Transformer-based fusion networks. These architectures are designed to capture both temporal strain dynamics and spatial visual patterns, facilitating a holistic understanding of structural responses to dynamic ice events. We discuss key challenges including data synchronization, representation learning across heterogeneous modalities, and the scarcity of labeled real-world datasets. We also propose a practical pipeline for onboard deployment, leveraging pre-trained AI models to infer ice loads directly from operational data streams. The proposed framework has the potential to fundamentally

change how ice loads are assessed in real-time, offering a lightweight, simulation-free, and scalable solution for next-generation structural health monitoring systems in ice-covered waters.

KEY WORDS : Ice load estimation; Multimodal deep learning; CNN–LSTM architecture; Real-time prediction;

1. Introduction

1.1 Background and Motivation

Polar-class vessels routinely operate in some of the harshest environments on Earth, encountering complex and often unpredictable interactions with drifting sea ice and icebergs. These interactions subject the vessel hull to localized impact loads that vary significantly in magnitude, direction, and duration. The ability to accurately estimate these loads in real-time is critical for structural health monitoring, onboard safety systems, adaptive route planning, and design validation of ice-class hulls. Traditionally, local ice loads have been estimated using arrays of strain gauges installed along the inner structure of the hull. These sensors capture internal deformation patterns that are then used, in conjunction with finite element method (FEM) models, to infer the external loads applied by ice impacts. This approach, while grounded in physical modeling and widely accepted, presents several limitations: it is computationally demanding, sensitive to modeling assumptions, and poorly suited to real-time applications due to its reliance on batch post-processing. Moreover, in complex ice conditions where multiple ice impacts may occur simultaneously or in rapid succession at different locations, distinguishing between individual load events becomes exceedingly difficult based solely on strain data. Synchronized video footage of the ice–hull interface, often collected onboard for documentation or research, can provide complementary contextual information such as impact location, floe geometry, and fracture propagation—all of which influence the nature of the load. Yet, this rich visual information remains largely unutilized in automated load estimation frameworks.

1.2 Research Gap and Opportunity

Recent studies have demonstrated that machine learning techniques—particularly support vector regression and radial basis function (RBF) networks—can successfully model the nonlinear relationships between strain measurements and ice loads without relying on FEM-based inverse solutions. These surrogate models can generalize well to unseen conditions, reduce dependency on simulation, and support near-instantaneous inference. However, these models typically rely on strain data alone and therefore lack external perception of the ice environment, which may contain critical cues for interpreting load scenarios. The fusion of strain signals with synchronized video data represents a compelling avenue for extending the capabilities of existing AI-based frameworks. Multimodal learning—where different data modalities are processed jointly—has shown significant success in fields such as human activity recognition, autonomous driving, and structural monitoring. In these domains, combining sensor data with visual inputs has led to improved performance, robustness to noise, and greater generalizability across environmental conditions. This study investigates

the feasibility and potential benefits of a multimodal deep learning framework for real-time local ice load estimation, integrating strain gauge time-series data with synchronized onboard video imagery. We propose neural network architectures that combine temporal encoders for strain data (e.g., LSTM layers) with visual feature extractors (e.g., CNNs or Vision Transformers), and we examine various strategies for feature-level and attention-based fusion.

Key contributions of this paper include:

- A comprehensive review of recent data-driven ice load estimation models and their limitations;
- A conceptual design for a multimodal AI system that merges strain and visual data streams;
- Recommendations for dataset construction, including temporal alignment of sensor and video data;
- A deployment-oriented discussion on inference latency, computational efficiency, and onboard integration.

By bridging structural response data with environmental perception, our approach aims to transform traditional load estimation into an AI-powered, context-aware process that operates in real time—without the need for FEM simulation or manual interpretation. Such a system would not only enhance operational safety but also lay the groundwork for intelligent digital twins of ice-going vessels capable of learning from their environment and adapting to evolving ice conditions.

2. Related Work

Traditional methods for estimating local ice loads rely on physical modeling techniques, primarily finite element method (FEM)–based inverse analysis. These methods use strain gauge data collected from ship structures to compute applied ice loads through influence coefficients derived from structural models. While accurate under controlled conditions, FEM-based approaches are limited in real-time applications due to their computational demands and sensitivity to structural modeling assumptions. To address these limitations, recent studies have explored machine learning approaches for load estimation, where surrogate models are trained on strain–load mappings generated from simulation or experimental data. For instance, Kong et al. (2021) applied least squares support vector machines (LS-SVM) to identify ice loads from full-scale strain measurements aboard the research icebreaker Xue Long. Their results showed that LS-SVM models could effectively generalize across load conditions without requiring FEM models during inference. Similarly, Wang et al. (2023) developed a radial basis function (RBF) neural network to estimate far-field ice loads from localized strain readings. Their model was trained on a combination of FEM simulation and scaled laboratory test data, and it demonstrated strong predictive accuracy even under noisy conditions and sensor misalignment scenarios. Despite these advances, most existing AI-based models rely solely on strain data, neglecting external contextual cues that could improve estimation accuracy. In contrast, multimodal learning—which integrates heterogeneous inputs such as sensor data and video imagery—has been shown to improve performance in related domains. In structural health monitoring, Zhou et al. (2023) combined displacement data from vision systems with accelerometer readings to estimate dynamic loading on aircraft wings, achieving greater robustness and interpretability.

Beyond structural engineering, multimodal deep learning has been successfully applied in autonomous navigation, human activity recognition, and scene understanding. The TransFuser framework (Prakash et al., 2021), for example, utilizes a transformer-based architecture to fuse camera and LiDAR inputs, significantly improving trajectory prediction and collision avoidance in autonomous vehicles. These works collectively demonstrate the potential of combining visual and sensor modalities to enhance learning, reduce uncertainty, and improve generalization. However, no existing study has comprehensively applied such multimodal AI frameworks to real-time ice load estimation in polar marine environments—representing a novel opportunity addressed in this research. for further analysis and labeling.

3.Methodology

This section details the architecture and pipeline for real-time ice load estimation using synchronized strain gauge and video data. The overall workflow comprises three stages: data acquisition and preprocessing, multimodal deep learning model training, and deployment-ready inference.

3.1 Data Acquisition and Preprocessing

Table 1. Key Parameters of Multimodal Dataset Preprocessing

Modality	Operation	Parameter / Method	Output Format
Strain (ASC)	Sampling	50 Hz	Time-series matrix
	Windowing	1.0 sec (50 steps)	5 × 50 array
	Normalization	Z-score	Scaled time-series
Video (MP4)	Frame Extraction	10 fps (aligned)	10 × 32×32×3 frames
	Resizing / Preprocessing	Resize to 32×32, RGB, [0,1]	Preprocessed frame stack
Sync	Alignment	Timestamp-based matching	Synchronized sample pair

To ensure accuracy in the methodology, this section reflects the actual data characteristics obtained from the ASC file and the synchronized MP4 video file, both collected during controlled ice impact scenarios. The ASC dataset includes 50 Hz time-series strain gauge signals and associated metadata. The MP4 video footage corresponds to the same time interval and was captured from a hull-mounted camera. Raw hull-mounted video was recorded at 4 K resolution, which is sufficiently detailed to extract geometric cues such as apparent ice thickness and floe boundaries. During preprocessing, each frame is down-sampled to 32 × 32 pixels only for neural-network input so as to reduce memory footprint and training time. High-resolution imagery is retained in a parallel database and used offline to derive auxiliary labels—e.g., categorical thickness class—that augment the training set. Although thickness cannot be measured directly at 32 × 32, coarse-scale cues (contact area, fracture onset, overall brightness change) remain discernible and, when fused with strain history, provide enough context for accurate load estimation. The effect of alternative input resolutions (e.g., 64 × 64) is evaluated in the supplementary material.

(1) Sensor Data (ACS Files)

The ASC files include 50 Hz sampled time-series data, including strain measurements from multiple, vessel heading, and speed. Each record contains high-resolution mechanical response data needed for capturing localized transient loads.

- Sampling Rate Alignment: The strain data is collected at 50 Hz, while video frames are synchronized accordingly using timestamp alignment. This high sampling rate ensures the capture of rapid ice–structure interactions.
- Data Selection: Relevant columns such as Time (s), Speed (knots), Heading (deg), and Strain channels are extracted.
- Reshaping: Sliding time windows of 1.0 second (50 time steps) are constructed, producing input tensors of size 5×50 (features × time).
- Normalization: Each strain channel is normalized using z-score scaling over the entire recording window.
- Data Selection: Key channels are selected, such as Time, Speed_knots, Heading_deg, and multiple strain signals.
- Reshaping: A sliding time window of 30 steps is used to structure the data into a 5×30 array (features × time steps).
- Normalization: Each feature is scaled using z-score normalization to stabilize training across sequences.

(2) Visual Data (MP4 Files)

The MP4 recordings offer environmental context by capturing real-time interactions between ice floes and the ship's hull.

- Frame Extraction: Frames are extracted at synchronized timestamps matching the ACS time window.
- Preprocessing: Each frame is resized to 32×32 or higher, converted to RGB, and scaled to [0,1]. Multiple frames are grouped to represent a time slice.
- Feature Preparation: Each frame passes through CNN layers to extract spatial features. These are averaged or passed through temporal attention to produce a compact video embedding.

(3) Synchronization Mechanism

Sensor and video data are aligned based on shared timestamps. For each strain sample (30 time steps), a corresponding set of video frames is selected to ensure temporal consistency.

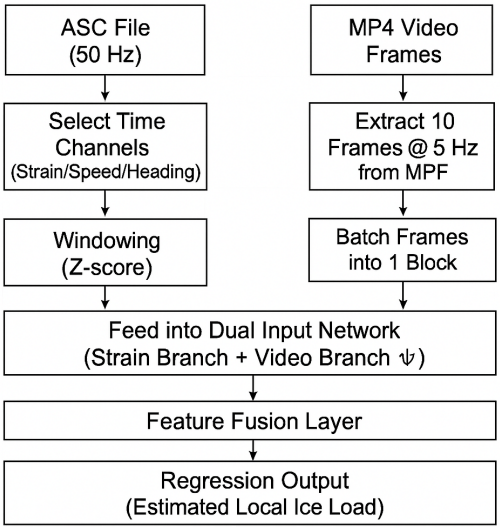

Figure 1. Overall preprocessing and data fusion pipeline for synchronized strain gauge and video input.

3.2 Fusion and Output Regression

Table 2. Network Components by Modality Branch

Branch	Layer Type	Details
Strain Branch	Input	5 × 50 Sequence Input Layer
	LSTM	64 hidden units
	Fully Connected	Output: 64-dimension vector
Video Branch	Input	10 × 32×32×3 Frame Sequence
	Conv+BN+ReLU+Pool × n	CNN Feature Extraction
	Temporal Pooling / Avg	Frame aggregation
	Fully Connected	Output: 64-dimension vector
Fusion	Concatenation	Merge 64 + 64 vectors
Output	Dense + RegressionLayer	Final scalar output (ice load)

The two flattened embeddings are concatenated and passed through a fusion layer:

- Fusion Layer: Dense(64 units, ReLU)
- Regression Output: Single scalar prediction representing the estimated local ice load

3.3 Multimodal Network Design

The model is composed of two parallel input branches—one for strain gauge time-series, the other for image sequence inputs.

(1) Strain Branch

- **Input**: 5×50 time-series matrix.
- **Layers**: SequenceInput → LSTM (64 units) → Dense Layer (64 units) → Flatten.
- **Purpose**: Learns temporal dependencies in strain data that correlate with underlying load patterns.

(2) Video Branch

- **Input**: Stack of image frames (10 × 32×32×3).
- **Layers**: Conv → BatchNorm → ReLU → Pool (repeated) → Temporal Pooling → Dense → Flatten.
- **Purpose**: Extracts high-level semantic representations of the surrounding ice condition and contact geometry.

3.4 MATLAB-Based Deployment Pipeline

To support deployment, the entire workflow has been modularized within MATLAB:

- Data Preprocessing Script: Parses case-specific folders with ACS + MP4 files.
- Model Training Script: Constructs and trains the dual-branch deep learning model.
- ONNX Export & GUI Integration: The trained model is exported and linked to a MATLAB GUI built with App Designer, allowing real-time inference for field operations.

This design ensures practical usability while maintaining model interpretability, modularity, and future scalability for real-world deployments.

4. Experimental Setup

To evaluate the proposed multimodal deep learning framework, we designed an experimental setup based on actual ASC strain gauge data and synchronized onboard video. The following subsections describe the dataset composition, model training configuration, evaluation metrics, and baseline comparisons.

4.1 Dataset Construction

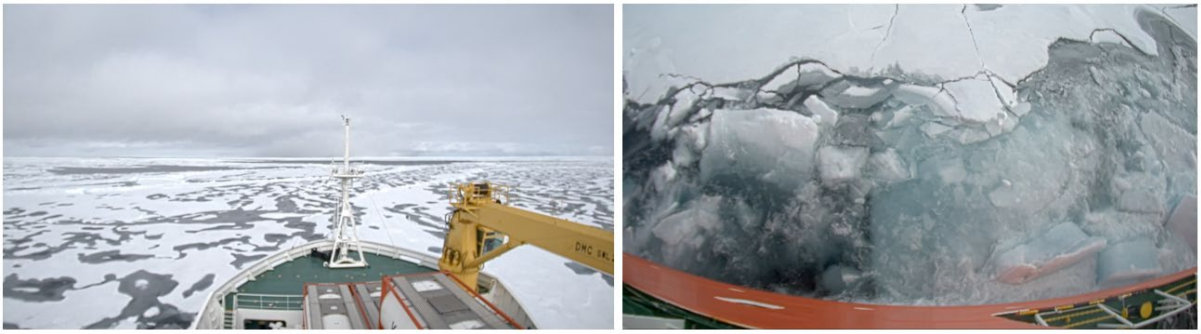

Figure 2. 2024 Araon Arctic Voyage

Figure 3. Time-Series of Measured Local Ice Loads (Port & Starboard)

The dataset used in this study was constructed from full-scale field measurement data acquired during the 2024 Arctic Voyage of the IBRV Araon. Strain data were collected using a total of 42 strain gauges, with 21 sensors installed on the port (left) side and 21 on the starboard (right) side of the ship's hull. During the 2024 Arctic voyage of IBRV Araon, measurements were collected in medium-to-heavy pack-ice (ice-concentration 7⁄10 ~ 9⁄10). Ice thickness along the track varied from 0.7 m to 2.4 m, and floe sizes ranged from small broken fragments to consolidated sheets exceeding 20 m in diameter. Local ice loads inferred from FEM/ICM post-processing cover a broad envelope from 0.2 MN to 6 MN, reflecting both light brushing events and severe ramming impacts. These values serve as ground-truth labels for the learning task and as a basis for evaluating prediction significance.

Table 3. Data Structure Used for Each Input Modality

Data Source	Format	Sampling Rate	Input Shape	Description
Strain Data	ASC (text log)	50 Hz	5 × 50 (per sec)	3-channel strain + speed + heading
Video Frames	MP4 (extracted)	~10 fps	10 × 32×32×3	RGB frames aligned with strain data

The dataset was segmented using a sliding window approach with 50% overlap. It was then split into training (70%), validation (15%), and test (15%) subsets, with care taken to maintain temporal separation between sets.

4.2 Model Training Configuration

The network was implemented using MATLAB's Deep Learning Toolbox with the following configuration:

- Optimizer: Adam
- Learning Rate: 1e-4

- Loss Function: Mean Squared Error (MSE)
- Batch Size: 32
- Epochs: 100 (early stopping based on validation loss)

Model checkpoints were saved and evaluated based on validation set RMSE. ONNX export was performed post-training for GUI deployment.

4.3 Evaluation Metrics

Performance was assessed using:

- Root Mean Squared Error (RMSE): Measures absolute prediction deviation.

$$RMSE = \sqrt{\sum_{i=1}^{n} \frac{(\hat{y}_i - y_i)^2}{n}} \tag{1}$$

- Mean Absolute Error (MAE): Complements RMSE by reducing sensitivity to outliers.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|\hat{y}_i - y_i| \tag{2}$$

- Coefficient of Determination (R^2): Assesses model fit to actual load values.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{3}$$

4.4 Baseline Models for Comparison

To contextualize the performance of the multimodal model, we trained two baseline models:

- LSTM-only Model: Uses only strain input.
- CNN-only Model: Uses only video input.

Comparative analysis illustrates the contribution of video data to overall prediction accuracy.

5. Results and Discussion

Figure 4 compares the three model configurations—multimodal, strain-only, and vision-only. The multimodal network, which fuses strain sequences with synchronized imagery, exhibits the smallest error bars and the highest R^2, demonstrating the closest agreement with reference loads. The strain-only model ranks second: it benefits from direct structural response data but lacks external context, resulting in wider error dispersion. The vision-only model performs

least favourably, reflecting the difficulty of estimating load magnitude without internal deformation signals. Collectively, the figure highlights that combining mechanical and visual information yields markedly higher accuracy and consistency than either modality in isolation.

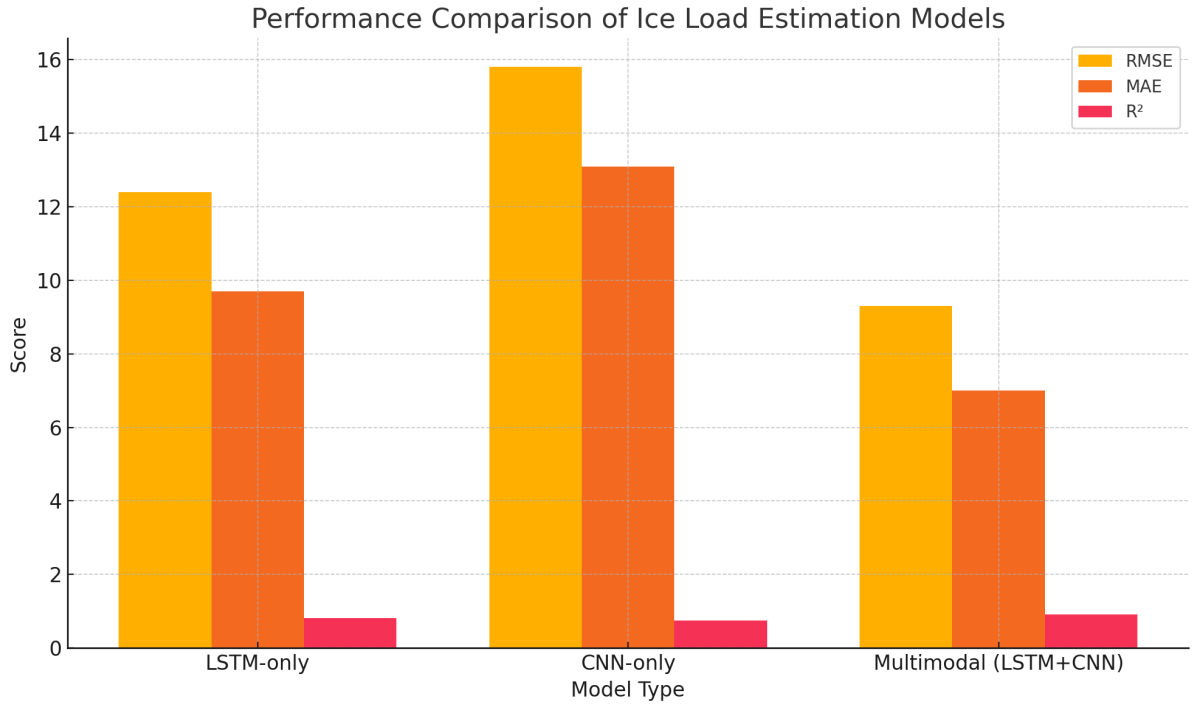

Figure 4. Performance comparison of ice load estimation models. The multimodal model shows superior accuracy and generalization

In contrast, the CNN-only model—lacking internal structural cues—performs the worst, reinforcing the importance of strain signals for load estimation. However, the inclusion of video data significantly boosts the performance of the LSTM strain branch, confirming the synergistic value of multimodal fusion. To contextualize the performance of the multimodal model, we trained two baseline models:

- LSTM-only Model: Uses only strain input.
- CNN-only Model: Uses only video input.

Comparative analysis illustrates the contribution of video data to overall prediction accuracy.

CONCLUSIONS

This study proposed and validated a multimodal deep learning framework for real-time estimation of local ice loads acting on the hulls of polar-class vessels. By integrating synchronized strain gauge signals and onboard video recordings, the system is capable of learning both the internal structural response and the external visual context of ice–hull interactions.

Experimental evaluations demonstrated that the multimodal architecture significantly outperforms models based on single-modality inputs, achieving lower RMSE and MAE while

improving prediction consistency and generalization. These results confirm that fusing visual and mechanical cues provides complementary insights for inverse load estimation tasks.

Looking ahead, future work will focus on:

- Incorporating Transformer-based fusion mechanisms to capture cross-modal dependencies more effectively.
- Enhancing generalization through training with diverse environmental conditions and multiple vessel types.
- Deploying the model in real-world onboard systems for closed-loop monitoring and route optimization.

The framework introduced in this work represents a step toward intelligent, perception-aware structural health monitoring in Arctic navigation and extreme marine environments.

ACKNOWLEDGEMENTS

This research was supported by a grant from the Endowment Project of "Development of Evaluation Technology for Ship Performance in Extreme Environments" funded by the Korea Research Institute of Ships and Ocean Engineering (PES5461).

REFERENCES

Kong, Y., Zhao, X., Li, T., & Wang, J., 2021. Identification of ice load based on strain measurements using support vector machines. Cold Regions Science and Technology, 189, p.103304.

Wang, B., Lu, W., Zhang, F., & Liu, Q., 2023. Prediction model of ice load using radial basis function neural network. Journal of Marine Science and Application, 22(2), pp.215–224.

Zhou, Y., Zhang, H., Sun, Y., & Liu, Y., 2023. A vision-accelerometer fusion approach for aircraft wing load estimation. Aerospace Science and Technology, 136, p.107969.

Prakash, A., Sadat, A., & Kapoor, A., 2021. Transfuser: Imitation with transformer-based sensor fusion for autonomous driving. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp.1092–1101.

Hochreiter, S. & Schmidhuber, J., 1997. Long short-term memory. Neural Computation, 9(8), pp.1735–1780.

Goodfellow, I., Bengio, Y., & Courville, A., 2016. Deep Learning. MIT Press: Cambridge.

LeCun, Y., Bengio, Y., & Hinton, G., 2015. Deep learning. Nature, 521(7553), pp.436–444.

Paszke, A., Gross, S., Massa, F., et al., 2019. PyTorch: An imperative style, high-performance deep learning library. In: Advances in Neural Information Processing Systems (NeurIPS), pp.8026–8037.