Athens Journal of Sciences- Volume 12, Issue 2, June 2025 – Pages 71-92

High Speed Craft Peak Acceleration Prediction using Machine Learning Models

By Chunan Zhang[*], Xiulei Cao[±], Michael Chen[°] & Hongmei Zhu[•]

The ability to predict the peak impact of slams, which occur during high-speed motion of planing watercraft, could lead to significant opportunities for improving the safety of passengers and equipment aboard. We apply Random Forest and Convolutional Neural Networks to achieve this goal using peak acceleration, freefall duration, and velocity change from high-speed crafts to quantify the amplitude of wave impact. We analyze different characteristics separately in the freefall and impact regions for each wave. We report numerical results of both models, demonstrating the accuracy of these predictions.

Keywords: *high-speed craft, machine learning, random forest, convolutional, neural networks*

Introduction

The impacts of waves on high-speed working watercraft can be highly life-threatening for passengers and cause serious damage to equipment on board. The collision of planing hulls and waves produces mechanical shocks that can result in acute and chronic injuries, including major spinal fractures for passengers (Kumar & Saran 2016, Gohari et al. 2014, Ensign et al. 2000). Adaptive mechanical shock-mitigating suspension seats have the potential to protect passengers by adjusting the suspension system when shocks occur (Wice 2015, Riley et al. 2010, Gohari & Tahmasebi 2014, Heidarian & Wang 2019). To implement mechanical suspension adjustment, an algorithm executed by a microprocessor must be deployed with the system so that the amplitude of incoming waves can be measured, so the system can make the necessary adjustments in real-time. Several studies (Ji 2015, Gohari et al. 2014) have applied neural network machine learning models for analyzing suspension seats. In Ji (2015), an efficient neural network (NN) algorithm was constructed to identify the vibration attenuation properties of five suspension seats. Each of the NN seat models predicted vertical seat-pan Root-Mean-Square (RMS) accelerations based on chassis accelerations and a measure of driver anthropometrics. In Gohari & Tahmasebi (2015), the vibration transmitted from vehicles to the driver's body over the long term was addressed by the active force control loop, which integrates an artificial neural network to approximate the mass of the seat and human body. In Gohari et al. (2014), a new artificial neural network model was introduced that could predict spine acceleration from excitation signals and human body mass and height. The model achieved 96% accuracy, making it useful for real-time and

[*]PhD Student, York University, Canada.
[±]Postdoctoral Researcher, York University, Canada.
[°]Professor, York University, Canada.
[•]Professor, York University, Canada.

https://doi.org/10.30958/ajs.12-2-1 doi=10.30958/ajs.12-2-1

offline analysis. Additionally, the study optimized off-road seat suspension using the developed neural network model and three meta-heuristic algorithms.

A recently developed ruggedized marine data acquisition system, equipped with accelerometers and displacement sensors, can monitor shock and vibration. Using this system, the acceleration and seat displacement of a working or laboratory watercraft can be measured and collected at the millisecond level. This data can be used to build predictive models for wave impact. Previous technical reports by the US Navy describe the dynamics of watercraft hitting waves and important characteristics of significant waves (Riley et al. 2014). Wave impacts are dominated by downward acceleration, as a crucial cause of spinal fractures is the collision in the vertical direction. The integration of generalizable force-correcting machine learning methods has shown promising results in predicting ship responses in varying sea states, further supporting the feasibility of data-driven maritime impact prediction (Marlantes 2024). However, the relationship between injury severity and the magnitude of downward acceleration varies with hull design, operating conditions, passenger mass, and other factors. Due to confidentiality considerations, such information is not publicly available to researchers. This variability presents a challenge but also encourages researchers to use artificial intelligence tools to build predictive models.

Researchers study the dynamics of individual waves using vertical acceleration signals (Wice 2015, Salehpour et al. 2011), and analyze wave impacts (Coats & Riley 2018, Riley et al. 2014, Riley et al. 2018). Consequently, features commonly extracted from watercraft are examined in these models. Several articles have applied machine learning methods to predict vibrations or signals in various applications (Kumar & Saran 2016, Kavsaoğlu et al. 2015, Lepine & Rouillard 2018, Ji 2015, Lin et al. 2020, Gohari et al. 2014, Gohari & Tahmasebi 2014). Our contribution is focusing specifically on predicting wave impacts from wave features using machine learning models tailored for high-speed craft. Recent studies have also explored machine learning techniques to predict planing hull motions, improving seakeeping performance and safety in high-speed crafts (Shehata 2023).

This study presents Random Forest (RF) for classification and Convolutional Neural Networks (CNNs) for regression. For RF, severe waves that cause physical injury can be classified by a predefined threshold of the wave impact. RF model requires less computational power and memory space for implementation in onboard microprocessor. For CNNs, wave severity is directly represented as the wave impact. CNNs with a common mean square error function could achieve a higher accuracy. Furthermore, we propose a new loss function for CNNs to further improve its model robustness.

Literature Review

This literature review summarizes key studies related to vibration and shock analysis, suspension systems, high-speed craft studies, and the application of machine learning and artificial intelligence in these domains.

Vibration analysis and shock detection are crucial for vehicle and equipment safety, particularly in road and marine environments. Evaluating shock detection algorithms for road vehicle vibration analysis involves using machine learning to combine various shock detection methods, improving reliability and accuracy (Lepine & Rouillard 2018). The ability to characterize vibrations transmitted from seats to the human body has significant implications for health, particularly regarding low back injuries among vehicle operators (Kumar & Saran 2016). Additionally, the evaluation of suspension seats under multi-axis vibration excitations using neural networks helps in predicting seat performance and reducing vibration exposure (Ji 2015). Studies on off-road vehicle seat suspension optimization use artificial neural network models to predict spine acceleration in vertical vibration, emphasizing the importance of seat suspension design in mitigating vibration impacts (Gohari et al. 2014). Meta-heuristic optimization algorithms are also applied to improve seat suspension systems, highlighting the effectiveness of genetic algorithms, particle swarm optimization, and harmony search in reducing transmitted vibrations (Gohari & Tahmasebi 2014).

Suspension systems play a vital role in enhancing ride comfort and safety in vehicles. Optimal selection of active suspension parameters using artificial intelligence involves multi-objective genetic algorithms to balance conflicting objectives, such as minimizing vertical acceleration and relative displacement (Salehpour et al. 2011). Adaptive artificial neural network (ANN) surrogate models for nonlinear hydraulic adjustable dampers in semi-active suspension systems demonstrate the benefits of neural networks in improving damping performance and reducing vibration (Lin et al. 2020). Active off-road seat suspension systems using intelligent active force control (AFC) integrated with ANN offer robust vibration control, particularly for heavy-duty vehicles (Gohari & Tahmasebi 2015). AFC techniques are also applied to control the roll movement of spray boom structures, showing superior performance in attenuating unwanted oscillations compared to traditional Proportional-Integral-Derivative (PID) controllers (Tahmasebi et al. 2013).

Research on high-speed craft often focuses on practical measures to enhance safety and performance. Studies evaluating the effectiveness of shock isolation seats in marine environments use laboratory tests to demonstrate their ability to reduce wave impact loads before installation in high-speed planing craft (Riley et al. 2018). Comparing shock isolated seats to rigid seats during wave impacts reveals that shock isolation can mitigate severe impacts, even when dynamic amplification occurs (Coats & Riley 2018). These studies provide essential insights into the practical implementation and benefits of shock mitigation technologies in high-speed marine craft.

Statistical modeling and machine learning are increasingly applied to enhance the safety and performance of high-speed craft. Spatial dynamic modeling of high-speed craft suspension seating uses a Newton-Euler approach to predict vertical acceleration responses, offering a flexible base for future examination of suspension systems (Wice 2015). A generalized approach for computing accelerations using full-scale high-speed craft trials data helps achieve repeatable and accurate analyses across different data analysts (Riley et al. 2010). Acceleration responses mode decomposition for quantifying wave impact load on high-speed planing craft provides valuable parameters such as peak acceleration and impact duration, essential for naval architecture and marine engineering investigations (Riley et al. 2014). Ride severity profiles for evaluating craft motions introduce new criteria for human comfort and performance, crucial for improving ride quality in marine environments (Riley et al. 2015).

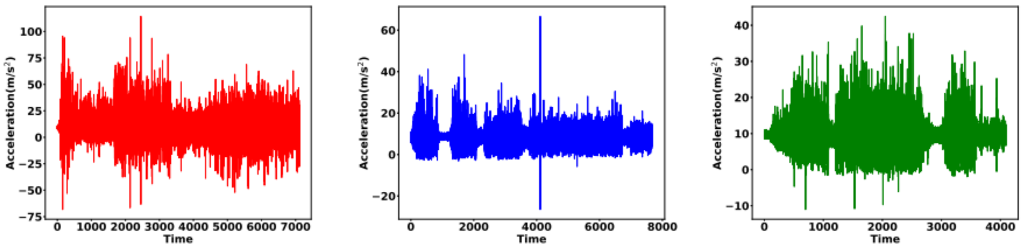
In summary, the extensive research on vibration and shock analysis, suspension systems, and the application of machine learning and artificial modeling across vehicles, marine, and other domains underscore the significant advancements and potential of these techniques in enhancing safety, comfort, and performance. The integration of advanced algorithms and models continues to drive innovation, offering promising solutions to complex challenges in various industries.

Materials

Acceleration Data

Three data sets are collected from three boats, referred to as boats A, B, and C respectively. Deck-mounted accelerometers are typically installed to record accelerations in three directions: horizontal, transversal, and vertical. Vertical acceleration is particularly crucial for the health and safety of the boat occupants (Wice 2015, Coats & Riley 2018, Riley et al. 2014, Riley et al. 2018). Therefore, in this paper, only vertical acceleration is analyzed. The vertical acceleration records for boats A, B, and C are presented in Figure 1.

Figure 1. *Vertical Acceleration for Boats A, B, and C*



(a) Boat A (b) Boat B (c) Boat C

In common practice, the time series of the accelerations are normalized by dividing by 9.8 m/s², which is the acceleration due to gravity. Normalizing the data in this manner converts the acceleration values into units of g (where 1g = 9.8 m/s²). By using unit g, it becomes straightforward to assess the severity of the accelerations in a way that is intuitive and easily understandable, particularly when discussing the impact on human health and equipment safety. Normalizing the data helps in maintaining consistency in analysis and comparison across different studies and practical applications. The varying weight of the boats, with boat A being the lightest and most volatile and boat C being the heaviest and least volatile, further highlights the differences in acceleration patterns and their potential impacts.

Data Filtering

To accurately predict peak impacts, it is essential to filter out high-frequency noise from the acceleration signals. Several low pass filtering techniques can be applied to achieve this (Coats & Riley 2018, Riley et al. 2010, Riley et al. 2014, Riley et al. 2015, Riley et al. 2018, Zhao & Wang 2019). In this study, we used different filtering methods to illustrate their effects on the acceleration signals.

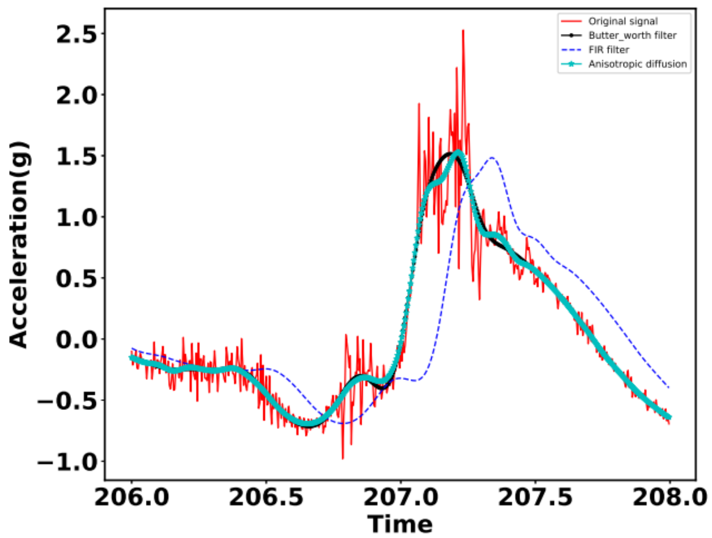
The original acceleration signal is shown by the red curve in Figure 2. The black curve illustrates an estimate of the rigid body acceleration using a 5 Hz, 4th-order Butterworth filter (Riley et al. 2010, Riley et al. 2014, Riley et al. 2018). This filter is known for its smooth frequency response and minimal phase distortion, making it suitable for applications where maintaining the integrity of the signal's shape is important. The Butterworth filter is chosen for its ability to provide a flat frequency response in the passband and a smooth transition to the stopband, ensuring that the signal's essential characteristics are preserved while noise is effectively filtered out.

The blue dashed line represents the output of a 5 Hz Finite Impulse Response (FIR) filter with coefficients of length 67 (Coats & Riley 2018). FIR filters are widely used due to their linear phase response, which ensures that all frequency components of the signal are delayed by the same amount. However, this can introduce a significant time delay, as noted by the time shift of approximately 0.268 seconds in the filtered signal.

Additionally, an Anisotropic Diffusion filter is applied, shown by the green curve in Figure 2. Anisotropic Diffusion filtering is a technique that reduces noise while preserving significant features of the signal, such as edges or abrupt changes. This filter adapts the smoothing process to the local structure of the signal, making it particularly effective for retaining the sharpness of the wave slam events.

The peak acceleration for the filtered wave slam event using these methods is reduced to 1.5 g, compared to 2.5 g in the unfiltered signal. This demonstrates the effectiveness of the filters in removing high-frequency noise while preserving the essential characteristics of the signal.

Figure 2. *The Effects of Low-Pass Filtering*



Low-pass signal filtering can be tuned to different frequency spectra. After inspecting many records for various small crafts at different speeds, it was found that a 10 Hz low-pass filter removes sufficient high-frequency oscillation without excessively removing the peak rigid body response signal. However, for this study, a 5 Hz low-pass filter was chosen because it better correlates the features in the freefall region with the peak impact in the impact region.

The rationale for using a 5 Hz filter is supported by previous research. In Coats & Riley (2018), the authors presented Fourier spectra for vertical accelerations at the Longitudinal Center of Gravity (LCG) recorded on three different crafts at varying speeds. They confirmed that the largest spectral amplitudes correspond to frequencies below 10 Hz. This practical evidence suggests that rigid body wave encounters are predominantly below the 2 Hz threshold, making the 5 Hz low pass filter a reasonable choice. Figure 2 illustrates that the Butterworth filter and Anisotropic Diffusion filter effectively capture the preliminary profile of the original acceleration signal. While the FIR filter produces a similar profile, the time delay it introduces can be problematic for real-time applications.

The objective of this paper is to apply machine learning methods to predict peak impacts in the impact region based on features in the freefall region. Once the predicted peak impact is obtained, the damping factor of the seats on the boat can be adjusted rapidly. To ensure sufficient time for prediction and adjustment of the damping factor, the FIR filter, despite its advantages, may not be the best choice due to its inherent time delay.

In summary, the choice of filtering method depends on the specific requirements of the application. The Butterworth filter and Anisotropic Diffusion filter are preferred in this study due to their ability to preserve signal integrity and provide timely responses, crucial for the real-time adjustment of damping factors on high-speed boats.

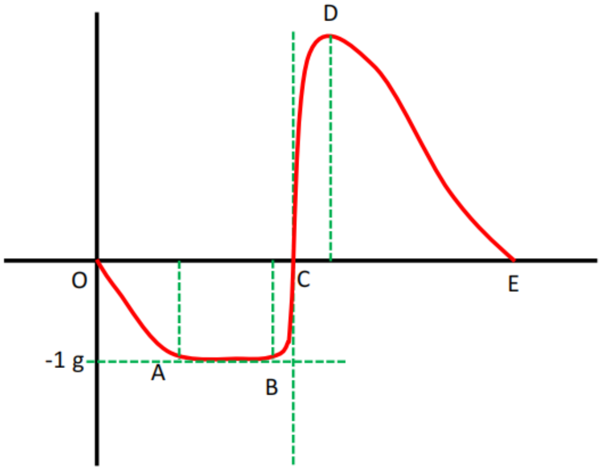
Feature Extraction Process

During the feature extraction process from the filtered acceleration signal, several key features are derived. These features are crucial for accurately predicting peak impacts using machine learning methods. The primary features extracted include the maximum value of the first derivative, the maximum value of the second derivative, the first integration of the filtered acceleration signal (velocity), and the double integration of the filtered acceleration signal (displacement). Additionally, specific features such as the minimum of the first derivative, maximum of the second-order derivative, freefall duration, minimum value in the plateau, and the area of the freefall region are obtained. These features are used as inputs for prediction models.

Definition of Features

For one wave, the acceleration signal can be divided into two main parts: the freefall region and the impact region. Below are the definitions of the features used in this paper.

Figure 3. *One Wave Plot*



Freefall Region

The freefall region is the section of negative acceleration between zero-crossings. In Figure 3, this corresponds to the part between points O and C.

- Freefall Duration 1 (sec): The time from the downward acceleration zero-crossing before the wave impact until the upward acceleration zero-crossing, represented by the interval between points O and C.
- Freefall Duration 2 (sec): The time from the downward acceleration zero-crossing before the wave impact until the slam occurs, represented by the interval between points O and B.
- Plateau Duration (sec): The time during which the acceleration is around -1g, if present, represented by the interval between points A and B.
- Freefall Min: The minimum downward acceleration before the impact, occurring around point A in Figure 3.
- Freefall Area: The cumulative integral of the acceleration signal during the freefall section.
- Derivative 1st Order: The first-order derivative of the signal with respect to time.
- Derivative 1st Order Min: The minimum value (most negative) of the first-order derivative during the freefall section.
- Derivative 2nd Order: The second-order derivative of the signal with respect to time.
- Derivative 2nd Order Max: The maximum value of the second-order derivative during the freefall section, typically occurring immediately after exiting the plateau region or between Freefall Min and the next zero-crossing.

Impact Region

The impact region is the section of positive acceleration between zero-crossings.

- Peak Impact: The maximum positive acceleration during the impact phase.

Correlation of Features in Freefall and Impact Regions

Correlation is a statistical technique that can show whether and how strongly pairs of variables are related. In this section, the correlation of features in the freefall and impact regions is presented. Understanding these correlations can support the selection of features for the prediction models. Tables 1-3 present the correlation coefficients for Boats A, B, and C respectively, with the peak impact larger than 1g.

Table 1. *Feature Correlation Under 5 Hz Using the Butterworth Filter for Boat A*

	peak_impact
freefall_min	-0.51
plateau_duration	0.28
freefall_duration_1	-0.12
freefall_duration_2	-0.14
freefall_area	-0.35
derivative_1st_order_min	-0.33
derivative_2nd_order_mzx	0.72

Table 2. *Feature Correlation Under 5 Hz Using the Butterworth Filter for Boat B*

	peak_impact
freefall_min	-0.56
plateau_duration	0.51
freefall_duration_1	-0.25
freefall_duration_2	-0.30
freefall_area	-0.52
derivative_1st_order_min	-0.72
derivative_2nd_order_mzx	0.84

Table 3. *Feature Correlation Under 5 Hz Using the Butterworth Filter for Boat C*

	peak_impact
freefall_min	-0.61
plateau_duration	0.58
freefall_duration_1	0.13
freefall_duration_2	0.27
freefall_area	-0.58
derivative_1st_order_min	-0.34
derivative_2nd_order_mzx	0.78

These tables provide insights into the relationships between features extracted from the freefall region and peak impact. By analyzing these correlations, more informed decisions can be made regarding which features to use for the Random Forest prediction model, thereby enhancing its accuracy and reliability.

Methodology

In this section, several machine learning methods are employed to predict the peak impact using the filtered acceleration signals. Both classification and regression techniques are utilized. Specifically, Random Forest is used for classification, while two Convolutional Neural Network (CNN) models are applied for regression. All these methods are implemented using Python.

Classification Methods

Severe waves that cause physical injury can be classified by a predefined threshold of the wave impact; as such, transforming the regression problem into a classification problem. In this approach, a Random Forest model is applied to classify the peak impacts. This model utilizes a subset of extracted features as described earlier.

<u>Random Forest</u>

Random Forest is a supervised learning algorithm that leverages ensemble learning for classification tasks. This method constructs multiple decision trees during training, with each tree operating independently and without interaction with others. The final output is determined by the mode of the classes.

In this study, the Random Forest algorithm is employed for classification to predict peak impacts based on features extracted from the freefall region of the acceleration signals. The key features considered are the minimum value of the first-order derivative, the maximum value of the second-order derivative, the freefall duration, and the cumulative integral during the freefall section. The first-order derivative minimum indicates the steepest descent in acceleration, while the second-order derivative maximum represents the highest rate of change in acceleration. The freefall duration measures the time interval between zero-crossings in the downward acceleration phase, and the freefall area quantifies the overall impact energy through the cumulative integral of acceleration during the freefall section.

To construct the Random Forest model, multiple decision trees are built using subsets of the data and features. Each tree is trained independently, learning to classify peak impacts based on the provided features. During the classification process, each tree votes for a class, and the final prediction is made based on the majority vote. This ensemble approach helps to reduce overfitting and improves the generalization capability of the model.

The training process involves dividing the dataset into training and validation sets, ensuring that the model learns effectively and generalizes well to unseen data. Hyperparameters such as the number of trees, maximum depth, and minimum samples per leaf are carefully tuned to optimize performance. The model's effectiveness is evaluated using confusion metrics, which provide a comprehensive assessment of its classification capabilities. The confusion matrix includes details on true positives, false positives, true negatives, and false negatives, offering a clear insight into the model's performance in correctly classifying wave impacts.

In this application, the Random Forest algorithm offers a robust and interpretable method for classifying peak impacts. It serves as a benchmark for comparison with more complex models like Convolutional Neural Networks (CNNs), highlighting its utility in handling the variability and complexity of wave impact data on high-speed crafts.

Regression Methods

Since wave severity is directly represented as the wave impact, this section approaches the prediction of continuous peak impact, necessitating the use of regression methods. This approach allows for a more precise and nuanced prediction of wave impacts, capturing the continuous nature of the severity levels rather than categorizing them into discrete classes. We construct two Convolutional Neural Network (CNN) models to achieve this. The first CNN model assumes that all peak impacts are equally important, and thus it employs the mean square error (MSE) as its loss function. This approach treats each prediction error equally, focusing on minimizing the overall deviation between predicted and actual values. However, in real-world applications, higher peak impacts are more critical to the safety and comfort of occupants on the boats. Therefore, the second CNN model incorporates a customized loss function designed to give more weight to higher peak impacts. This weighted loss function ensures that the model pays greater attention to accurately predicting larger impacts, which are more likely to cause significant damage or injury. By prioritizing these critical events, the second model aims to improve the practical utility of the predictions in real-time safety applications.

<u>Convolutional Neural Networks</u>

Convolutional Neural Networks (CNNs) are a class of deep learning models that are particularly well-suited for tasks involving spatial hierarchies, such as image recognition and time-series analysis. They operate by applying convolutional filters to input data, which allows the network to detect local patterns that are invariant to translation. These patterns are then combined across multiple layers to form more complex representations, enabling the CNN to learn intricate structures within the data.

Convolutional layers apply a set of convolutional filters to the input data. Each filter scans the input signal to detect specific patterns, such as peaks and troughs in acceleration data. The convolution operation preserves the spatial relationship between data points by learning filter coefficients that activate when specific patterns are detected. After each convolutional layer, an activation function, typically ReLU, is applied to introduce non-linearity into the model. This helps the network to learn complex patterns and interactions between the input features. Pooling layers, such as max pooling, are used to reduce the dimensionality of the data by down sampling the output of the convolutional layers. This process helps in reducing the computational load and mitigating the risk of overfitting by summarizing the presence of features.

Towards the end of the network, fully connected layers are added to combine the features learned by the convolutional and pooling layers. These layers perform high-level reasoning based on the detected patterns and provide the final output of the network. In a regression task, the output layer typically consists of a single neuron that provides the predicted value of the peak impact. The output is a continuous value representing the intensity of the predicted impact.

During each iteration of backpropagation training, each neuron in the network computes an output value based on its inputs. The parameter sensitivities, or gradients, are then used to iteratively update the CNN parameters until a certain stopping criterion, such as minimal loss or a maximum number of iterations, is achieved. This process enables the network to learn from the data and improve its predictions over time.

We define two CNN models in this section: the first is a CNN with a uniform weight, where the loss function treats all prediction errors equally. The second model uses a weighted loss function that assigns greater importance to higher peak impacts, thus improving the model's performance in predicting critical, high-impact events. This dual approach allows us to compare the effectiveness of different loss functions in enhancing the accuracy and reliability of peak impact predictions.

Convolutional Neural Network with MSE

In this approach, we treat all the peak impacts equally, applying the mean squared error (MSE) as the loss function. MSE is commonly used in regression tasks and calculates the average of the squares of the errors—that is, the difference between the predicted and actual values. This loss function ensures that all prediction errors are treated the same, focusing on minimizing the overall deviation across all predictions. Mathematically, the MSE loss function can be defined as:

$$loss_{MSE} = \frac{1}{n}\sum_{i}^{n}(y_{i,pred} - y_{i,true})^2$$

Convolutional Neural Network with Logistic Function

In real-world applications, particularly for high-speed crafts, predicting higher peak impacts is more critical than lower ones. Therefore, we designed a customized loss function that assigns greater importance to higher peak impacts. This approach involves defining two loss functions for the CNN model: the traditional mean squared error (MSE) and a customized logistic function-based loss.

For the CNN model, the customized loss function is designed to add more weight to higher peak impacts, ensuring that the model focuses more on accurately predicting these critical events. This is crucial for applications where larger impacts pose greater risks to passenger safety and equipment integrity. The customized loss function can be defined as:

$$loss_{customized} = \sum_{i}^{n}\left(y_{i,pred} - y_{i,true}\right)^2\left(\frac{1}{1 + e^{-\mathrm{k}(y_{i,true} - n)}}\right)$$

where k=7 is a constant that determines the steepness of the weight function, and n is the threshold value above which the impacts are considered high. The customized loss function increases the weight for errors corresponding to true values greater than n, thus prioritizing the accurate prediction of high peak impacts.

This dual approach, using both MSE and the customized loss functions, allows us to compare the effectiveness of these loss functions in enhancing the model's ability to predict peak impacts. By focusing on higher peaks, the customized loss function aims to improve the practical utility of the model, making it more suitable for real-time safety applications in high-speed crafts.

Results and Discussion

This section presents results calculated by different machine learning methods: Random Forest for classification models and two Convolutional Neural Network models for regression.

Random Forest Classification

To predict the peak impact for all the waves in the impact region, several features from the freefall region are selected. In this study, five features are chosen: the minimum value in the freefall region (freefall_min), the minimum of the first-order derivative during the freefall region (freefall_der1nd_min), the maximum of the second-order derivative during the freefall region (freefall_der2nd_max), the area of the freefall region (freefall_area), and the duration between two consecutive zero crossings in the freefall region (freefall_duration).

For each boat, 85% of the data is randomly selected as the training set, and the remaining 15% is used as the testing set. In this paper, the stratified sampling method is applied to ensure that the training and testing sets maintain the same distribution of the target variable, which is crucial given the imbalanced nature of the data. The stratified sampling method used in this paper is defined as follows: assume that for a boat, the peak impacts in the remaining set vary from 0g to 3g. This set is then uniformly divided into six subsets with peak impact ranges of (0g to 0.5g), (0.5g to 1g), (1g to 1.5g), (1.5g to 2g), (2g to 2.5g), and (2.5g to 3g).

As demonstrated in Figure 4 - accuracy and confusion matrix for boats A, B, and C, the accuracy for the test dataset is 96.67% for boat A, 98.33% for boat B, and 99.31% for boat C. These accuracy values indicate that the Random Forest algorithm can predict the peak impact with a high degree of accuracy. It is important to note that boat A is the lightest boat, making its peak impact the most volatile. This increased volatility in peak impacts for boat A results in slightly lower accuracy compared to boats B and C. This demonstrates the model's sensitivity to the variability in the data, with more consistent data leading to better predictive performance.

Figure 4. *Accuracy and Confusion Matrix for Boats A, B, and C*

Boat A

Accuracy for Test Dataset: 96.67%

	Predicted: <= 1G	Predicted: > 1G	
Actual: <= 1G	1225	20	1245
Actual: > 1G	24	54	78
	1249	74	

Boat B

Accuracy for Test Dataset: 98.33%

	Predicted: <= 1G	Predicted: > 1G	
Actual: <= 1G	2149	15	2164
Actual: > 1G	22	30	52
	2171	45	

Boat C

Accuracy for Test Dataset: 99.31%

	Predicted: <= 1G	Predicted: > 1G	
Actual: <= 1G	995	1	996
Actual: > 1G	6	8	14
	1001	9	

In the context of feature importance, relative feature importance ranking for each model is extracted and presented as shown in Figure 5.

The maximum of the second-order derivative during the freefall region (freefall_der2nd_max) consistently emerges as a critical feature across all models, particularly for Boats A and B. This feature's significance highlights its role in capturing the rapid changes in acceleration that are indicative of peak impacts. Other features like freefall duration, peak freefall, and freefall area also contribute significantly but vary in their importance across the different boats. This variation in feature importance further illustrates the nuanced differences in how peak impacts manifest in boats of different weights and designs.

Figure 5. *Relative Feature Importance Ranking*

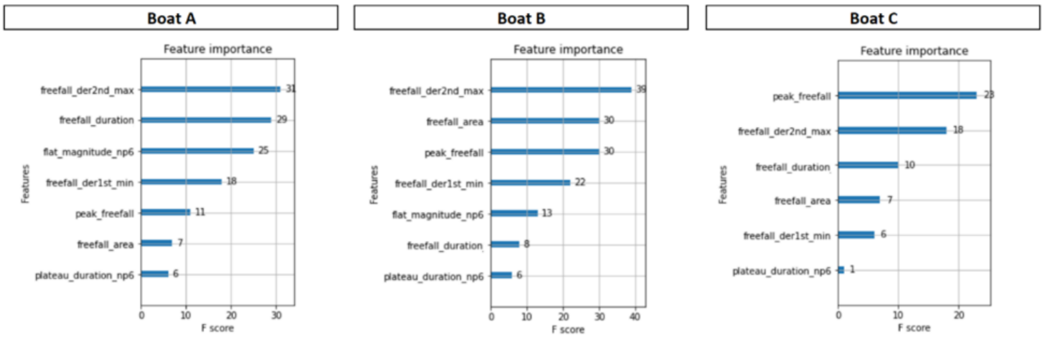
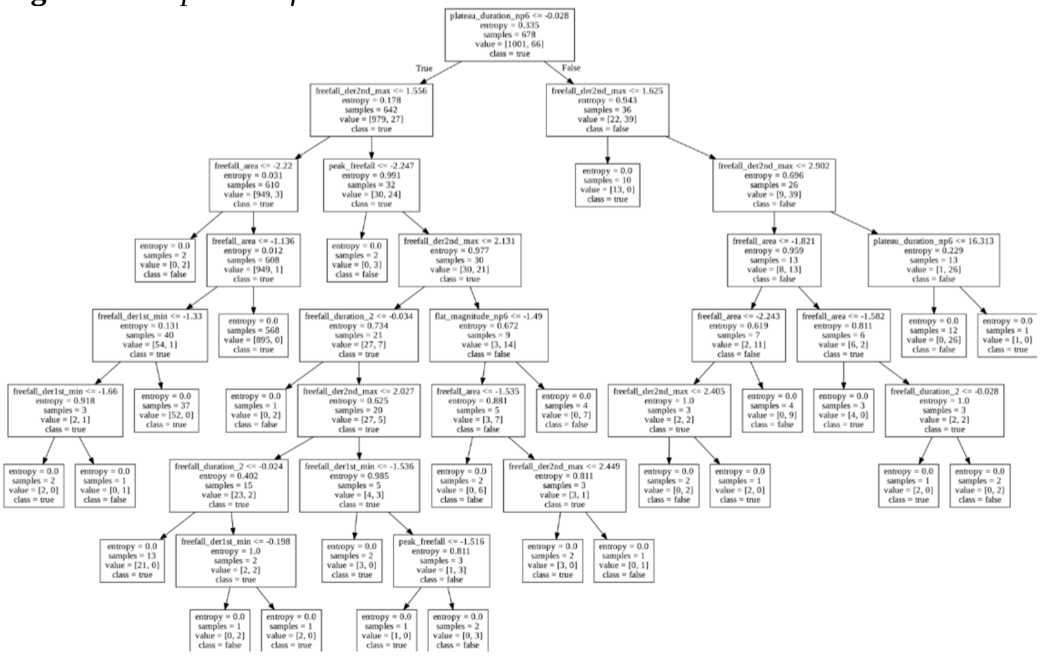


Figure 6 illustrates a representative decision tree from the Random Forest model applied to Boat A's dataset. This decision tree exemplifies how the model partitions data based on key features to predict peak impact severity. At each node, the tree evaluates specific features such as freefall minimum acceleration or freefall duration—and determines a threshold to split the data. For instance, a node might assess whether the freefall minimum acceleration exceeds a certain value, directing the data down different branches based on this criterion. This hierarchical structure enables the model to capture complex interactions between variables, leading to accurate classifications of impact severity levels. By aggregating the predictions from numerous such trees, the Random Forest model enhances its robustness and generalization capabilities, effectively predicting peak accelerations experienced by high-speed crafts. The visualization in Figure 6 provides insight into the model's decision-making process, highlighting the importance of specific features and thresholds in classifying impact events.

Figure 6. *Sample Tree for Boat A*



To summarize this section, the detailed analysis provided by the confusion matrices and feature importance rankings offers valuable insights into the model's performance and the underlying data characteristics. These results not only validate the use of Random Forest for this application but also highlight areas for potential improvement and further investigation, particularly in understanding the impact of boat weight and design on the accuracy of peak impact predictions.

Convolutional Neural Networks (CNN)

In this subsection, Convolutional Neural Networks (CNNs) are applied to the filtered acceleration signals. The peak impacts and the corresponding 0.2 seconds ahead of slam series are collected, with the time ahead of each slam set to 0.2 seconds as determined by the designers. These time series ahead of the slam are used as inputs for the CNN model, with the output being the peak impacts.

To train and test the CNN model, 10% of the filtered data for each boat is randomly selected as the testing set. The remaining 90% is divided into two parts: 85% for training and 15% for validation. Given the significant variation in peak impacts (e.g., from 0g to 3g), stratified sampling is employed to ensure the CNN model learns from all impact ranges. The same stratified sampling method as presented in the Random Forest (RF) model is used. This ensures that the distribution of wave impact severity levels is consistent across both the training and testing datasets, thereby improving the robustness and reliability of the regression model's predictions. By maintaining the same sampling strategy, we ensure comparability and consistency in evaluating the model's performance. In each subset, 85% of the data is randomly selected for training and 15% for validation.

In real applications, high peak impacts (e.g., larger than 1g) are more critical to the safety and comfort of occupants on the boats. Therefore, the CNN model in this paper employs two different loss functions. Using the MSE loss function, the CNN model is trained to minimize the average of the squares of the errors between predicted and actual values, treating all errors equally. However, recognizing the practical importance of accurately predicting high peak impacts, a second CNN model is trained with a customized loss function that assigns greater weight to higher impacts, specifically those above a certain threshold n.

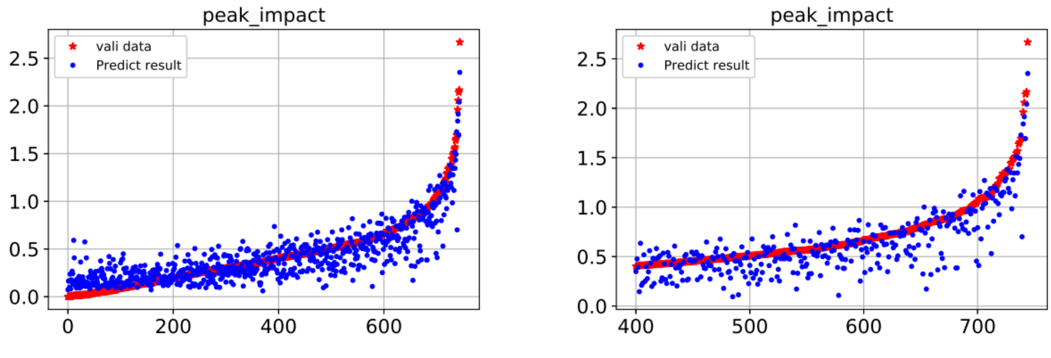
The experimental results (Figures 7-18) presented below compare the actual versus predicted values for all boats using both the Mean Squared Error (MSE) and a customized loss function, with n set to 1 for all boats.

Performance results based on both the validation and testing datasets are included. This dual evaluation approach provides a comprehensive understanding of the model's effectiveness and generalizability. The validation dataset helps in tuning the model parameters, while the testing dataset offers an unbiased assessment of the model's predictive capability and robustness in real-world scenarios. As illustrated in the figures, the model demonstrated robust performance on the testing datasets.

Additionally, when comparing the CNN models using MSE and the customized loss function, the results demonstrate that the customized loss function produces more accurate and conservative predictions, particularly for wave impacts exceeding the 1g threshold. This improved accuracy is crucial for ensuring the safety and performance of high-speed crafts, as it allows for better prediction of severe wave impacts. This demonstrates that the customized loss function's enhanced sensitivity to larger impacts makes it a valuable tool for refining predictive models in our applications.

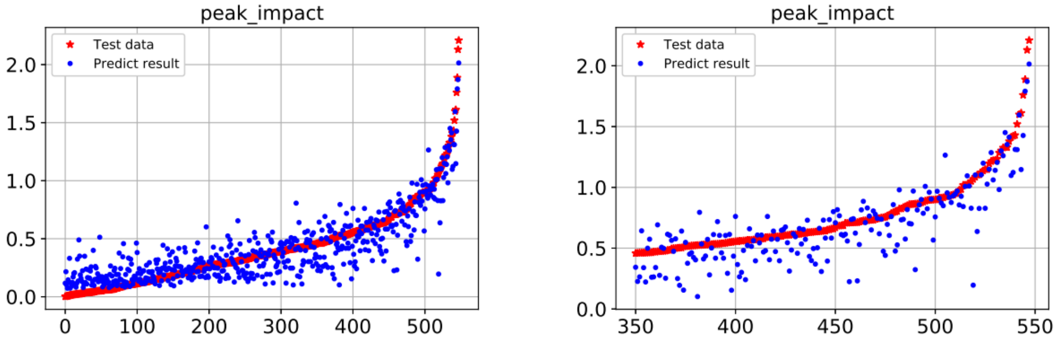
Furthermore, the figures illustrate the actual versus predicted values for all peaks, as well as for peaks exceeding 1g. These comparisons underscore the effectiveness of the customized loss function in enhancing model accuracy, particularly for significant wave impacts, highlighting the importance of its design.

Figure 7. *Plot of Predicted and True Values for Boat A with Mean Squared Error Loss Function in the Validation Set*



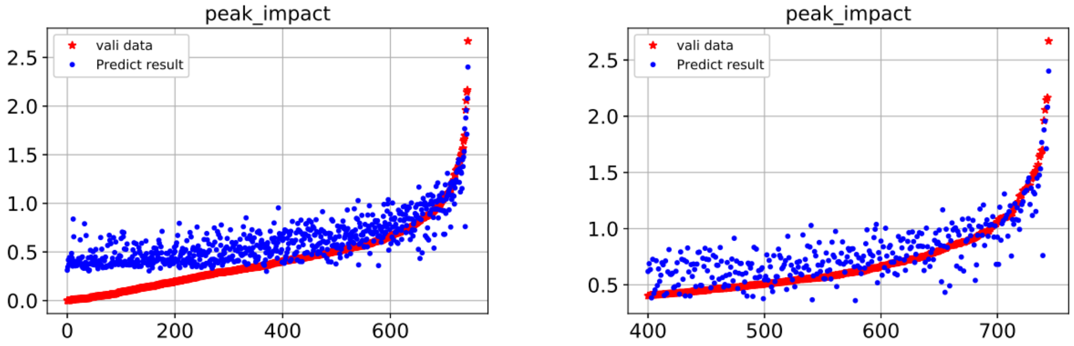
(a) All peak impacts (b) Peak impact larger than 1 g

Figure 8. *Plot of Predicted and True Values for Boat A with Mean Squared Error Loss Function in the Test Set*



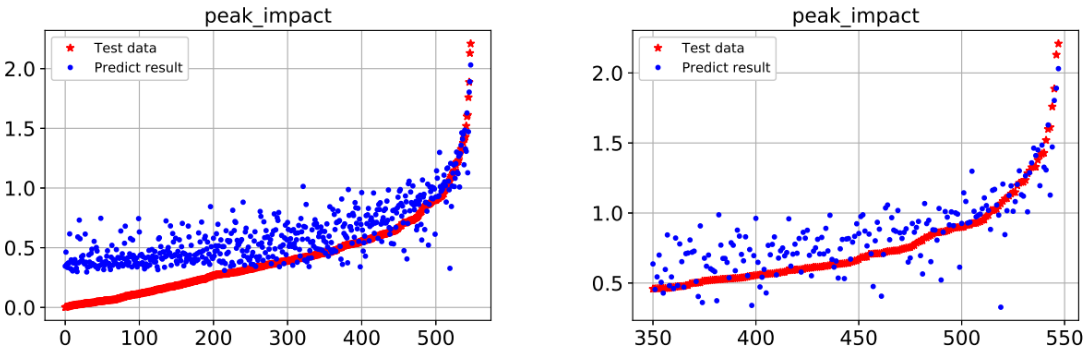
(a) All peak impacts (b) Peak impact larger than 1 g

Figure 9. *Plot of Predicted and True Values for Boat A with Customized Loss Function in the Validation Set*



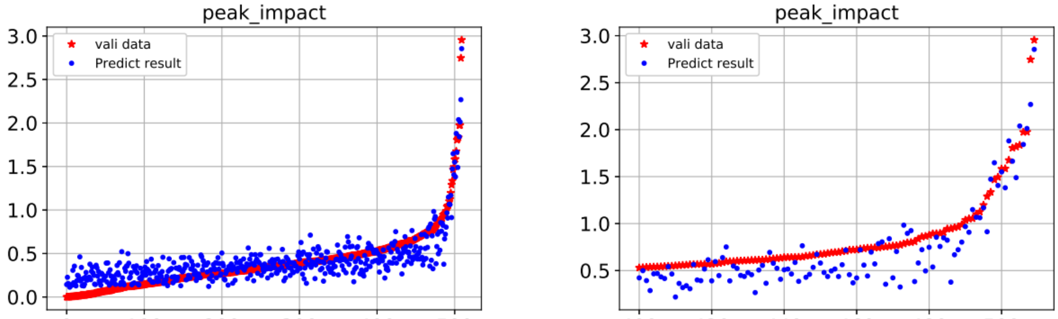
(a) All peak impacts (b) Peak impact larger than 1 g

Figure 10. *Plot of Predicted and True Values for Boat A with Customized Loss Function in the Test Set*



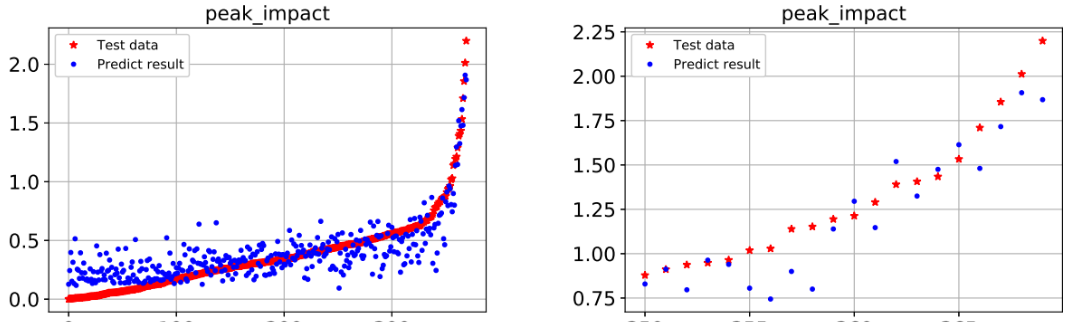
(a) All peak impacts (b) Peak impact larger than 1 g

Figure 11. *Plot of Predicted and True Values for Boat B with Mean Squared Error Loss Function in the Validation Set*



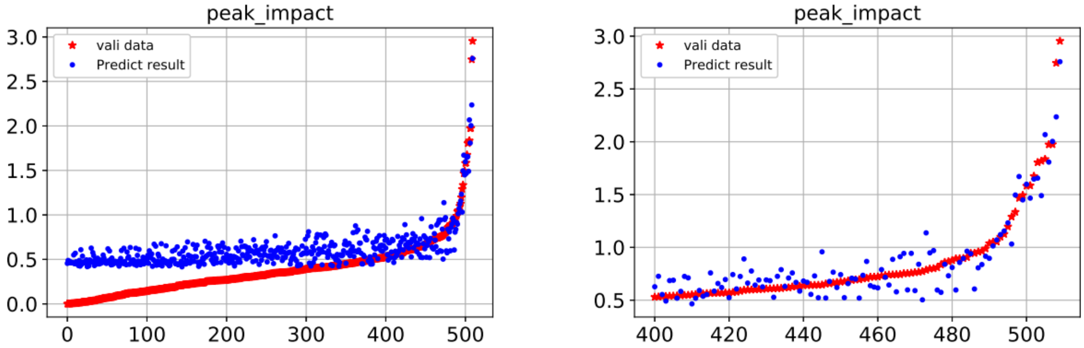
(a) All peak impacts (b) Peak impact larger than 1 g

Figure 12. *Plot of Predicted and True Values for Boat B with Mean Squared Error Loss Function in the Test Set*



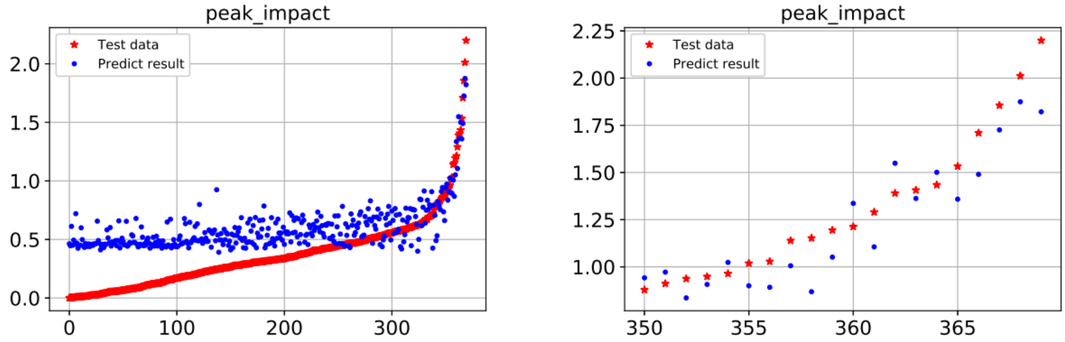
(a) All peak impacts (b) Peak impact larger than 1 g

Figure 13. *Plot of Predicted and True Values for Boat B with Customized Loss Function in the Validation Set*



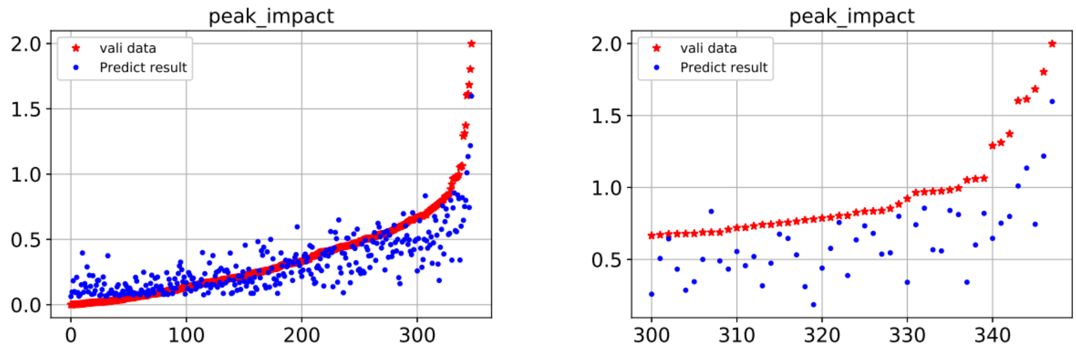
(a) All peak impacts (b) Peak impact larger than 1 g

Figure 14. *Plot of Predicted and True Values for Boat B with Customized Loss Function in the Test Set*



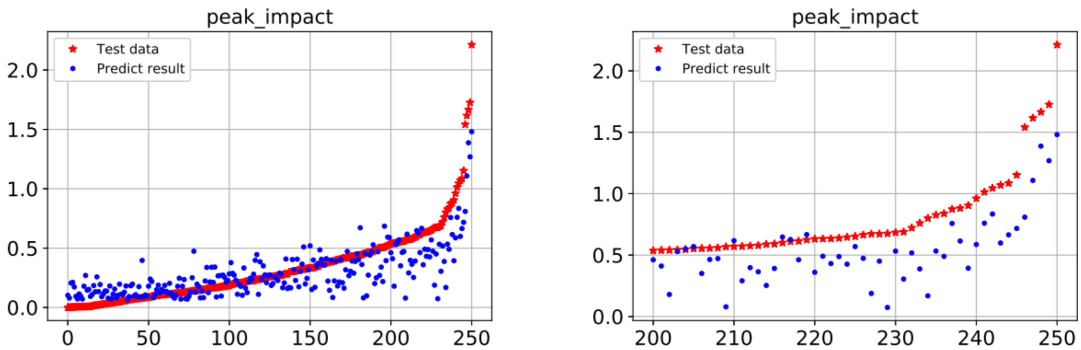
(a) All peak impacts (b) Peak impact larger than 1 g

Figure 15. *Plot of Predicted and True Values for Boat C with Mean Squared Error Loss Function in the Validation Set*



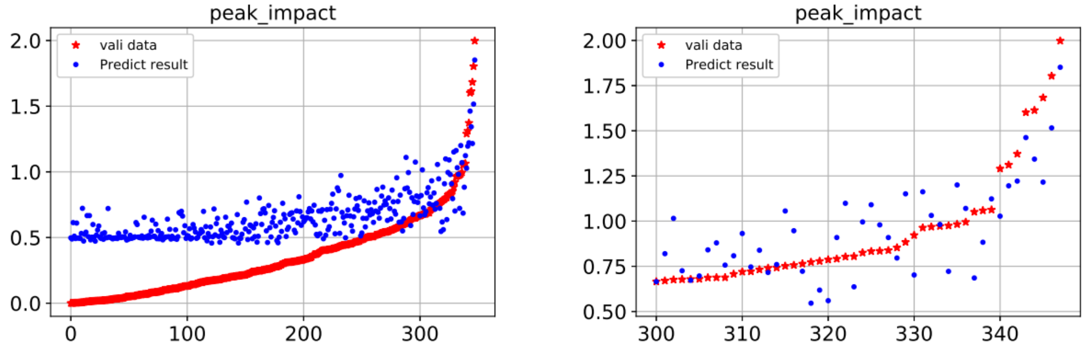
(a) All peak impacts (b) Peak impact larger than 1 g

Figure 16. *Plot of Predicted and True Values for Boat C with Mean Squared Error Loss Function in the Test Set*



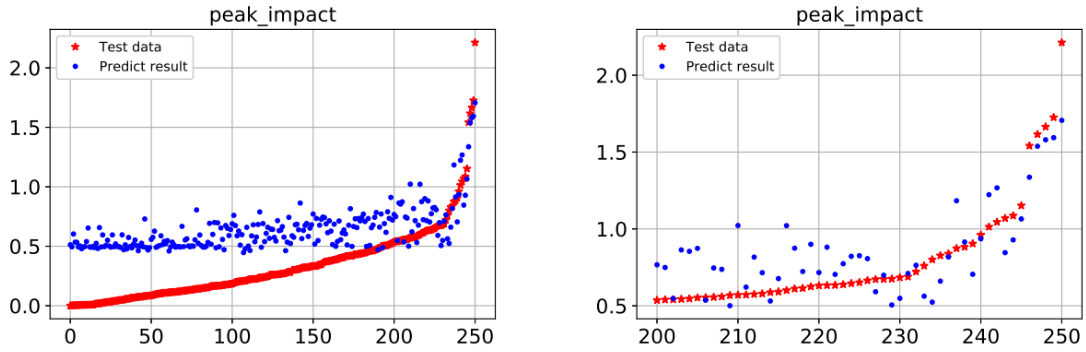
(a) All peak impacts (b) Peak impact larger than 1 g

Figure 17. *Plot of Predicted and True Values for Boat C with Customized Loss Function in the Validation Set*



(a) All peak impacts (b) Peak impact larger than 1 g

Figure 18. *Plot of Predicted and True Values for Boat C with Customized Loss Function in the Test Set*



(a) All peak impacts (b) Peak impact larger than 1 g

To summarize this section, the results demonstrate that while both models achieve high accuracy, the customized loss function model shows improved performance in predicting higher peak impacts, crucial for passenger safety and equipment protection. This model's enhanced precision and recall for severe impacts make it a valuable tool for high-speed craft operations, emphasizing the importance of tailored loss functions in real-world applications.

Conclusion

This study investigated the application of machine learning models, specifically Random Forest and Convolutional Neural Networks (CNNs), to predict peak acceleration impacts on high-speed watercraft. By focusing on both classification and regression modeling methods, we aimed to enhance the predictive capabilities and safety measures for passengers and equipment aboard these crafts. Vertical acceleration data from three boats (A, B, and C) are collected and normalized to units of g for consistency. Filtering methods, including a 5 Hz Butterworth filter, a 5 Hz FIR filter, and an Anisotropic Diffusion filter, are used to remove high-frequency noise. Recent advancements have demonstrated the effectiveness of data-driven approaches in thrust prediction for underwater propulsion systems, which can be adapted for high-speed craft impact modeling (Lee 2024).

The Butterworth filter is preferred for its minimal phase distortion. Feature extraction involves identifying key parameters such as the maximum values of the first and second derivatives, integration results, freefall duration, and other specific features, which were vital for predicting peak impacts using machine learning models. Correlation analysis provides more insight into the relationship between extracted features and peak impact.

Our results demonstrate that the Random Forest model effectively classifies peak impacts based on features extracted from the freefall region of the acceleration signals. The model achieves high accuracy across different boats. This indicates the model's robustness in handling the variability and complexity of wave impact data. The relative feature importance ranking results highlight the importance of feature extraction from filtered acceleration signals. Key features such as the minimum value of the first derivative, maximum value of the second-order derivative, freefall duration, and freefall area were crucial in enhancing the predictive accuracy of the models.

The CNN models are employed to predict peak impacts as continuous variables. The first CNN model, utilizing the mean squared error (MSE) loss function, treats all prediction errors equally. The second CNN model employs a customized loss function that assigned greater weight to higher peak impacts, reflecting their critical importance in real-world applications. The results indicate that while both models are highly accurate, the customized loss function model excels in predicting higher peak impacts, essential for passenger safety and equipment protection. Its improved accuracy and reliability for severe impacts make it a valuable tool for high-speed craft operations, underscoring the importance of tailored loss functions.

In conclusion, this study demonstrates the effectiveness of machine learning models in predicting peak acceleration impacts on high-speed watercraft. These models enhance passenger safety and mitigate equipment damage, offering a promising approach to enhancing maritime safety and performance. Future research should focus on extending the study's objectives to include additional aspects of the impact region, such as impact duration in addition to peak impact. Predictive analytics using machine learning has also been applied in ship motion trajectory forecasting and risk assessment, demonstrating its broader utility in maritime operations (Zhang 2023). It should also incorporate more sophisticated data processing techniques and expand the scope of predictive models to cover a broader range of operational scenarios.

References

Coats TW, Riley MR (2018) *A Comparison of Shock Isolated Seat and Rigid Seat Acceleration Responses to Wave Impacts in a High-Speed Craft*. Naval Surface Warfare Center Carderock Division, Naval Architecture and Engineering Department.

Ensign W, Hodgdon JA, Prusaczyk WK, Ahlers S, Shapiro D, Lipton M (2000) *A survey of self-reported injuries among special boat operators.* Naval Health Research Centre, 00–48.

Gohari M, Tahmasebi M (2014) Off-road vehicle seat suspension optimisation, Part II: Comparative study between meta-heuristic optimisation algorithms. *Journal of Low Frequency Noise, Vibration and Active Control* 33(4): 443–454.

Gohari M, Tahmasebi M (2015) Active off-road seat suspension system using intelligent active force control. *Journal of Low Frequency Noise, Vibration and Active Control* 34(4): 475–489.

Gohari M, Rahman RA, Tahmasebi M, Nejat P (2014) Off-road vehicle seat suspension optimisation, part I: derivation of an artificial neural network model to predict seated human spine acceleration in vertical vibration. *Journal of Low Frequency Noise, Vibration and Active Control* 33(4): 429–441.

Heidarian A, Wang X (2019) Review on seat suspension system technology development. *Applied Sciences* 9(14): 2834.

Ji X (2015) *Evaluation of suspension seats under multi-axis vibration excitations-a neural net model approach to seat selection*. Doctoral Dissertation. The University of Western Ontario (Canada).

Kavsaoğlu AR, Polat K, Hariharan M (2015) Non-invasive prediction of hemoglobin level using machine learning techniques with the PPG signal's characteristics features. *Applied Soft Computing* 37(Dec): 983–991.

Kumar V, Saran VH (2016) Seat-to-Head Transmissibility and Reading Discomfort of the Seated Subjects Exposed to Whole Body Vibration. *International Journal of Acoustics and Vibration* 21(1): 67–74.

Lee JA (2024) Data-Driven Approaches for Thrust Prediction in Underwater Flapping Fin Propulsion Systems. In *arXiv preprint arXiv:2406.01947*.

Lepine J, Rouillard V (2018) Evaluation of shock detection algorithm for road vehicle vibration analysis. *Vibration* 1(2): 220–238.

Lin J, Li H, Huang Y, Huang Z, Luo Z (2020) Adaptive artificial neural network surrogate model of nonlinear hydraulic adjustable damper for automotive semi-active suspension system. *IEEE Access* 8(Jun): 118673–118686.

Marlantes KE (2024) Predicting Ship Responses in Different Seaways using a Generalizable Force Correcting Machine Learning Method. In *arXiv preprint arXiv:2405.08033*.

Riley MR, Haupt KD, Jacobson DR (2010) *A Generalized Approach and Interim Criteria for Computing A sub 1/n Accelerations Using Full-Scale High-Speed Craft Trials Data*. Available at: https://www.semanticscholar.org/paper/A-Generalized-Approach-and-Interim-Criteria-for-A-1-Riley-Haupt/1ab5e48b282f9ad65109e10c1861bf78c2cabcec.

Riley MR, Coats TW, Murphy H (2014) *Acceleration response mode decomposition for quantifying wave impact load in high-speed planing craft.* Naval Surface Warfare Center Carderock Division Report NSWCCD-23-TR-2014/007.

Riley MR, Haupt KD, Ganey HN (2015) *Ride Severity Profile for Evaluating Craft Motions.* Naval Surface Warfare Center Carderock Division Report NSWCCD-80-TR-2015/002.

Riley MR, Haupt KD, Ganey HC, Coats TW (2018) *Laboratory test requirements for marine shock isolation seats.* Naval Surface Warfare Center Carderock Division Report NSWCCD-80-TR-2015/010 Rev A.

Salehpour M, Jamali A, Nariman ZN (2011) Optimal selection of active suspension parameters using artificial intelligence. *International Journal of Automotive Engineering* 1(4): 244–255.

Shehata AA (2023) *A*n attempt to predict planing hull motions using machine learning methods. In *IOP Conference Series: Materials Science and Engineering.*

Tahmasebi M, Rahman RA, Mailah M, Gohari M (2013) Roll movement control of a spray boom structure using active force control with artificial neural network strategy. *Journal of Low Frequency Noise, Vibration and Active Control* 32(3): 189–201.

Wice AM (2015) *Spatial dynamic modelling of high speed craft suspension seating*. Doctoral Dissertation. Carleton University.

Zhang MA (2023) A machine learning method for the prediction of ship motion trajectories in real operational conditions. *Ocean Engineering* 283(Sep): 114905.

Zhao Y, Wang X (2019) A review of low-frequency active vibration control of seat suspension systems. *Applied Sciences* 9(16): 3326.