Robot terrain classification based on improved Hilbert-Huang transform and long short-term memory network

Ziyu Guo

School of Physical Sciences and Technology, Central China Normal University, Wuhan, China

guoziyu@mails.ccnu.edu.cn

Abstract. Nowadays, with the rapid development of science and technology, the innovation and application of robot technology has become an important force to promote industrial development, which requires robots to have a high degree of autonomy and adaptability. Among them, terrain classification technology is one of the key technologies to achieve this goal. In order to improve the ground adaptability of robots in complex environments, this paper proposes a terrain classification algorithm based on improved Hilbert-Huang transform(HHT) combined with ensemble empirical mode decomposition(EEMD) and long short-term memory network(LSTM). Firstly, the signal data is processed by EEMD, and then the frequency domain features of the signal are extracted by Hilbert transform to expand the feature dimension. Finally, the features are learned and classified by the LSTM model, which effectively improves the classification accuracy. In this paper, we conducted sufficient experiments to compare the effects of EMD and EEMD and the effects of different neural network models, and verified the contribution of the Hilbert-Huang transform to improve the classification performance through ablation experiments, which proves the effectiveness and reliability of our proposed algorithm, and provided powerful technical support for the robot to adapt to the ground information in the complex environment.

Keywords: machine learning, terrain classification, Hilbert-Huang transform, long short-term memory networks.

1. Introduction

As industrial and information technologies rapidly evolve, robots are widely used in high-risk, high-complexity industrial field**s** [1]. In order to improve work efficiency and reduce work risks, robots are increasingly deployed in extreme environments such as field surveys, planetary exploration, etc [2]. These complex and changeable environments place higher demands for the robot 's ground adaptability level, and thus we urgently need an effective method to classify the ground where the robot is located, so that the robot can accomplish tasks in different environments.

Common ground-based classification methods in the field can be categorized into traditional machine learning algorithms and deep learning algorithms based on neural networks. Tsai et al. [3] combined aerial radar data and photographic images to extract geometric features and spectral features, followed

by a classification study of ground cover types using decision trees and support vector machines. Khairul Azmi Mahadhir et al. [4] used support vector machine (SVM) algorithm to identify different terrains by analyzing vibration signals, and then implement terrain classification for agricultural robots. Wenzhao Liao [5] used genetic algorithm to optimize the random forest model, and the optimized model has achieved a classification accuracy of 92.6%, which effectively improves the accuracy of ground classification. M. G. Harinarayanan Nampoothiri et al. [6] used machine learning techniques to develop a 100% accurate Ensemble-Subspace KNN based classification model, which can identify 11 terrain types for autonomous robots in real time. Although the machine learning algorithms can achieve good results, these algorithms usually need to manually design features, which is a tedious and difficult process, and it performs poorly when analyzing high-dimensional and complex ground data.

In recent years, with the improvement of arithmetic power and the development of artificial intelligence, scholars have begun to use neural-network-based deep learning approaches: Xinyun Zou et al. [7] proposed a reserve network (r-SNN) method based on spiking neuron networks for terrain classification using sensor data and camera images, and achieved an accuracy of more than 95 %. Amirreza Shaban et al. [8] employed a deep neural network called BEVNet for terrain classification, which can directly predict terrain classes of local maps from sparse LiDAR input data. Junghee Lee et al. [9] used convolutional neural network (CNN) as the main classification algorithm to convert one-dimensional spectral vectors into two-dimensional feature representations for ground cover classification studies. Ahmadreza Ahmadi et al. [10] used gated recurrent neural networks for semi-supervised robotic ground classification. Although deep learning algorithms can automatically learn the ground features, the deep learning algorithm applied by most scholars only processes in the time domain space and does not consider frequency domain features, which makes the network unable to capture the deep information behind the signal, making it difficult to make further breakthroughs in classification.

To tackle the above problems, this paper proposes a ground classification algorithm based on improved HHT algorithm based on EEMD and LSTM. The HHT algorithm is utilized to capture the frequency domain information in the signal that is difficult to be directly mined by the model. After that, the powerful feature extraction capability of the LSTM model is utilized for further learning and classification. This method can fully exploit the frequency domain features of the signal and take advantage of LSTM to achieve more accurate ground classification results.

2. Method

2.1. Hilbert-Huang transform

2.1.1. Empirical Modal Decomposition.
Hilbert-Huang Transform (HHT) can process nonlinear and non-stationary signals [11], which obtains the frequency domain information such as the instantaneous frequency of the signal and is very suitable for our signal data set. Specifically, the HHT contains two steps of empirical modal decomposition and Hilbert spectral analysis, where:

Empirical modal decomposition (EMD) is an adaptive method for analyzing time series data. It decomposes a complex time series into a series of intrinsic modal functions (IMFs), and then performs Hilbert transform on each IMF to extract features. However, the traditional EMD decomposition will produce problems such as modal aliasing and endpoint effects. Therefore, we use ensemble empirical modal decomposition (EEMD) to replace the traditional EMD to analyze the signals. EEMD adds noise several times and performs average processing on the basis of the EMD, so it can effectively avoid the modal aliasing problem of the EMD. Figure 1 shows the IMFs obtained by EMD decomposition of the input signal, and figure 2 shows the IMFs obtained by EEMD decomposition. It can be seen that the decomposition result of EEMD is more stable.
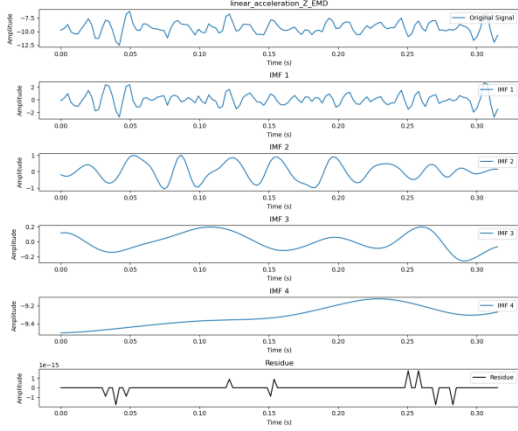
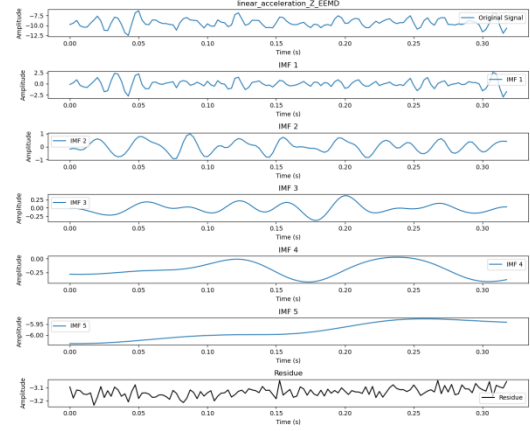**Figure 1.** EMD decomposition results.

Figure 2. EEMD decomposition results.

The procedure for EEMD decomposition is delineated as follows:

(1) Add N sets of white noise $n_i(t)$ with standard normal distribution to the original input signal $x(t)$ to obtain a new signal $x_i(t)$:

$$x_i(t) = x(t) + n_i(t) \tag{1}$$

(2) For the new signal $x_i(t)$, EMD decomposition is performed to obtain J IMF components $c_{i,j}(t)$ and residual function $r_i(t)$:

$$x_i(t) = \sum_{j=1}^{J} c_{i,j}(t) + r_i(t) \tag{2}$$

where $c_{i,j}(t)$ is the j-th intrinsic mode function obtained by EMD decomposition of the new signal $x_i(t)$, and $r_i(t)$ is the residual function of the new signal $x_i(t)$ obtained by the EMD decomposition .

(3) Perform an ensemble average operation on the aforesaid IMF components to achieve the final IMF of EEMD decomposition of the original input signal $x(t)$:

$$c_j(t) = \frac{1}{N} \sum_{i=1}^{N} c_{i,j}(t) \tag{3}$$

where $c_j(t)$ is the j-th intrinsic mode function acquired through EEMD decomposition of the original input signal $x(t)$.

Ultimately, the original input signal $x(t)$ is represented by the following formula:

$$x(t) = \sum_{j=1}^{J} c_j(t) + r(t) \tag{4}$$

2.1.2. Hilbert spectral analysis.

The original input signal $x(t)$ is decomposed by EEMD to extract J intrinsic mode functions. Then, Hilbert transform is used for time-frequency analysis of each IMF component $c_j(t)$, and the instantaneous frequency and instantaneous amplitude of the IMF are calculated to obtain the frequency domain characteristics of the signal.

The Hilbert transform of $c_j(t)$ is defined as:

$$H[c_j(t)] = \hat{c}_j(t) = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{c_j(t)}{t - \tau} d\tau \tag{5}$$

After the Hilbert transform, the analytic signal corresponding to $c_j(t)$ is defined as:

$$z_j(t) = c_j(t) + i c_j(t) = a_j(t) exp[i\theta_j(t)] \tag{6}$$

$$a_j(t) = [c_j^2(t) + \hat{c}_j^2(t)]^{1/2} \tag{7}$$

$$\theta_j(t) = arctan \frac{\hat{c}_j(t)}{c_j(t)} \tag{8}$$

$a_j(t)$ is the envelope of the signal $z_j(t)$, and $\theta_j(t)$ is the phase of the signal $z_j(t)$.

The instantaneous frequency of $c_j(t)$ is:

$$\omega_j(t) = \frac{d\theta_j(t)}{dt} \tag{9}$$

Thus, the frequency domain characteristics of the input signal are obtained

Empirical modal decomposition (EMD) is an adaptive time series analysis method. It decomposes a complex time series into a series of intrinsic modal functions (IMFs), and then performs Hilbert transform on each

2.2. Long short-term memory networks

LSTM is a recurrent neural network with special structure and function, which controls the flow of information through input gates, forgetting gates, and output gates, so that LSTM can selectively retain or forget information, overcoming the gradient disappearance and gradient explosion problems faced by traditional RNNs, and being able to handle long sequence data more effectively[12]. The input data set is 3810*128 long sequence data, which is suitable for LSTM model to process and complete the classification problem. Figure 3 illustrates the structure of the LSTM network.
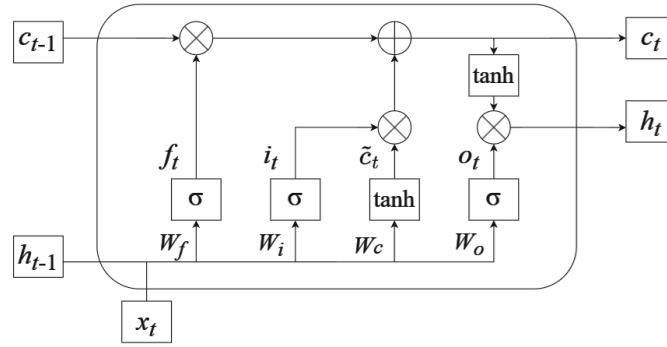

Figure 3. LSTM network structure.

(1) Input gate controls which information from the current input should be stored in the cell state:

$$i_t = \sigma(W_i \cdot [h_{t-l}, x_t] + b_i) \tag{10}$$
$$\tilde{c}_t = \tanh(W_c \cdot [h_{t-l}, x_t] + b_c) \tag{11}$$

(2) Forget gate decides what information should be discarded from the cell state:

$$f_t = \sigma(W_f \cdot [h_{t-l}, x_t] + b_f) \tag{12}$$

(3) Output gate controls what information about the cell state will be included in the next hidden state:

$$o_t = \sigma(W_o \cdot [h_{t-l}, x_t] + b_o) \tag{13}$$
$$h_t = o_t \odot tanh(c_t) \tag{14}$$

(4) Cell state:

$$c_t = f_t \odot c_{t-l} + i_t \odot \tilde{c}_t \tag{15}$$

In the above equation, W_i、W_c、W_f、W_o are the weight matrix, and b_i、b_c、b_f、b_o are the bias parameters, and σ is the sigmoid function, and $\odot$ is the dot product.

3. Experiment

3.1. Experimental hardware and software environment

In this experiment, NVIDIA GeForce RTX 3060 is used as the GPU, Python version is 3.11.7. The neural network and model were constructed using the PyTorch 1.13.1 framework, which is developed in the PyCharm environment.

3.2. Dataset description

The data used in this experiment were obtained from[13], the collector deployed 10 sensors in different ground environments for signal collection. Each set of signals collected by each sensor contains 128 sampling points, so each set of signals is a matrix with the shape of (128,10). Finally, 3810 sets of signal data were collected, and each set of data represents the characteristics of a specific ground type, which corresponds to 9 different ground types.

Figure 4 shows an example set of data collected by 10 sensors, and figure 5 shows the distribution of the nine different ground categories across all signals.
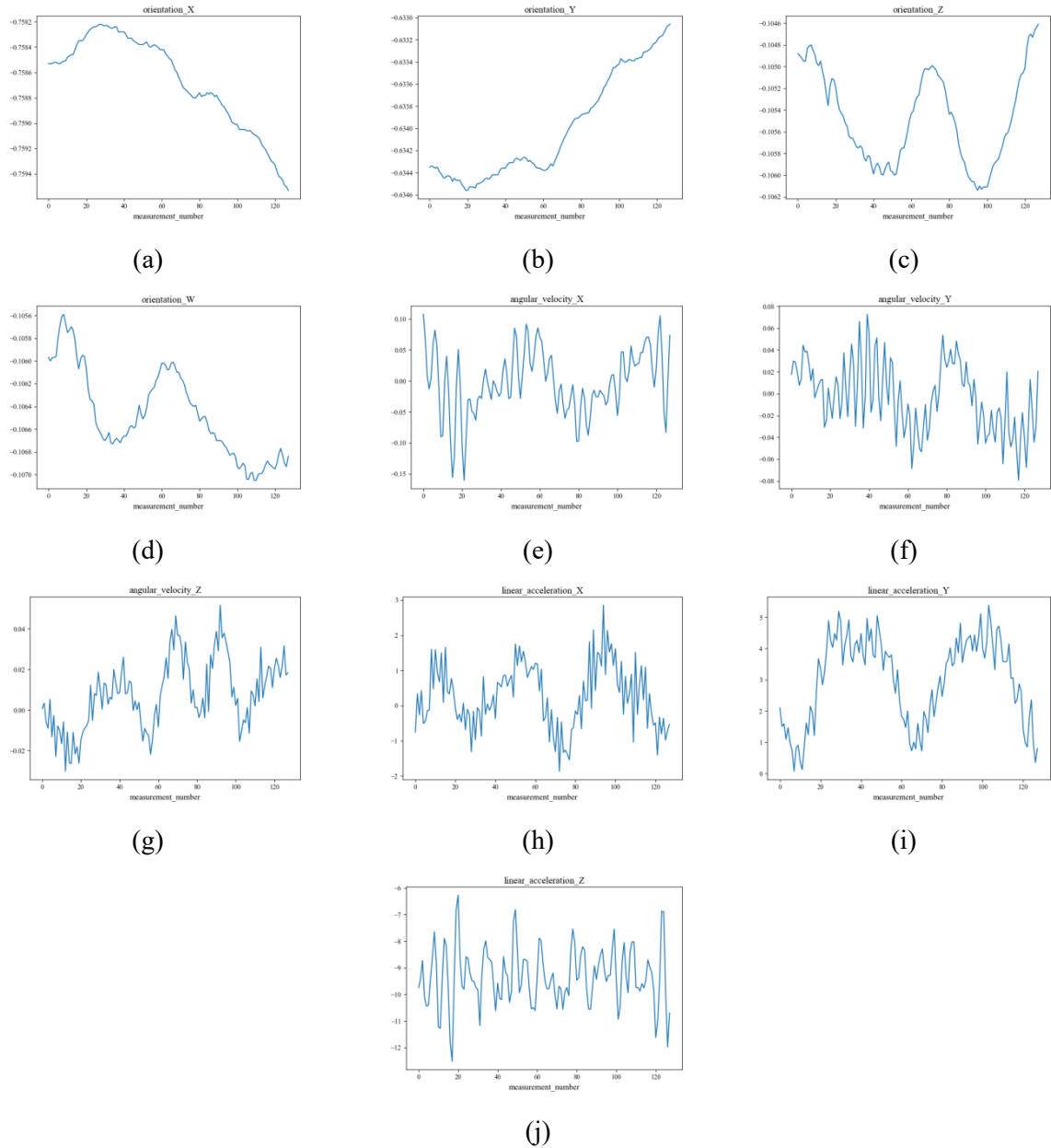

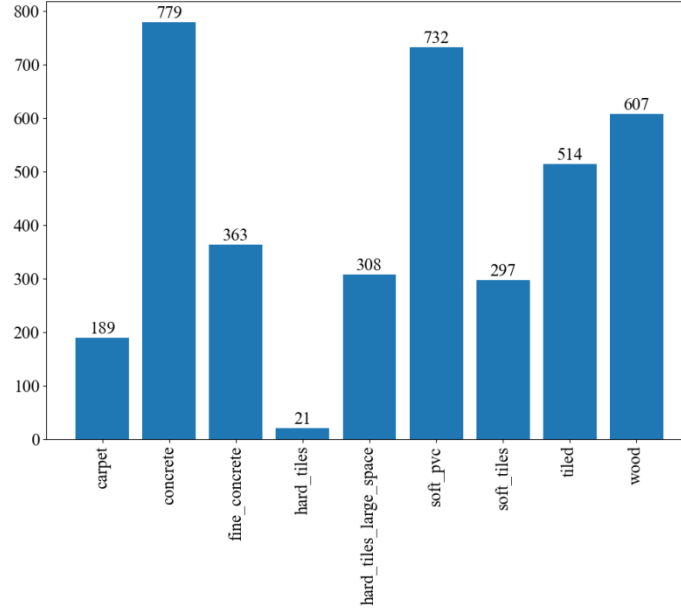**Figure 4.** Waveforms of a set of signals acquired by 10 sensors.

Figure 5. Distribution of ground categories.

3.3. Experimental Measurement Indicators

When using the neural network model for classification, we construct a 9*9 confusion matrix to assess the performance of the model. Each element in the confusion matrix C_{ij} represents the count of samples from class i that are categorized by the model as class j. The diagonal elements represent the count of samples accurately classified by the model, whereas the non-diagonal elements represent the number of samples misclassified. Based on the matrix, we can calculate the assessment indicators such as accuracy, precision, recall and F1 score in classification problems:

Accuracy measures the model's overall prediction efficacy by comparing the number of correct predictions to the total samples, as outlined in the following calculation:

$$Accuracy = \frac{\sum_{i=1}^{n} C_{ii}}{\sum_{i=1}^{n} \sum_{j=1}^{n} C_{ij}} \tag{16}$$

Precision measures the ratio of all samples predicted to be positive by the model that are actually positive, reflecting the reliability of the model's predictive outcomes, as outlined below:

$$Precision_i = \frac{C_{ii}}{\sum_{j=1}^{n} C_{ij}} \tag{17}$$

Recall measures the fraction of actual positive samples that are accurately classified as such by the model, reflecting the model's capacity to detect positive samples, and is computed as follows:

$$Recall_i = \frac{C_{ii}}{\sum_{j=1}^{n} C_{ji}} \tag{18}$$

The F1 Score (F1 Score) is the reconciled average of the precision and recall, which is employed to evaluate the balance of accuracy and completeness of the model. The computation formula is delineated below:

$$F1_i = 2 \times \frac{Precision_i \times Recall_i}{Precision_i + Recall_i} \tag{19}$$

3.4. Comparative experiments

3.4.1. Comparison of classification models.

In order to select the final neural network model for classification, we compared the classification effects of LSTM and 1DCNN models. To ensure a fair comparison, we used the original data as the base dataset,

trained and tested with the two models separately, and used the five-fold cross-validation method for metrics measurement. The cross-validation outcomes are delineated in table 1.

Table 1. Metrics for evaluating the LSTM model and 1DCNN model with five-fold cross-validation.

Model	accuracy	precision	recall	f1-score
LSTM	0.7932	0.7529	0.7348	0.7398
1DCNN	0.3840^c	0.3642	0.3131	0.3134

Table 1 illustrates that the precision of the LSTM model outperforms the 1DCNN model in classification tasks, so we selected LSTM as our classification model in the subsequent experiments.

3.4.2. Comparison of Empirical Modal Decomposition Methods
In the previous section, we mentioned that the decomposition of HHT transform has two different ways, EMD and EEMD, and the latter is an improvement of the former. In order to verify the different effects of the two approaches, we keep the rest of the conditions unchanged, and compare classification effect after decomposing dataset with EMD and EEMD respectively. The five-fold cross-validation is still adopted, and table 2 shows the classification results.

Table 2. Classification results corresponding to EMD and EEMD decomposition.

Model	accuracy	precision	recall	f1-score
LSTM	0.7932	0.7529	0.7348	0.7398
1DCNN	0.3840^c	0.3642	0.3131	0.3134

Table 2 shows that the average accuracy of EEMD is higher than that of EMD, which proves the effectiveness of EEMD as an improvement.

3.5. Ablation experiments
To evaluate the improvement effect of the HHT transform on the classification performance, we conducted classification experiments on the dataset containing only the original signals and the expanded dataset with the transformed signals respectively, while keeping all other conditions constant. The cross-validated results are shown in figure 6. The figure illustrates that our proposed HHT method can improve the classification accuracy, proving the effectiveness of our method.
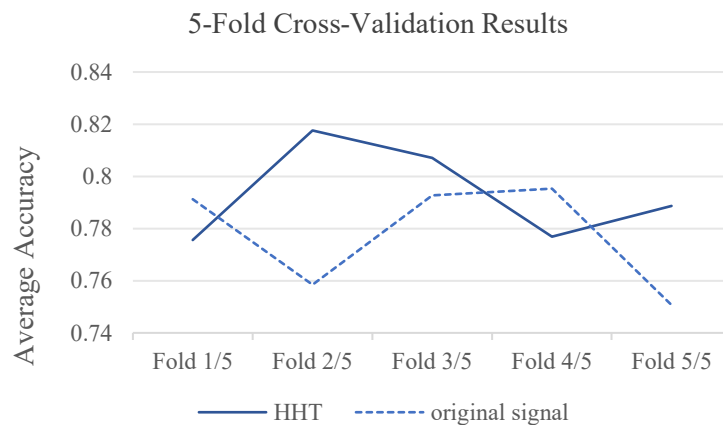

Figure 6. Average accuracy of five-fold cross-validation between original and expanded dataset.

4. Conclusion
Employing the dataset on the Kaggle website as the experimental dataset, this paper proposes a robot ground classification algorithm that combines EEMD and LSTM, which effectively realizes efficient

classification of complex ground environments. The original dataset was collected by 10 sensors, containing a total of 3810 sets of signals and 9 ground types, and each set of signals is a matrix of size (128, 10). We conducted two sets of comparison experiments between EMD/EEMD and LSTM/1DCNN: the classification accuracy obtained using the original dataset combined with the LSTM model is 77.77%, while the classification accuracy using the improved HHT combined with EEMD algorithm and the LSTM model is improved to 79.32%, which proves the effectiveness of the EEMD improvement; the classification accuracy of the improved HHT combined with EEMD algorithm and 1DCNN model is only 38.40%, proving that LSTM is more suitable than 1DCNN for the task in this dataset. Additionally we performed ablation experiments to demonstrate the improvement and enhancement of the HHT transform for the results. Future work can further optimize the robot ground classification problem from the perspective of combining multiple deep learning methods and optimizing data processing for a wider range of application scenarios and higher robustness.

References

[1]   Wong C, Yang E, Yan XT, Gu D. Autonomous robots for harsh environments: a holistic overview of current solutions and ongoing challenges. Syst Sci Control Eng. 2018 Jan 1;6(1):213–9.

[2]   Chai H, Li Y, Song R, Zhang G, Zhang Q, Liu S, et al. A survey of the development of quadruped robots: Joint configuration, dynamic locomotion control method and mobile manipulation approach. Biomim Intell Robot. 2022 Mar;2(1):100029.

[3]   Tsai MD, Tseng KW, Lai CC, Wei CT, Cheng KF. Exploring Airborne LiDAR and Aerial Photographs Using Machine Learning for Land Cover Classification. Remote Sens. 2023 Apr 26;15(9):2280.

[4]   Mahadhir KA, Tan SC, Low CY, Dumitrescu R, Amin ATM, Jaffar A. Terrain Classification for Track-driven Agricultural Robots. Procedia Technol. 2014;15:775–82.

[5]   Liao W. Ground classification based on optimal random forest model. In: 2023 IEEE International Conference on Control, Electronics and Computer Technology (ICCECT) [Internet]. Jilin, China: IEEE; 2023 [cited 2024 May 7]. p. 709–14. Available from: https://ieeexplore.ieee.org/document/10141122/

[6]   Nampoothiri MGH, Anand PSG, Antony R. Real time terrain identification of autonomous robots using machine learning. Int J Intell Robot Appl. 2020 Sep;4(3):265–77.

[7]   Zou X, Hwu T, Krichmar J, Neftci E. Terrain Classification with a Reservoir-Based Network of Spiking Neurons. In: 2020 IEEE International Symposium on Circuits and Systems (ISCAS) [Internet]. 2020 [cited 2024 Apr 30]. p. 1–5. Available from: https://ieeexplore.ieee.org/abstract/document/9180740

[8]   Shaban A, Meng X, Lee J, Boots B, Fox D. Semantic Terrain Classification for Off-Road Autonomous Driving.

[9]   Lee J, Han D, Shin M, Im J, Lee J, Quackenbush LJ. Different Spectral Domain Transformation for Land Cover Classification Using Convolutional Neural Networks with Multi-Temporal Satellite Imagery. Remote Sens. 2020 Mar 30;12(7):1097.

[10]   Ahmadi A, Nygaard T, Kottege N, Howard D, Hudson N. Semi-Supervised Gated Recurrent Neural Networks for Robotic Terrain Classification. IEEE Robot Autom Lett. 2021 Apr;6(2):1848–55.

[11]   Ma C, Xie S, Bi CW, Zhao YP. Nonlinear dynamic analysis of aquaculture platforms in irregular waves based on Hilbert–Huang transform. J Fluids Struct. 2023 Feb;117:103831.

[12]   Guan Y, Liu J. Research on the Method of Remaining Useful Life Prediction of Lithium-ion Battery Based on LSTM. In: 2021 3rd International Conference on Advances in Computer Technology, Information Science and Communication (CTISC) [Internet]. Shanghai, China: IEEE; 2021 [cited 2024 Jun 29]. p. 226–31. Available from: https://ieeexplore.ieee.org/document/9527645/

[13]   CareerCon 2019 - Help Navigate Robots [Internet]. [cited 2024 Jul 7]. Available from: https://kaggle.com/competitions/career-con-2019