Data-driven Prediction of Underwater Navigation Fitness Zone Classification

Nan Yang [1, *], Huiyong Zhang [2], Haotian Niu [3]

[1] Faculty of Intelligence and Engineering, Shenyang City University, Shenyang, China, 110112

[2] Faculty of Life and Health Management, Shenyang City University, Shenyang, China, 110112

[3] Business School, Shenyang City University, Shenyang, China, 110112

* Corresponding Author Email: 18641828766@163.com

Abstract. This article presents a pioneering approach to enhancing underwater navigation accuracy and stealth through the utilization of gravity anomaly data. Developed by researchers from leading universities in China, the study aims to revolutionize autonomous marine navigation systems in unknown sea areas. By integrating advanced data processing, feature extraction, model construction, and optimization techniques, a cutting-edge navigation system has been formulated, showcasing remarkable improvements in navigation precision.The core innovation lies in leveraging bicubic interpolation to refine data resolution, enabling a meticulous evaluation of regional suitability and segmentation of sea zones. Subsequently, a gradient-based suitability prediction model is constructed, which is validated through migration prediction, underscoring its effectiveness and reliability. This research not only addresses the limitations of traditional terrain matching navigation methods, plagued by data accuracy and resolution constraints, but also paves the way for advancements in oceanic exploration and technological competitiveness.The introduction of gravity anomaly data, coupled with machine learning and visualization tools, marks a significant step forward in the development of adaptive navigation zones. This innovative system holds immense potential to enhance the autonomy, accuracy, and concealment of underwater vehicles, thereby facilitating safer and more efficient oceanic missions in the future.

Keywords: Gravity anomaly data, Autonomous marine navigation system, Adaptive region.

1. Introduction

As humanity delves deeper into the exploitation and exploration of marine resources, the autonomous navigation capabilities of underwater vehicles (UVs) have become paramount. The intricate and dynamic oceanic environment, characterized by rugged terrain, currents, tides, water temperature, and salinity, poses significant challenges to precise UV navigation. While traditional terrain-matching navigation methods can offer some navigational functionality, they heavily rely on the accuracy and resolution of terrain data, leading to frequent mismatches in unknown or complex maritime regions, thereby compromising navigation accuracy and even causing mission failures.

In recent years, advancements in ocean observation technologies and big data processing capabilities have sparked research interest in multi-sensor data fusion for navigation. Among these, gravity anomaly data stands out as a crucial geophysical field data source. It reflects variations in crustal density, revealing the undulations of submarine terrain and finding extensive applications in marine geological exploration and underwater topographic mapping. However, the potential of gravity anomaly data in UV autonomous navigation remains largely untapped.

A review of the literature reveals that scholars worldwide have extensively studied autonomous navigation techniques for UVs. Traditional methods, such as inertial navigation, acoustic navigation, and terrain-aided navigation, each have their merits and drawbacks but struggle to maintain high precision and robustness in the complex and varied oceanic environment. With the rapid development of machine learning and artificial intelligence technologies, data-driven navigation methods have gradually gained traction. These methods extract multi-source data features from the marine environment, construct predictive models, and enable autonomous path planning and optimization for navigation [1].

2. Research methodology

2.1. Data preprocessing and fitness calibration

The data in this study come from the actual investigation and simulation. Data preprocessing is one of the key steps in this study. We firstly check and process the gravity anomaly data for missing values and duplicates, and then use the double-cubic interpolation method to improve the resolution of the data. By calculating the data gradient and combining with visualization techniques, we can more accurately assess the suitability of each region and reasonably divide the sea area [2].

2.2. Fitness classification prediction system construction

Based on the preprocessed data, we identified the key feature attributes, and constructed and trained a classification prediction system using machine learning models (e. g., decision trees, neural networks, etc.). The system is able to realize accurate prediction of area suitability output results and provide accurate navigation support for underwater vehicles.

2.3. Model performance and applicability assessment

In order to validate the performance and applicability of the model, we conducted a comprehensive evaluation of the gravity anomaly dataset. By comparing the relationship between different gradients and fitness, we evaluate the predictive ability of the model and conclude the performance and applicability of the model on the gravity anomaly dataset.

3. Modeling and solving

3.1. Gravity anomaly datum data refinement and regional suitability calibration

The gravity anomaly benchmark data are preprocessed and interpolated to improve the resolution, key feature attributes are extracted to reasonably divide the regions, and the suitability calibration of each region is completed based on the feature attributes, so as to realize the refined analysis of the gravity anomaly data and the assessment of the regional suitability.

3.1.1. Visualization

There is no loss of data, then you can carry out a preliminary visualization, because it is a preliminary visualization, you should consider how to carry out the most reasonable regional division and interpolation.

Use python [3] to write code to achieve visualization as shown in Figure 1 below:

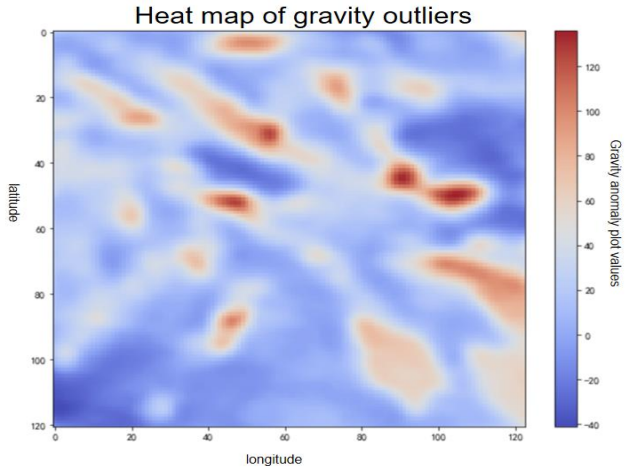

Figure 1. Heat map of gravity anomalies [4].

A heat map of the gravitational anomalies is already available, and interpolation can be used to improve the resolution of the data. In other words, interpolation can be used to estimate the gravity

anomalies that are not directly observed, thus obtaining finer data. This method can help to better understand the distribution of gravity anomalies in different regions.

3.1.2. Interpolation [5]

The principle of bicubic interpolation is to determine the pixel points to be interpolated in the target image, the(X, Y). According to the mapping relationship between the source image and the target image, determine the corresponding position of the pixel point to be interpolated in the source image$(x. y)$. In the source image to$(x. y)$ as the center, the nearest 16 pixels are selected as the target pixels for calculation$(X. Y)$ value of the parameters. The weights of these 16 pixel points are calculated using the double cubic interpolation basis function to obtain the target pixel point, the$(X. Y)$ pixel value. The formula is j$P = f(u, v) = \sum$i $= 03\sum$j $= 03$wijxiyi. This is the basic formula for double cubic interpolation, where p is the value of the new pixel, the$(u. v)$ are the coordinates of the inserted point, the $(x. y)$ is the coordinates of the surrounding 16 pixels. This formula is able to estimate the new pixel values more accurately by double-cubic interpolation, producing smoother image edges and more accurate colors.

The advantages of bicubic interpolation in image processing mainly include.

(1) Excellent interpolation effect: the double-three interpolation method can produce smoother image edges than bilinear interpolation, while retaining the original image details and

Clarity. This is because it takes into account the influence of more surrounding points and interpolates three times instead of once.

(2) Higher computational efficiency: Although double-cubic interpolation requires more computational resources and time, in some cases, more efficient algorithms and computational platforms can be utilized to improve computational speed. At the same time, compared with some other complex interpolation techniques, the computation of double-cubic interpolation is not particularly high.

(3) Better edge fidelity: Double-cubic interpolation can better preserve the details and shapes of the original image when dealing with image edges, reducing the blurring or distortion of the edges.

(4) Stronger anti-aliasing ability: double triple interpolation for anti-aliasing is also excellent, can reduce the jaggedness of the image, so that the image is more smooth and natural.

(5) Better visualization: By double cubic interpolation, a continuous interpolation function can be obtained, which has continuous first-order partial derivatives and continuous cross derivatives everywhere. This means that the interpolation result is visually smoother and more natural.

Figure 2 is the python code and visualization. After interpolation processing, higher resolution gravity anomaly data can be obtained. Next, different regions can be divided according to the interpolation results, and the suitability of each region can be marked. In other words, the geological features and properties of different areas can be determined based on these data to better interpret and evaluate the data.

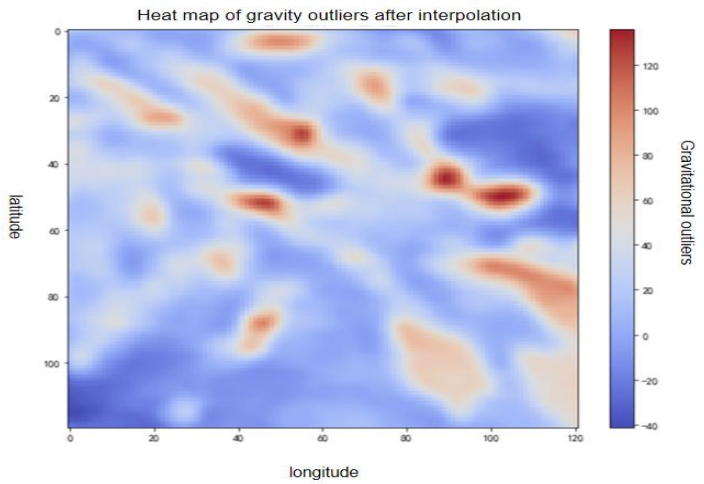

Figure 2. Heat map of gravity anomalies after interpolation process.

3.1.3. Adaptability calibration

After the preliminary data processing and interpolation visualization are completed, the appropriate suitability criteria can be defined to facilitate the observation of the distribution of gravity outliers. It can be broken down into the following steps:

(1) Calculate the gradient of outliers

(2)Assign value labels based on gradient definitions

(3)Apply adaptability to different points

Calculate the gradient [6]:

The gradient is a vector that represents the direction of the maximum growth rate of a function at each point. For the function$f(x,y)$, its gradient can be expressed as: gradient $f = (\partial f \div \partial x, \partial f \div \partial y)$among them $\partial f \div \partial x$is the partial derivative of the function f with respect to x. The partial derivatives of, indicate that f exist x Rate of change in the direction. $\partial f \div \partial y$ is the partial derivative of the function f with respect to y The partial derivatives of, indicate that f exist y direction; the magnitude of the gradient, which represents the rate of change of the function in the direction of the fastest change at the point, can be calculated by the following formula: The size of the gradient $|grand(f)|\sqrt{((\partial f \div \partial x)^2 + (\partial f \div \partial y)^2)}$. The magnitude of the gradient is used to find the direction in which the function grows fastest, and this can be used to determine its fitness.

The following is the calculation and visualization of Python and gets the following Figure 3:

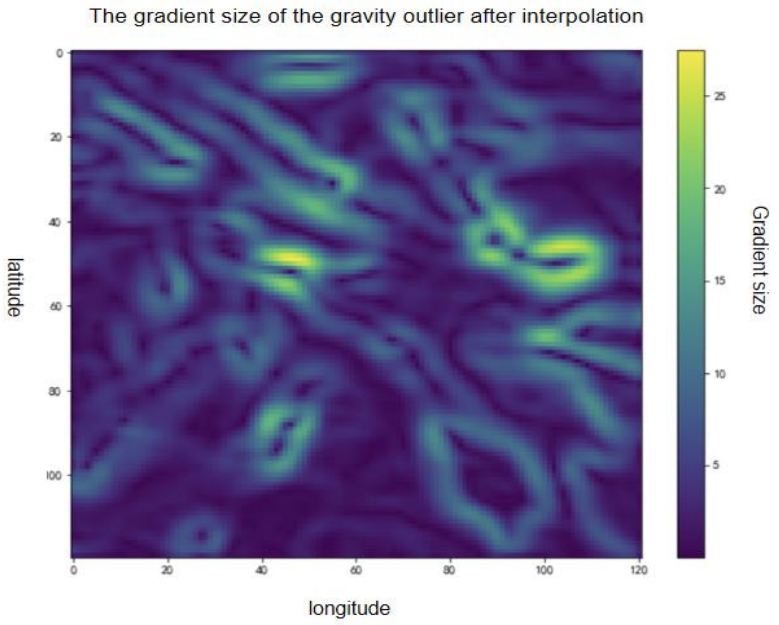

Figure 3. Gradient magnitude of gravity anomalies after interpolation.

The gradient plot shows the rate-of-change distribution of interpolated gravity anomalies. Areas with larger gradients appear brighter in the figure, and gravity anomalies change more significantly in those areas of the surface.

Adaptability Labeling

It can be tagged as follows.

(1) High fitness: regions with gradient values higher than 70%.

(2) Medium fitness: regions with gradient values above 30% and below 70%.

(3) Low fitness: regions with gradient values below 30%.

Adaptability visualization.

After completing the fitness labeling, a scatterplot with different colors was used to represent the different fitness points, as well as the fitness labels (Figure 4).

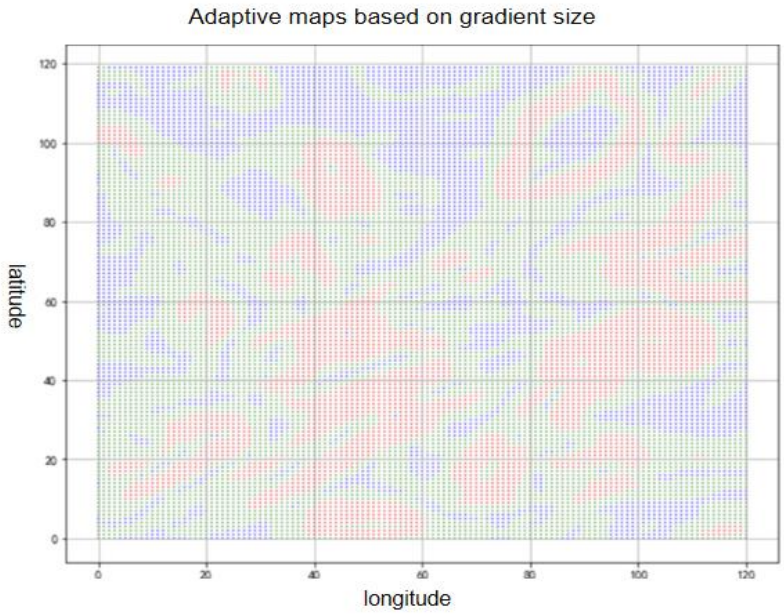

Figure 4. Gradient-size-based fitness maps.

The fitness map is a direct display of the distribution of regions with different fitness classes based on the gradient of the gravity anomaly.

The red color represents the high adaptation area. The green color represents the medium adaptation area. The blue color represents the low adaptation area.

3.2. Performance comparison of using machine learning [7] models to predict regional fitness label Y

Choose an appropriate machine learning model, such as decision tree, random forest, support vector machine, neural network, etc., to predict the region al fitness label Y. Divide the data set into a training set and a test set, use the training set to train the model, and then use the test set to evaluate the performance of the model.

3.2.1. SVM classification model introduced

SVM(Support Vector Machine) is a commonly used machine learning model. It is a second-class classification model in which the basic model is defined as the linear classifier with the largest intervals over the feature space [8]. The goal is to find a hyperplane in the feature space, so that the two types of data are as far away from the hyperplane as possible, so that the new data classification is more accurate.

Features of SVM include:

Nonlinear mapping: The theoretical basis of the SVM method is nonlinear mapping, which uses the inner product kernel function to map samples to a high-dimensional space, thereby transforming the nonlinear problem into a linear problem. Optimal hyperplane. The goal is to find the optimal hyperplane in the feature space, so that the two types of data are furthest away from the hyperplane, maximizing the classification interval.

Support vectors [9]: Support vectors are those points that are closest to the separating hyperplane, and SVM is trained to find these support vectors.

Kernel Functions: SVM can solve nonlinear problems by using kernel functions to map samples to high-dimensional spaces. Commonly used kernel functions include linear kernel functions, polynomial kernel functions, Gaussian radial basis functions, etc.

The advantages of SVM include:

Strong generalization ability [10]: SVM can avoid overfitting and has good generalization ability.

Processing high-dimensional data: SVM can process high-dimensional data and is suitable for various types of data.

High computational efficiency: The training process is relatively simple, and the SVM computational efficiency is relatively high.

The disadvantages of SVM include:

Sensitive to parameters and kernel functions: The selection of SVM's parameters and kernel functions has a great impact on the classification results and needs to be carefully adjusted.

Problems may be encountered when processing large-scale data: The training process of SVM involves matrix operations, and the problem of excessive computation may be encountered when processing large-scale data.

In conclusion, it is an effective classification model for all types of data and problems. When usingSVM, you need to pay attention to selecting the appropriate parameters and kernel functions to get the best classification results.

3.2.2. Prediction of results

The label Y and feature X have been obtained in the previous fitness label.

Next, we need to use gradient and adaptability prediction. The result of this prediction model is the regional adaptation classification prediction model (system F) required by the question. The model parameters are shown in Table 1.

Analysis Process.

The above table shows the configuration of each parameter of the model as well as the model training time

The output result 2: confusion matrix heat map is shown in Figure. 5.

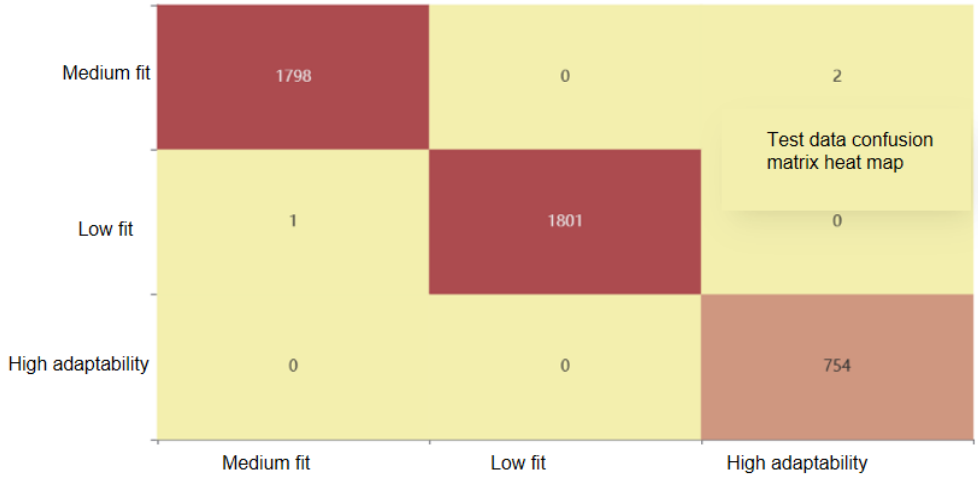

Figure 5. Heat map of the confusion matrix.

Output 3: The results of the model evaluation are shown in Table 1.

Table 1. Results of model evaluation.

	Accuracy	Recall	Accuracy	F1
The Training set	0.999	0.999	0.999	0.999
test set	0.999	0.999	0.999	0.999

The table above shows the prediction evaluation metrics for the cross-validation set, the training set and the test set, which measure the prediction results of the support vector machine through quantitative metrics. This table shows how cross-validation can be used to tune the hyperparameters of a model to obtain a more stable and accurate model. A cross-validation set is a method used to evaluate the hyperparameters of a model. By performing cross-validation multiple times on a dataset, the stability and accuracy of the model can be assessed and the hyperparameters can be adjusted to optimize the performance of the model.

Accuracy: the proportion of correctly predicted samples, the larger the better.

Recall: the proportion of positive samples predicted, the larger the better.

Precision rate: the proportion of results predicted to be positive that are actually positive, the larger the better.

F1: Harmonic average of precision and recall.

Output 4: The results of the predictive evaluation of the test data are shown in Table 2.

Table 2. Results of the test data prediction assessment.

Prediction Outcome Y	Adaptability	Prediction Outcome Probability_Medium Adaptability	Probability of prediction result _ low adaptability	Predict the probability of the result and have high adaptability	gradient
Low fit	Low fit	3.6759144956938125e-7	0.9996962971074116	0.00030333530113891996	1.17963134371935
Low fit	Low fit	0.035552145362956346	0.9633648095613778	0.001083045075666001	1.7636406670274
Medium fit	Medium fit	0.999999848907506	9.99878794587665e-8	5.110461469582168e-8	3.00029991144412
Medium fit	Medium fit	0.9999998464033211	9.986306567968423e-8	5.3733613318443296e-8	2.47629602269458
Medium fit	Medium fit	0.9999998421638917	9.929811480654109e-8	5.853799344926344e-8	3.02839474036856
Medium fit	Medium fit	0.9999998371644564	9.789899371303646e-8	5.351194368788237e-8	3.41606703719574
Medium fit	Medium fit	0.999999833143304	9.594705095765108e-8	6.493655000408733e-8	3.69495246050607
Medium fit	Medium fit	0.9999998318638827	9.510378165805198e-8	7.09096450775373e-8	3.87775942912433
Medium fit	Medium fit	0.9999998346830136	9.680498074183392e-8	7.303233568913084e-8	3.93398929215219
Medium fit	Medium fit	0.9999998416088455	9.918626747253647e-8	6.851200577297755e-8	3.80960086973893
Medium fit	Medium fit	0.9999998473315337	9.992577836118514e-8	5.9204887096102605e-8	3.4524352280676

The above table is only a pre view of the results, only the data is displayed.

The table above shows the support vectors several, the (SVM) Classification results for the test data, the classification result value is the classification group that has the maximum prediction probability. It can be seen that the prediction accuracy is 99.9%. The

3.3. Predicting regional fitness labels for gravity anomaly data using trained machine learning models

In previous tasks, one or more machine learning models have been successfully selected, trained, and evaluated to predict the region fitness label Y. Now, with gravity anomaly data, these data contain features related to region al suitability. The goal is to use the best previously trained model to predict the region al fitness label Y for these gravity anomaly data.

3.3.1. Interpolation visualization

Still the interpolation and visualization steps are shown in Figure 6.

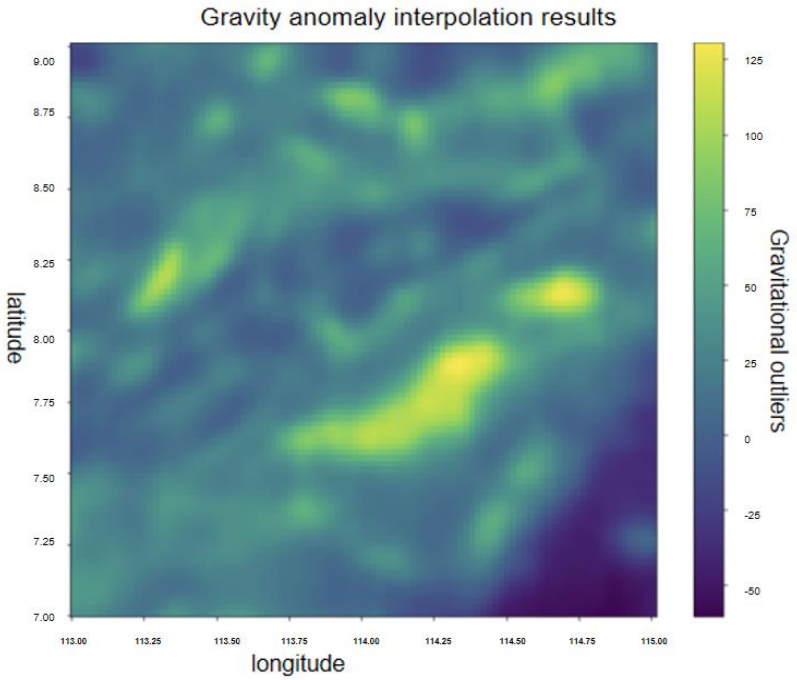

Figure 6. Gravity outlier results.

3.3.2. Gradient size calculation

Using the same calculation, the gradient size can be calculated and visualized as shown in Figure7.

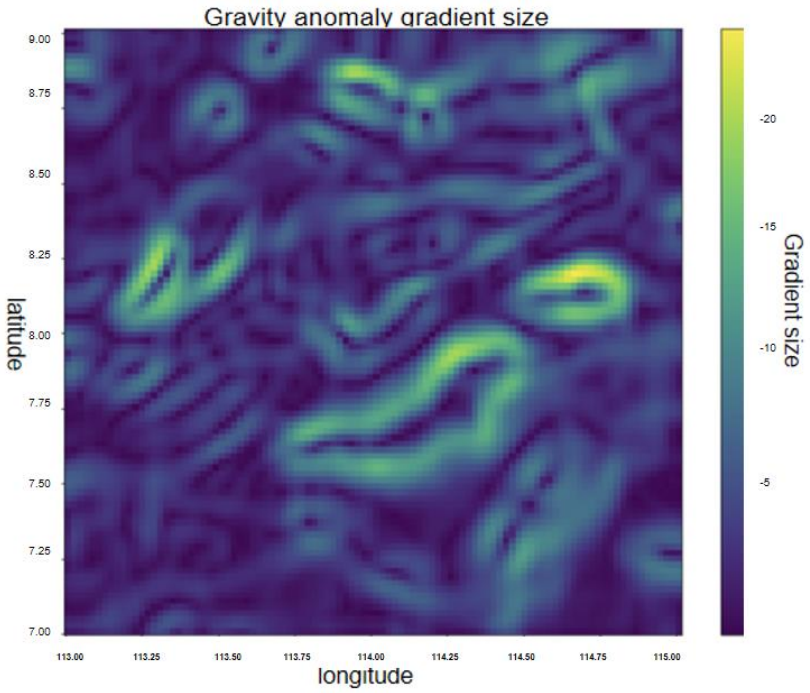

Figure 7. Gravity anomaly gradient size.

3.3.3. Migration prediction

The final prediction results are shown in Table 3.

Table 3. Shows the final forecast results.

Prediction Outcome Y	Prediction Outcome Probability_Medium Adaptability	Probability of prediction result _ low adaptability	Predict the probability of the result and have high adaptability	gradient
Low fit	0.00000191918113999994	0.9993046192294788	0.0006934615893812658	1.54825013851155
Medium fit	0.9999998288377726	0.35626331631112695	0.0007766803411680489	2.73185851706549
Medium fit	0.6431740371680038	1.0665220113329524e-7	0.0005626465208692089	3.00029991144412
Medium fit	0.9999998480902825	9.99624543947365e-8	5.086790928396163e-8	3.30522638746427
Medium fit	0.9999998415461216	9.987872337149594e-8	5.1947263240429716e-8	3.45642656526864
Medium fit	0.9999998456221955	9.955581953;39941e-8	5.351194368788237e-8	3.12296635191614
Medium fit	0.9999998488227237	9.917301727158578e-8	5.6768166423431554e-8	2.51028660653192
Medium fit	0.5780088529213298	9.979440891;13016e-8	5.928086128807926e-8	0.793122415152553
Low fit	6.492580618299389e-8	9.99859042613692e-8	0.00026330649449970544	1.11655283638608
Low fit	2.7702173668036675e-7	9.96102824988129e-8	5.633399308891479e-8	3.27393899662088
Medium fit	0.9999998288377726	6.621558645140071e-8	1.0494664096678644e-7	4.88856189434749

3.3.4. Applicability analysis

The applicability of the established forecasting system is relatively high, with the following points.

(1) Consistency of a data interpolation:

The prediction function of this system is based on test data, which has gone through the same difference processing steps to ensure the consistency of the input data. This preprocessing method is crucial to the generalization and transfer capabilities of the model, because it can improve the performance of the model on different data sets and enable it to better adapt to new environments and tasks.

By keeping the data consistent, it can ensure that the data used in the training and testing process of the model have the same features and distribution, thus avoiding the bias of the model in the generalization process. Such bias may cause the model to perform poorly on new datasets, so maintaining data consistency is an important step in improving the generalization ability of the model.

In addition, migration capability refers to the performance of the model on different datasets. If a model is trained on only one particular dataset, it may not be able to adapt to other datasets. Therefore, maintaining data consistency can improve the migration ability of the model so that it can perform well on different datasets.

(2) The universality of gradient features:

The magnitude of the gravity anomaly gradient is a generalized metric for describing changes in the local gravity field, regardless of the geological formation or geophysical setting. Since the gradient captures the core properties of the gravity anomaly, gradient-based prediction models show good adaptability to new regions.

(3) Generalization ability of c model:

If the prediction model shows excellent generalization ability on the data with 99.9% accuracy, it indicates that the model has a stable performance on the independent test set. Therefore, it can be expected that the model has strong prediction ability on new data as shown in Table 4.

Table 4. Model results.

	Accuracy	Recall	Precision	F1
Training set	0.999	0.999	0.999	0.999
Test set	0.999	0.999	0.999	0.999

(4) Continuity of geological and geographical conditions:

If the data originate from the same or similar locations and geographical conditions, then the applicability of the forecasting system F to new data will be enhanced because of this continuity. Due to the similarities between data, the model can better understand and adapt to new data, thereby improving its prediction accuracy on new data. This continuity can lead to better generalization performance and transfer capabilities, allowing the model to better adapt to different environments and tasks.

4. Conclusion

This article introduces an underwater navigation adaptation zone classification and prediction system based on gravity anomaly data. This system significantly improves the autonomous navigation accuracy and concealment of underwater vehicles in unknown sea areas through data processing, feature extraction, model construction and optimization. Bicubic interpolation technology was used to improve data resolution, build a gradient adaptability prediction model, and verify its effectiveness through migration prediction. This system not only solves the mismatch problem in traditional terrain matching navigation, but also achieves high-precision and high-reliability underwater autonomous navigation through machine learning and visualization technology. This system provides important support for ocean exploration and military applications, showing great potential in the development of the ocean field.

References

[1] Zhang Yujie, Li Houpu, Qiu Mengxin. Research on forward and inversion algorithm of three-dimensional gravity anomaly based on LinkNet network [J/OL]. Progress in Geophysics, 2024, 1-14.

[2] Zhang Haotian. Exploration of Adaptability of Linear Spatial Design in the Context of Territorial Spatial Planning: A Case Study of Lixi Town Tourist Scenic Area in Yingde City [J]. The House Collective, 2024, (19): 108-110+114.

[3] Zhang Jiaojiao, Du Yu. Real-time monitoring system of "Tianyuan" system based on Python [J]. Network Security Technology and Application, 2024, (07): 55-59.

[4] Wang Aoming, Li Shanshan, Fan Diao, et al. Research on real-time acquisition and simulation technology of underwater dynamic gravity data [J]. Marine Surveying and Mapping, 2022, 42(01): 18-24.

[5] Qi Shanlu. Research on GPR data processing visualization and disease risk assessment method [D]. Yantai University, 2023.

[6] Liu Zhijin, Liang Chuntao, Chen Chunmei, et al. Research on the application of wave field gradient method in Alaska[J/OL]. Geophysical and Geochemical Exploration Calculation Technology, 2024,1-14.

[7] Wu Zheng, Li Quanan, Chen Xiaoya, et al. Research progress of machine learning in magnesium alloy application[J/OL]. Chinese Journal of Engineering, 2024, 1-16.

[8] Tang Pengyu, Wang Zhen. Study on the residual time of sweat fingerprint on glass based on hyperspectral combined with multiple models [J/OL]. Progress in Laser and Optoelectronics, 2024, 1-7.

[9] You Y, Meng Yunlong, Wu Jingtao, et al. Estimation of automobile motion state based on whale optimization algorithm-support vector regression [J]. China Mechanical Engineering, 2024, 35(06): 973-981+992.

[10] Wang Yusheng, Yang Jufen, Liu Zhigang. Research on Multi-classification of Driver's Mental Fatigue State Based on Multi-feature Fusion of EEG Signals [J/OL]. Automotive Engineering Journal, 2024,1-11.