

# A Hybrid Computational Framework for Regional Classification in Gravity-Aided Navigation: Integrating Hierarchical Clustering and Random Forests

Yichen Yang<sup>1,\*</sup>, Jiayuan Du<sup>2</sup>

<sup>1</sup> Shenyang Aerospace University, Shenyang, Liaoning, China

<sup>2</sup> Dalian Polytechnic University, Dalian, Liaoning, China

\* Corresponding author: Yichen Yang (Email: 1918988932@qq.com)

**Abstract:** Data-driven gravity-aided underwater navigation requires reliable region-of-acceptance (RoA) classification to ensure robust localization. We formulate RoA prediction on geospatial gravity-anomaly fields as a pattern-recognition task and design a hybrid computational pipeline that couples unsupervised calibration with supervised inference. This paper proposes: (i) feature normalization over longitude–latitude–anomaly triplets; (ii) Ward hierarchical clustering to calibrate RoA labels; and (iii) a random-forest classifier trained on the calibrated labels, benchmarked against SVM, KNN, and decision-tree baselines. The trained model is directly transferred to a second dataset to assess cross-dataset generalization. Experiments show high accuracy and balanced precision/recall on held-out data, and 0.99 accuracy in the transfer evaluation, while maintaining low inference latency suitable for deployment. The framework offers a practical, generalizable solution for RoA prediction that can be integrated with downstream path-planning and decision modules.

**Keywords:** Hierarchical Clustering; Random Forest Classification; Domain Adaptation.

## 1. Introduction

Accurate regional classification in gravity-aided navigation is a critical challenge for underwater localization. The task can be formalized as pattern recognition on geospatial gravity anomaly fields, where heterogeneous anomaly distributions complicate calibration of reliable regions of acceptance (RoA). This issue is demanding due to the sensitivity of navigation systems to anomaly variations and the difficulty of ensuring consistent performance across zones [1,2].

Recent advances in machine learning and geospatial data mining have provided new opportunities for anomaly-based regional prediction. Hierarchical clustering has shown advantages in capturing structural similarity in spatial datasets [3], while ensemble classifiers such as random forests have demonstrated robustness against noise and non-linear feature interactions [4]. However, the integration of unsupervised calibration and supervised inference for anomaly-driven navigation remains underexplored, especially with respect to transferability across datasets.

This paper develops a hybrid computational framework that couples Ward-based hierarchical clustering with random forest classification to construct reliable RoA predictors. The proposed approach not only improves intra-dataset accuracy but also exhibits strong generalization performance when transferred to new anomaly datasets, achieving accuracy rates up to 0.99. Such performance highlights the practicality of the method for deployment in real-world navigation systems, contributing a generalizable solution for data-driven anomaly classification [5,6].

## 2. Hierarchical Clustering Framework for Regional Calibration in Gravity-Anomaly Fields

### 2.1. Data Standardization and Distribution Visualization

The raw dataset consists of longitude, latitude, and gravity anomaly measurements, which were first visualized to reveal the spatial heterogeneity of the anomaly field. The distribution map highlights regions of significant fluctuations, providing an intuitive reference for subsequent clustering and classification tasks.

To ensure comparability among features with different scales, the three attributes were normalized prior to model construction. The standardization procedure is expressed as:

$$x_{ij} = \frac{x_{ij} - m_j}{M_j - m_j}, \quad \text{where } M_j = \max\{x_{ij}\}, m_j = \min\{x_{ij}\} \quad (1)$$

This transformation aligns the ranges of longitude, latitude, and anomaly values, thereby eliminating the bias introduced by heterogeneous magnitudes and establishing a consistent basis for hierarchical clustering.

As shown in Figure 1, the processed distribution map clearly demonstrates localized gravity variations, which are critical for the identification of coherent regions of acceptance.

To identify coherent regions within the gravity anomaly field, an agglomerative hierarchical clustering (AGNES) framework was applied using the Ward linkage criterion. This method iteratively merges clusters in a bottom-up manner by minimizing the total within-cluster variance, thereby ensuring compact and well-separated groupings. By selecting an appropriate threshold on the resulting dendrogram, distinct region labels were derived and appended to the dataset, providing a structured partition of the anomaly field.

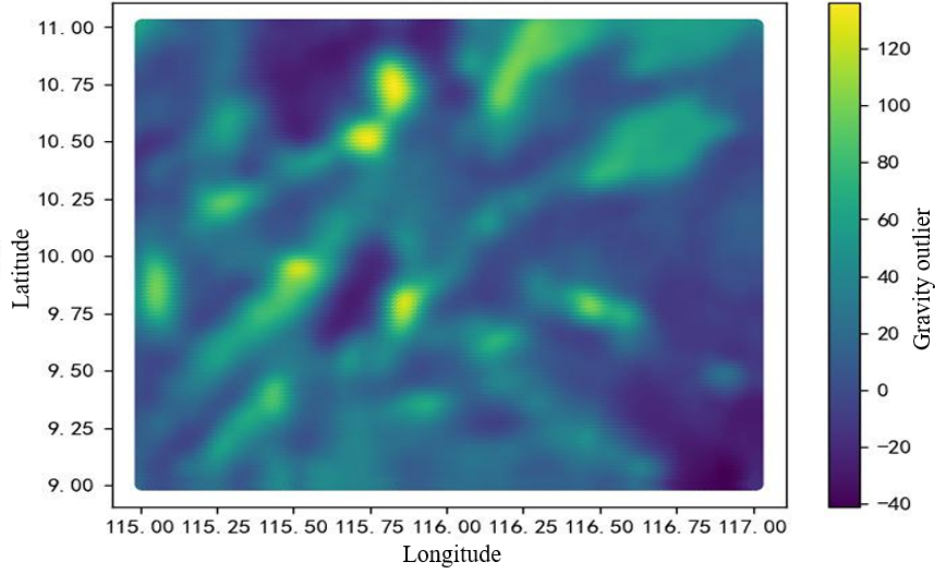


Figure 1. Distribution of gravity anomaly data across the study region

## 2.2. Hierarchical Clustering for Regional Partitioning

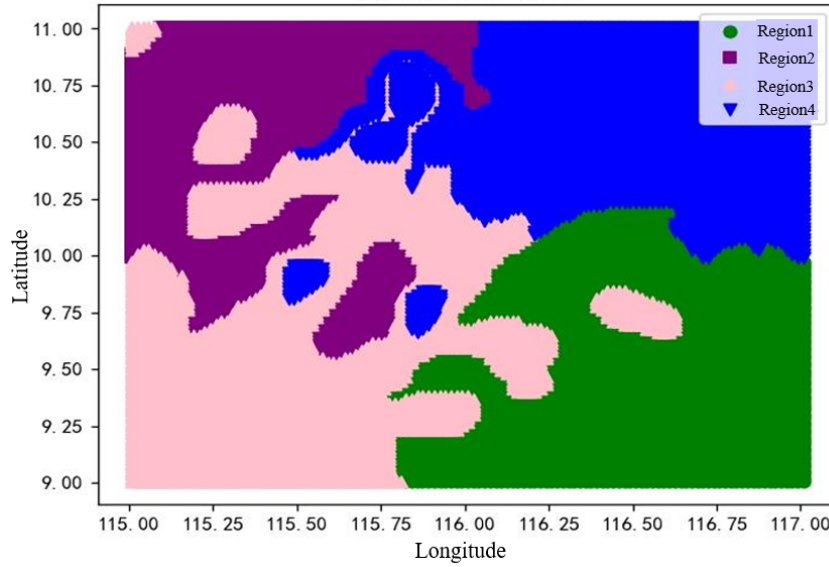


Figure 2. Visualization of regional partitioning based on ward hierarchical clustering

The clustering results reveal four major regions with distinct geospatial patterns, each exhibiting consistent gravity anomaly characteristics. Such regionalization enhances interpretability and supports subsequent classification tasks by providing reliable calibration labels. As shown in Figure 2, the four regions are visualized with distinct colors, highlighting clear spatial boundaries and heterogeneity across the study area.

## 3. Comparative Classification Modeling with Multiple Algorithms

For the classification task, longitude, latitude, and gravity anomaly values were selected as the input feature set  $X$ , while the region labels derived from clustering served as the target variable  $Y$ . This design ensures that both spatial coordinates and anomaly intensity contribute jointly to the prediction of regional categories. To obtain robust estimates,

the dataset was randomly partitioned into training and testing subsets, with 80% used for model fitting and 20% reserved for evaluation, ensuring reproducibility through fixed partitioning.

A random forest (RF) classifier was constructed as the primary model. RF is an ensemble method that aggregates the predictions of multiple decision trees, each trained on bootstrap samples with feature randomness. For an ensemble of  $T$  trees, the predicted class  $\hat{y}$  for an input  $x$  is given by majority voting:

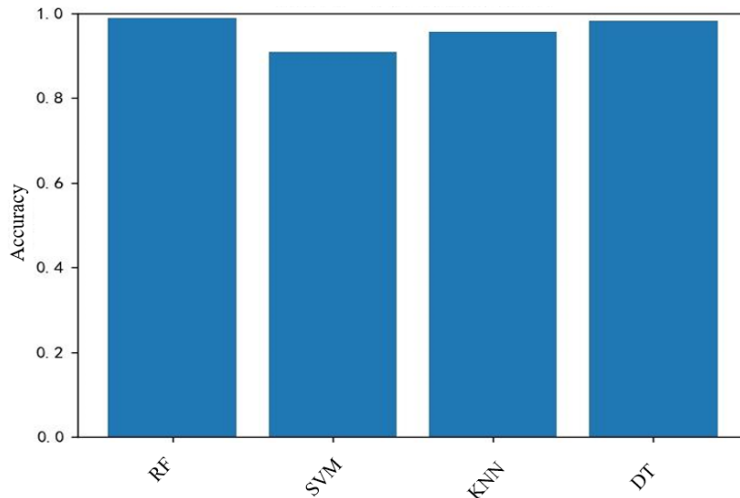
$$\hat{y} = \arg \max_{c \in C} \sum_{t=1}^T \mathbf{1}\{h_t(x) = c\}, \quad (2)$$

where  $h_t(x)$  denotes the prediction of the  $t$ -th tree. This ensemble structure reduces variance and improves generalization over single-tree models.

To benchmark performance, three widely used classifiers

were introduced for comparison. The support vector machine (SVM) seeks an optimal separating hyperplane that maximizes the margin between two classes, expressed as:

$$\min_{w,b} \frac{1}{2} \|w\|^2 \quad \text{s.t. } y_i (w \cdot x_i + b) \geq 1, \forall i, \quad (3)$$



**Figure 3.** Accuracy comparison of Random Forest, SVM, KNN, and Decision Tree classifiers

where  $w$  and  $b$  define the hyperplane. The k-nearest neighbor (KNN) algorithm performs classification by majority voting among the  $k$  closest training samples to a query point, with distance typically measured in Euclidean space. Finally, the decision tree classifier recursively partitions the feature space by selecting attributes that maximize information gain or minimize Gini impurity at each split, thereby forming interpretable hierarchical rules for classification.

The models were assessed using accuracy as the primary metric, alongside auxiliary indicators such as precision, recall, and  $F_1$ -score to provide a balanced evaluation of predictive performance. As shown in Figure 3, the random forest model achieved the highest accuracy, outperforming SVM, KNN, and the single decision tree classifier, thereby confirming its robustness and suitability for anomaly-based regional prediction.

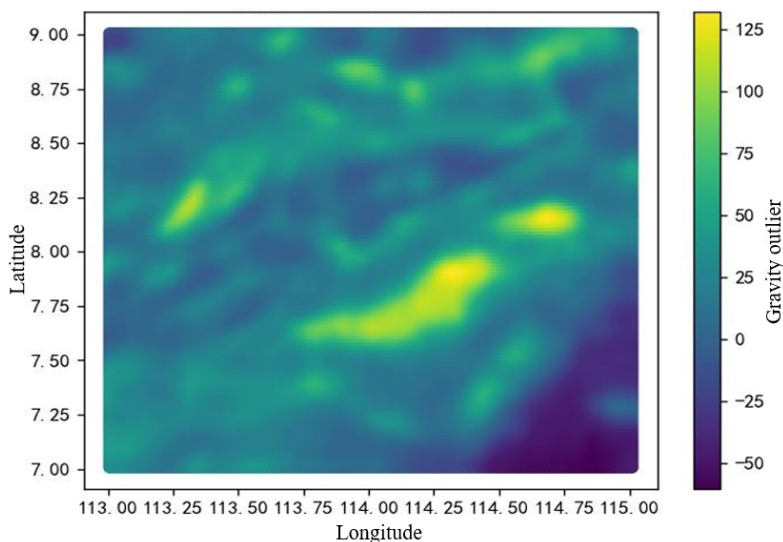
#### 4. Model Transferability and Applicability Assessment

To validate the generalization capability of the proposed

classification framework, a second dataset of gravity anomaly measurements was introduced. This dataset was visualized to examine the spatial distribution of anomalies and to highlight potential regional heterogeneity, as shown in Figure 4.

The previously established best-performing model from Problem 2 was directly transferred to this new dataset without additional retraining. Longitude, latitude, and gravity anomaly values were fed as input features, and the model produced predicted regional labels for each data point. The predicted labels were subsequently integrated into the dataset to facilitate further evaluation.

To assess the model’s applicability in a cross-dataset context, standard performance indicators were computed, including accuracy, precision, recall, and  $F_1$ -score. As summarized in Table 1, the model maintained consistently high predictive performance, with accuracy reaching 0.99. These results demonstrate both the robustness and transferability of the proposed regional classification framework, supporting its deployment in broader anomaly-based navigation scenarios.



**Figure 4.** Distribution of gravity anomaly data in dataset 2

**Table 1.** Evaluation metrics of the transferred classification model on dataset 2

Class	Precision	Recall	F1-score	Support
1	0.99	0.99	0.99	168
2	0.99	0.99	0.99	525
3	0.99	0.99	0.99	747
4	0.99	0.99	0.99	695
5	0.99	0.99	0.99	818

## 5. Conclusion

This paper proposes a hybrid computational framework that integrates Ward-based hierarchical clustering with random forest classification to address the problem of regional classification in gravity anomaly fields. The method achieves consistently high accuracy, reaching 0.99 in cross-dataset evaluation, and demonstrates robustness compared to alternative classifiers such as SVM, KNN, and decision trees. However, the framework may be limited by its reliance on fixed feature sets and by potential sensitivity to data imbalance in larger-scale applications. Future work will focus on incorporating advanced feature engineering and deep learning architectures to enhance adaptability, while extending the approach to real-time anomaly detection scenarios. Overall, the framework provides a scalable and generalizable solution for reliable anomaly-based regional prediction in computational navigation systems.

## Acknowledgments

The authors gratefully acknowledge the financial support from xxx funds.

## References

- [1] Pham N P. Deep learning for automatic geophysical interpretation with uncertainty quantification[M]. The University of Texas at Austin, 2022.
- [2] Lee I. Geospatial clustering in data-rich environments: Features and issues[C]//International conference on knowledge-based and intelligent information and engineering systems. Berlin, Heidelberg: Springer Berlin Heidelberg, 2005: 336-342.
- [3] Zhang D, Lee K, Lee I. Hierarchical trajectory clustering for spatio-temporal periodic pattern mining[J]. Expert Systems with Applications, 2018, 92: 1-11.
- [4] Talebi H, Peeters L J M, Otto A, et al. A truly spatial random forests algorithm for geoscience data analysis and modelling[J]. Mathematical Geosciences, 2022, 54(1): 1-22.
- [5] Kellenberger B, Tasar O, Bhushan Damodaran B, et al. Deep domain adaptation in earth observation[J]. Deep Learning for the Earth Sciences: A Comprehensive Approach to Remote Sensing, Climate Science, and Geosciences, 2021: 90-104.
- [6] Ott F. Representation learning for domain adaptation and cross-modal retrieval[D]. lmu, 2023.