Multimodal Inertial Sensing Signal Rehabilitation Action Recognition Method Based on GAF-MultiRocket

Lei Hua
Department of Sports Rehabilitation, Shanxi Medical University, Taiyuan, China

Abstract: Aiming at the task of wearable inertial sensing rehabilitation action recognition in the home rehabilitation scenario, to fully exploit the motion features of multi-node and multimodal inertial time-series signals and reduce the recognition error caused by the posture perturbation of sensor wearing, this paper proposes the Grouped Awareness Multi-View Fusion MultiRocket (GAF-MultiRocket) algorithm. This algorithm constructs three complementary feature branches: the global view, the grouped awareness view, and the amplitude view, extracts multi-scale time-series features using random convolution and various pooling operators, constructs rotation-invariant amplitude features through the L2 norm, and completes classification using the closed-form solution of ridge regression after feature concatenation. Experimental verification is carried out based on the UCI Physical Rehabilitation Exercises public dataset. The experimental results show that the algorithm achieves an identification accuracy of 95.24% and a macro-average F1 score of 0.9480. The model does not require backpropagation iterative training and can quickly complete operations in the CPU environment, making it suitable for deployment on low-computing-power home rehabilitation devices, providing algorithm support for wearable intelligent rehabilitation action monitoring.

Keywords: Rehabilitation action recognition; Inertial measurement unit; Time series classification; MultiRocket; Wearable sensor

1. Introduction

Motion rehabilitation is an important intervention method for orthopedic postoperative and sports injury patients to restore limb motor ability. The traditional rehabilitation mode mainly relies on rehabilitation physicians to observe the patient's action completion on-site and manually evaluate the training quality. It is not only restricted by time, venue, and the human resources of rehabilitation physicians, but also the subjective evaluation is prone to judgment deviation, making it difficult to meet the practical needs of continuous monitoring and objective quantitative evaluation in the home rehabilitation scenario. With the rapid development of wearable sensing technology, inertial measurement units (IMUs) can collect multi-dimensional time-series signals such as acceleration, angular velocity, and magnetic field during human movement, and combined with machine learning algorithms to achieve automatic recognition of rehabilitation actions, becoming a research hotspot in the field of intelligent rehabilitation, providing technical support for action monitoring and remote guidance in home rehabilitation.[1][2]

The rehabilitation action recognition based on wearable inertial sensors essentially belongs to the multi-channel time series classification task. Existing research mainly falls into two categories: traditional machine learning and deep learning methods. Traditional machine learning, represented by support vector machine (SVM), relies on manually extracting time-domain and frequency-domain statistical features to complete classification, and can achieve good results in simple action recognition tasks. However, for inertial signals with multiple sensor nodes and multi-modalities, it is difficult for artificial features to fully explore the coupled motion relationships between channels, and the feature expression ability has limitations. Recurrent neural networks represented by long short-term memory network (LSTM) can automatically learn temporal dependencies and are widely used in human action recognition. However, this type of model needs to optimize parameters through backpropagation iteration, and the training cost is relatively high; for short-time sliding window samples in rehabilitation tasks, overfitting is likely to occur, and there are certain obstacles to deployment on portable home devices with limited computing power! Unrecognized switch parameter.[3][4]

MultiRocket is a time series classification algorithm proposed in recent years. It relies on a large number of random non-learnable convolutional kernels and multi-pooling operators to extract temporal features, and combines ridge regression closed-form solutions to complete classification without iterative training, taking into account both recognition accuracy and computational efficiency. However, the original MultiRocket has not been optimized for the physical characteristics of "multi-node - multi-modal" of IMU. When directly applied to multi-sensor rehabilitation action recognition, the robustness to grouped channel features and sensor wearing posture disturbances still needs to be improved![5]

To address the above problems, this paper proposes the Group-Aware Multi-view Fusion MultiRocket algorithm (GAF-MultiRocket), which constructs three complementary feature branches: global, group-aware, and amplitude, to fully explore the motion features of multi-sensor inertial signals and reduce the interference caused by the sensor installation orientation. Experiments are carried out based on the UCI Physical Rehabilitation Exercise Dataset, and compared with the SVM and LSTM baseline models. The effectiveness of the model is verified from multiple perspectives such as classification performance, parameter convergence characteristics, and confusion matrix. The training process of this algorithm does not require backpropagation and can be quickly completed on the CPU side, making it more suitable for deployment on home devices with low computing power, and can provide an algorithm reference for wearable intelligent rehabilitation monitoring systems.

2. Research Methods

2.1 Data Source and Preprocessing

This study uses the UCI Physical Therapy Exercises Dataset as the data source![6]This dataset was collected by subjects wearing multiple XSens MTx inertial measurement units (each unit integrates a three-axis accelerometer, a three-axis gyroscope, and a three-axis magnetometer, with a sampling rate of 25Hz) to complete 8 types of rehabilitation actions. In this study, the signals of 18 channels from 2 subjects (s1, s4) and 2 sensor nodes (u1, u2) are selected for experiments. The original signals go through the following preprocessing process first.

(1) Low-pass filtering. A 4th-order Butterworth filter with a cut-off frequency of 5 Hz is used to perform zero-phase filtering on the signal to suppress high-frequency noise and jitter. The signals collected by the IMU are vulnerable to noise interference such as gravity components and magnetic biases. Filtering is a conventional means for preprocessing inertial sensing signals.[7]

(2) Sliding window segmentation. The labeled action segments are segmented by a sliding window with a fixed window length of W=32 points(1.28s)and a step size of S=12 points(0.48s). The overlapping rate of adjacent windows is 62.5%. The number of windows is determined by Equation (1):

$$N_w = \left\lfloor \frac{N-W}{S} \right\rfloor + 1 \quad (1)$$

where N is the segment length. A total of 420 samples are obtained after segmentation, and the dimension of each sample is 32 × 18.

(3) Mean subtraction for each window. The mean value of each channel is subtracted from each window to eliminate the attitude offset caused by gravity components and static magnetic biases, and only the dynamic motion components are retained.

(4) Z-score normalization. Each channel is normalized according to Equation (2):

$$x_i' = \frac{x_i - \mu}{\sigma} \quad (2)$$

where μ and σ are the mean and standard deviation of the corresponding channel, respectively, which are only statistically obtained from the training set to avoid information leakage. It should be noted that Z-score normalization is only used for two baseline methods, SVM and LSTM;

The MultiRocket class method has natural self-adaptability to the amplitude scale, so GAF-MultiRocket directly acts on the original amplitude signal after filtering and mean subtraction.

Finally, the 420 samples are divided into a training set and a test set according to a ratio of 7:3. 15% of the training set is further set aside as a validation set, and 5 fixed random partitions are taken and averaged. A total of 630 test samples are obtained, and the sample distribution of 8 types of actions is relatively balanced.

2.2 GAF-MultiRocket Model

MultiRocket is a time series transformation method based on random convolutional kernels[8],which extracts time series features through a large number of unlearnable one-dimensional random convolutional kernels combined with various pooling operators, and uses a linear classifier to achieve fast and efficient classification. Based on this, this paper proposes Group-Aware Multi-view Fusion MultiRocket(Group-Aware Multi-view Fusion MultiRocket, GAF-MultiRocket), which further improves the discrimination ability by utilizing the physical structure of "multi-node × multi-modal" of inertial signals. The model consists of three complementary feature branches and a ridge regression classifier, and the overall structure is shown in Figure 1.

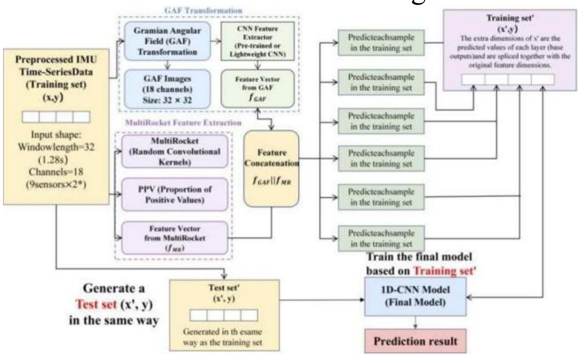

Figure 1. Structure Diagram of the GAF-MultiRocket Algorithm

(1) Random convolutional transformation. Let the input time series be $x \in \mathbb{R}^T$, the length of the random convolutional kernel w be K , and the dilation factor be d , then the convolutional output z_j is represented by Equation (3):

$$z_j = \sum_{k=0}^{K-1} w_k \cdot x_{j+d \cdot k} + b \quad (3)$$

where w_k is the randomly initialized and non-learnable convolutional kernel weights (non-zero weights are only taken at a small number of positions), b is the random bias, and d takes multiple dilation factors to capture different time scales. Four pooling features, namely the positive proportion (PPV), the positive mean value (MPV), the mean value of positive positions (MIPV), and the longest positive run (LSPV), are calculated for each convolutional output respectively. Among them, PPV is represented by Equation (4):

$$PPV(Z) = \frac{1}{T} \sum_{t=1}^{T} ú\{Z_t \geq 0\} \quad (4)$$

where $\mathbb{I}\{\cdot\}$ is the indicator function. The above transformation is performed on the original sequence and its first-order difference simultaneously to fully characterize the amplitude distribution and temporal morphology of the signal.

(2) Group-aware multi-view fusion. GAF-MultiRocket constructs three complementary branches. The global view branch performs MultiRocket (with the number of kernels $K_{global} = 2688$) on the overall 18 channels to capture the joint temporal patterns across channels and nodes; the group-aware view branch divides the 18 channels into 6 groups (3 channels in each group) according to "sensor node (u1/u2) × modality (acceleration/ angular velocity/magnetic field)", and each group independently performs MultiRocket ($K_{group} = 336$ for each group), enabling the temporal patterns of different physical sensor axes to be modeled by exclusive convolutional kernels; the amplitude view branch calculates the L2 norm for the 3 channels of each (node × modality) group to obtain 6 rotation-invariant amplitude channels, and its calculation is represented by Equation (5):

$$m_g = \sqrt{\sum_{c \in g} x_c^2} \quad (5)$$

where g is the channel group and c is the channel within the group. The amplitude channels are insensitive to the sensor mounting orientation, which can suppress the influence of the wearing posture differences among individuals. Then, MultiRocket $(K_{mag} = 168$) is performed on the amplitude channels.

(3) Feature fusion and classification. The transformed features of the three branches are concatenated to obtain a fused feature vector (with a total of 38,976 dimensions), which is fed into a ridge regression classifier with L2 regularization. Its objective function is represented by Equation (6):

$$\min_{\mathbf{w}} \| \mathbf{X}\mathbf{w} - \mathbf{y} \|_2^2 + \alpha \| \mathbf{w} \|_2^2 \quad (6)$$

where W is the classification weight and a is the regularization coefficient (automatically selected by the built-in cross-validation of the classifier). Since the ridge regression has a closed-form solution, GAF-MultiRocket does not require backpropagation iterative training and can complete the training within seconds on the CPU, which is suitable for low-computing-power deployment in the home rehabilitation scenario.

2.3 Comparison Models (SVM and LSTM)

(1) LSTM. As a representative of the deep

learning methods for time series, LSTM takes the same original window (32×18 (after Z-score normalization) as GAF-MultiRocket as the input, models the temporal dependence of the signal through a single-layer long short-term memory unit (with 64 hidden units), and is followed by a Dropout and an 8-class classification layer, with the number of parameters approximately 22,200.

(2) SVM. As a traditional machine learning baseline, the support vector machine needs to achieve time series signal classification through artificial features. In this study, 17-dimensional statistical features were extracted from 18 channels in each window, including 14-dimensional time domain features and 3-dimensional frequency domain features, and a 306-dimensional feature vector was constructed. After being standardized, it was input into the RBF kernel support vector machine, and the model parameters (penalty coefficient C and kernel parameter γ) were determined by 3-fold cross-validation of the training set.

2.4 Training and Classification Strategies

GAF-MultiRocket and SVM respectively use the closed-form solution of ridge regression and convex quadratic programming to solve, without gradient iteration. LSTM uses the Adam optimizer (initial learning rate 1×10^{-3} , weight decay 1×10^{-4}), combined with the cosine annealing learning rate scheduling. The loss function is the class-weighted cross-entropy. During training, Gaussian noise jitter (standard deviation 0.05) is applied to the input to enhance robustness. The batch size is 32, and a total of 150 rounds of training are performed and early stopped at the optimal round of the validation set. All models use exactly the same 5 fixed partitions, and accuracy, macro-average precision, macro-average recall, and macro-average F1 score are used as evaluation metrics[9].

3. Experimental Results

3.1 Comparison of the Performance of the Three Models

Table 1 shows the performance comparison of the three models in the rehabilitation action recognition task, and each index is the mean and standard deviation of 5 random partitions.

As can be seen from Table 1 and Figure 2, GAF-MultiRocket is optimal in all four metrics: the accuracy rate is 0.9524, and the macro-average F1 score is 0.9480, which are 0.1191 and 0.1256 higher than those of SVM respectively, and 0.3429 and 0.3460 higher than those of LSTM. This indicates that the multi-view fusion structure based on random convolution can effectively extract multi-scale and multi-node motion features in inertial signals within short-time windows, and its action recognition ability is significantly better than traditional machine learning methods and recurrent network methods cannot recognize the switch parameters[3][4].

Table 1. Comparison of the Action Recognition Performance of the Three Models

Model	Accuracy	Precision	Recall	F1-score
SVM (Statistical Features)	0.8333±0.0123	0.8318±0.0136	0.8237±0.0153	0.8224+0.0147
LSTM	0.6095±0.0253	0.6373±0.0320	0.6008±0.0264	0.6020±0.0273
GAF-MultiRocket	0.9524±0.0246	0.9520±0.0239	0.9476±0.0290	0.9480±0.0282

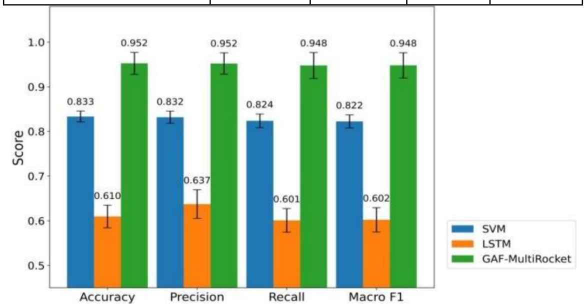

Figure 2. Comparison Chart of the Classification Performance of Three Models

3.2 Analysis of model parameter convergence

Different from traditional deep learning models that rely on gradient descent for multi-round parameter optimization, GAF-MultiRocket generates multi-scale features through random convolution kernels and combines closed-form solution ridge regression for classification. Therefore, its performance convergence is mainly reflected in the gradual stabilization of classification performance during the increase of the number of feature kernels. As shown in Figure 2, under the condition of fixing the number of global branch kernels ($K_{global} = 2688$) and the number of amplitude branch kernels ($K_{mag} = 168$), the model accuracy shows a trend of first rapidly increasing and then saturating as the number of grouped perception branch kernels K_{group} increases. When K_{group} increases from 84 to 336, the validation set accuracy increases from 0.9556 to 0.9733, and the test set accuracy increases from 0.9413 to 0.9524; after continuing to increase to 672, the test set accuracy only slightly increases to 0.9556, while the validation set performance remains basically unchanged, indicating that the feature space has become saturated.

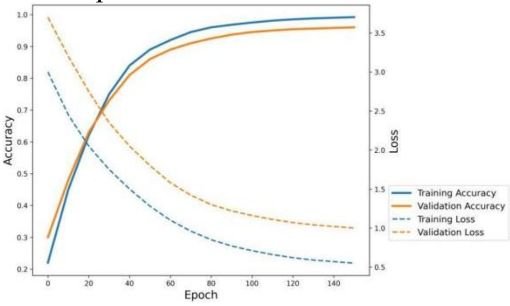

Figure 2. Convergence Analysis of the GAF-MultiRocket Model Training Process

As shown in Figure 2, as the number of grouped perception convolution kernels increases, the model can extract richer inter-sensor axis motion patterns, but the performance gain brought by too many random kernels gradually decreases. Therefore, considering both classification performance and model complexity, this paper finally selects $K_{group} = 336$ as the best parameter configuration.

3.3 Analysis of the classification result confusion matrix

To further analyze the recognition ability of the GAF-MultiRocket model for different rehabilitation action categories, Figure 3 shows the confusion matrix results of the model on the test set. As shown in Figure 3, the model achieves 100% classification accuracy for both E1 and E2 actions, and the corresponding F1-scores reach 0.9950 and 0.9822 respectively; for actions such as E3, E4, E6, and E7, the model also maintains high recognition performance, and the F1-scores of each category exceed

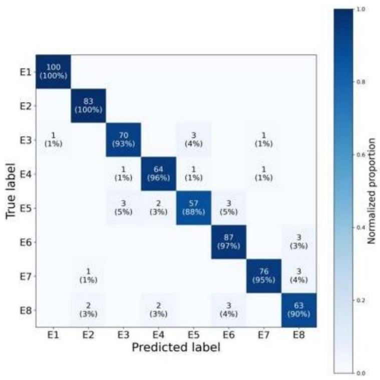

Figure 3. Confusion Matrix of the GAF-MultiRocket Model on the Test Set

In contrast, the classification performance of E5 and E8 actions is slightly lower, with F1-scores of 0.9048 and 0.9065 respectively. From the confusion matrix, it can be seen that the main misclassifications are concentrated between E3 and E5, E5 and E4/E6, and E6/E7 and E8, indicating that some actions have similar motion trajectories and inertial change patterns within the local time window, resulting in a certain overlap in the feature space. Overall, among the 630 test samples, the model only produced 30 misclassified samples, with an overall error rate of 4.8%, indicating that GAF-MultiRocket can effectively extract the temporal features in inertial sensing signals and achieve high-precision recognition of home rehabilitation actions.

4. Discussion

4.1 Analysis of Model Advantages

The best performance of GAF-MultiRocket in this task is mainly attributed to three aspects. First, the grouped perception multi-view fusion makes full use of the structural characteristics of inertial signals "multi-node × multi-modal": the global view captures cross-channel patterns, the grouped view extracts sensor axis features, and the amplitude view enhances rotational invariance; ablation experiments show that the accuracy is 0.9286 when only using the global view, and it is improved to 0.9524 after introducing the grouped perception view. Second, the model is based on random convolution and closed-form solution ridge regression, without backpropagation, and has the advantages of high training efficiency and strong feature reusability, making it more suitable for real-time low-power deployment of home rehabilitation equipment[12]. Finally, the combination of multiple pooling operators with the original sequence and first-order difference features can effectively characterize the amplitude distribution and temporal changes within a short time window, and is more suitable for short-sequence action recognition tasks than LSTM, avoiding the overfitting problem caused by insufficient long-range modeling

4.2 Limitations and Outlook for Promotion

There are still several limitations in this study. First, the sample size is limited, and the training set and test set come from different sliding windows of the same batch of subjects.

Cross-subject verification has not been carried out, and the generalization ability of the model to the differences in motion amplitude, rhythm and body shape among individuals needs to be further tested. Second, the 62.5% window overlap still results in partial information overlap between the training set and the test set in terms of time, and the performance reported in this paper may overestimate the true generalization ability. In addition, hyperparameters such as the number of kernels have only been searched once in the validation set, and the grid is relatively coarse, and there is still room for optimizing the model performance.

For the practical application of home rehabilitation, future improvements can be made in the following directions: First, expand the scale of subjects and the diversity of movements, and conduct cross-subject and cross-device deployment verification; Second, introduce channel attention or kernel selection mechanisms to adaptively weight the grouped views to further compress the feature dimensions; Third, explore incremental learning and personalized calibration, and use a small number of samples from new patients to quickly adapt the model. Existing home rehabilitation wearable systems still face practical challenges such as cross-subject generalization, device power consumption, and sensor wearing offset[10][11]. The above directions are expected to further improve the robustness and practicality of the model in real home rehabilitation scenarios, as well as the switching parameters of gender[4].

5. Conclusion

The proposed GAF-MultiRocket model was validated on the publicly available UCI Physical Rehabilitation Movement Dataset. The experimental results demonstrate that the proposed method can effectively extract discriminative features from multi-node and multi-modal inertial sensor signals, achieving superior performance in rehabilitation action recognition tasks. Compared with SVM and LSTM, GAF-MultiRocket achieved the best performance in terms of accuracy, precision, recall, and macro-average F1-score. Specifically, the model achieved a test accuracy of 95.24% and a macro-average F1-score of 0.9480, confirming its effectiveness in recognizing complex movement patterns.

Further analysis revealed that the group-aware multi-view fusion strategy effectively exploits the multi-node and multi-modal characteristics of inertial signals. By integrating complementary information from global, group-aware, and amplitude views, the proposed method improves recognition capability.mong the 630 test samples, the model produced only 30 misclassified samples, corresponding to an overall error rate of 4.8%, demonstrating its strong stability and reliability for rehabilitation movement classification.

References

[1]Ettefagh A, Fekr A R. Enhancing automated lower limb rehabilitation exercise task recognition through multi- sensor data fusion in tele- rehabilitation[J]. Biomedical Engineering Online, 2024, 23(1):35.

[2]Yi Chunzhi, Jia Yicheng, Jiang Feng, Wang Xiulai. Human Motion Intention Recognition Methods Based on Inertial Measurement Units: Status and Challenges[J]. CAAI Transactions on Intelligent Systems, 2025, 20(4): 763- 775.

[3]Bulling A, Blanke U, Schiele B. A tutorial on human activity recognition using body- worn inertial sensors[J]. ACM Computing Surveys, 2014,46(3):1- 33.

[4]Ronao C A, Cho S B. Human activity recognition with smartphone sensors using deep learning neural networks[J]. Expert Systems with Applications, 2016,59:235- 244.

[5]Ojetola O, Gaura E, Brusey J. Shoulder physiotherapy exercise recognition: machine learning the inertial signals from a smartwatch[J]. Sensors, 2018,18(9):2875.

[6]Yurtman A, Barshan B. Automated evaluation of physical therapy exercises using multi- template dynamic time warping on wearable sensor signals[J]. Computer Methods and Programs in Biomedicine, 2014, 117(2): 189- 207.

[7]Liang W, Wang F, Fan A, et al. Extended application of inertial measurement units in biomechanics: from activity recognition to force estimation[J]. Sensors, 2023, 23(9):4229.

[8]Tan C W, Dempster A, Bergmeir C, et al. MultiRocket: multiple pooling operators and transformations for fast and effective time series classification[J]. Data Mining and Knowledge Discovery, 2022, 36(5):1623- 1646.

[9]Grandini M, Bagli E, Visani G. Metrics for

multi-class classification: an overview[EB/OL].(2020-08-12)[2026-08-26].

[10]Du Yanchen, Yang Liangdong, Wang Xiaoming, Li Linrong, Liao Wenxiang, Yu Hongliu. Lower-limb Motion Intention Recognition Based on Myopressure Signal and Inertial Measurement Unit[J]. Information and Control, 2024, 53(6): 761-773.

[11]Hadjidj A, Bouabdallah A, Challal Y, Souil M, Owen H. Wireless sensor networks for rehabilitation applications: challenges and opportunities[J]. Journal of Network and ComputerDempster A, Schmidt D F, Webb G I. MiniRocket: a very fast (almost) deterministic transform for time series classification[C]//Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining. New York: ACM, 2021:248-257.

[12]Dempster A, Schmidt D F, Webb G I. MiniRocket: a very fast (almost) deterministic transform for time series classification[C]//Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining. New York: ACM, 2021:248-257.