Sensors and Materials, Vol. 38, No. 4 (2026) 2271–2285
MYU Tokyo

S & M 4436

Sensor Array Signal Classification Algorithm Based on Big Data Mining Fusion

Jiana Bi,* ChunCheng Wei, and Ning He

School of Software and Big Data, Changzhou College of Information Technology,
No. 22 Mingxin Middle Road, Science and Education City, Changzhou, Jiangsu 213164, China

(Received February 9, 2026; accepted April 21, 2026)

Keywords: sensor array, signal classification, big data mining, feature extraction, deep learning

With the rapid development of the Internet of Things and intelligent sensing technology, sensor array systems have been widely applied in various fields. However, traditional methods struggle to effectively process high-dimensional nonlinear signals, and existing deep learning models suffer from feature redundancy and excessive computational complexity. In this paper, we propose a sensor array signal classification algorithm fused with big data mining, achieving accurate and efficient classification through multistage optimization. Firstly, a hybrid feature engineering framework is constructed, combining signal time–frequency analysis with genetic-algorithm-rough-set-based feature selection to reduce feature dimensionality while retaining discriminative information. On this basis, an attention-mechanism-enhanced long short-term memory fully convolutional network classification model is designed, which dynamically focuses on key response segments through attention weights. Experimental results show that the proposed algorithm outperforms traditional methods on multiple public datasets. Moreover, the algorithm achieves good balance between parameter quantity and inference time, providing feasibility for deployment on resource-constrained edge sensor nodes. This algorithm relies on high-quality annotated data, suffers from insufficient model interpretability, and faces domain shift issues, with remaining room for optimization in edge deployment. These problems will be addressed by integrating few-shot learning, interpretable design, and lightweight technologies in future work.

1. Introduction

As an important means of modern information acquisition, sensor array technology plays a crucial role in numerous fields such as industrial control, medical diagnosis, environmental monitoring, and security surveillance.[1,2] Compared with a single sensor, a sensor array can synchronously acquire spatially distributed information and collect multichannel, multidimensional signal data, to form massive signal datasets. These data contain abundant target features and environmental information, significantly improving the reliability of target detection and recognition.[3,4] However, with increases in the number of sensors and sampling

*Corresponding author: e-mail: 544099426@qq.com
https://doi.org/10.18494/SAM6281

ISSN 0914-4935
https://myukk.org/

frequency, the amount of data generated by sensor arrays per second grows exponentially, presenting typical big data challenges. Traditional signal processing algorithms cannot meet the requirements of real-time processing and effective information extraction owing to their high computational complexity and poor adaptability. Therefore, integrating big data mining technology to design efficient and accurate sensor array signal classification algorithms is of great practical value for improving the overall performance of sensor array systems. Research on sensor array signal classification can be roughly divided into three categories.

First, we look at traditional signal processing algorithms, with the multiple signal classification (MUSIC) algorithm being representative. The MUSIC algorithm estimates the direction of signal arrival through performing eigenvalue decomposition on the covariance matrix of the sensor array signal to obtain the signal subspace and the orthogonal noise subspace, utilizing their orthogonality.[5] Tang *et al.*[6] improved the MUSIC algorithm for vector hydrophone arrays by establishing a signal model for the vector array and proposing a strategy combining noise compensation and whitening to restore the orthogonality of the noise subspace and enhance algorithm robustness. Unzueta and Romero[7] combined generalized normalization techniques with the MUSIC algorithm, achieving smaller errors; compared with other types of interferometric sensors, the developed system offers advantages such as its implementation simplicity and ability to measure signals from the acoustic environment around each sensor.

The second category includes classical machine learning algorithms, represented by support vector machine (SVM) and random forest (RF). SVM shows stable performance in sensor signal classification and is especially suitable for small-sample, high-dimensional feature scenarios, with strong generalization ability. RF can effectively handle high-dimensional features, exhibits certain robustness to outliers and overfitting, and can evaluate feature importance. Dyah *et al.*[8] conducted a comparative study of RF and SVM for gas sensor array classification. Laref *et al.*[9] proposed a generalized pattern search algorithm to optimize SVM hyperparameters and compared it with other optimization algorithms, demonstrating it to be an efficient and robust solution. Wang *et al.*[10] utilized model parameters obtained from training SVM to construct a classification function; the algorithm is simple and computationally fast, making it suitable for embedded processor applications. Gupta *et al.*[11] researched federated learning based on Bayesian weighted RF, enabling multiple sensor nodes to collaboratively train a global intrusion detection model without sharing local sensitive data. Liu *et al.*[12] pointed out that among many machine learning algorithms, RF, owing to its good anti-overfitting capability and insensitivity to noise, can achieve sufficient generalization ability with relatively little training data.

The third category is composed of deep learning algorithms, represented by convolutional neural networks (CNNs), long short-term memory (LSTM), and CNN-LSTM hybrids. CNN can effectively identify specific patterns in signal waveforms for classification. LSTM is wellsuited for processing sensor data with temporal dynamic characteristics. CNN-LSTM can simultaneously capture spatiotemporal features of signals, making it a powerful model for handling sensor array signals. Eisele *et al.*[13] achieved higher classification accuracy for array sensors when feeding preprocessed transducer signals and source maps simultaneously into a CNN, significantly outperforming traditional single-element sensors. Heng *et al.*[14] showed that training CNN models on sensor array datasets yields higher results than using single-sensor

datasets. Tello *et al.*[15] evaluated five classification models, with an improved LSTM model based on LeNet-5 with the highest performance, achieving 100% accurate classification using only 30 seconds of data from each 360-second sample of the sensor array. Ravi and Hardeep[16] reported simulation results indicating that the proposed LSTM model achieved comparable performance to a Nonlinear AutoRegressive method implemented as an artificial neural network in terms of system throughput. Li *et al.*[17] proposed a new detection framework combining a CNN-Transformer-LSTM deep learning model with Time2Vec encoding; comparative evaluation confirmed the superior accuracy and robustness of theirproposed model.

The performance of traditional signal processing algorithms degrades significantly in complex real-world environments, and they have high computational complexity and are sensitive to prior information, limiting their application in real-time, nonideal scenarios. Classical machine learning algorithms can effectively handle high-dimensional features and have certain robustness to outliers and overfitting, but their model interpretability is relatively complex, and their ability to handle long-range dependences in time-series data is limited. Deep learning algorithms can effectively identify specific patterns in signal waveforms for sensor signal classification, but they lack the ability to model long-term temporal dependences in sequence data and are inefficient in extracting local spatial features. The genetic algorithm–rough set–attention LSTM fully convolutional network (GA-RS-ALSTM-FCN) proposed in this paper is an advanced hybrid model integrating feature selection and deep learning. Its core is the ALSTM-FCN classifier, which simultaneously models the long-term temporal dependences and local spatial features of sensor signals through parallel LSTM and fully convolutional network branches, and introduces an attention mechanism to dynamically focus on key information segments. On this basis, the front-end of the model integrates a hybrid feature selection module combining GA and RS theory to optimize the input feature subset and remove redundant information, thereby improving model efficiency and generalization performance. The algorithm realizes end-to-end integration from feature optimization to classification modeling, achieving higher classification accuracy and robustness with more refined features and a more powerful model.

The signal features, classification objectives, and performance constraints of this study are determined by the working mechanisms and material properties of sensors, and the algorithm is designed to adapt to the output characteristics of sensors. Without the physicochemical responses of sensors and sensitive materials, the algorithm would lack processing objects and application carriers. Moreover, the algorithm performance directly determines the final recognition accuracy and practical level of the sensor array system, and the two are mutually supportive and inseparable. Most existing studies focus separately on feature selection or model structure optimization. This study innovatively integrates GA-RS hybrid feature selection with attention-enhanced LSTM-FCN to form an end-to-end efficient classification framework. It reduces dimensionality while enhancing key temporal feature extraction, balancing accuracy and computational efficiency for better adaptation to edge sensor deployment. The proposed algorithm can, in contrast, support the research and development of novel sensors and sensitive materials. On the one hand, it extracts weak response features of sensors via high-precision classification, providing data support for optimizing material sensitivity and selectivity. It

reduces the hardware cost of sensor arrays through feature dimensionality reduction and lightweight inference, promoting the design of low-power and miniaturized devices. Moreover, the algorithm can mine potential laws of sensor responses, offering an efficient analysis tool for the verification of novel sensing mechanisms and concepts, and improving the overall performance of sensing systems from materials to applications.

2. Models and Methods

The proposed sensor array signal classification algorithm consists of three core modules: signal preprocessing and feature extraction, feature selection and dimensionality reduction, and the signal classification model. The overall framework is shown in Fig. 1. Signal preprocessing and feature extraction module: First, the raw sensor array signals undergo preprocessing, including wavelet denoising, baseline correction, and normalization. Subsequently, features are extracted from multiple dimensions: time-domain features (12 dimensions), frequency-domain features (15 dimensions), time–frequency features (20 dimensions), and nonlinear features (10 dimensions), outputting an initial feature vector of 57 dimensions.

Feature selection and dimensionality reduction module: A hybrid feature selection method combining GA and RS is adopted. The GA evaluates the quality of feature subsets through a fitness function and performs iterative optimization; RS theory is used to calculate feature dependence for further reduction of redundant features. The output is an optimized feature subset (approximately 35 dimensions).

Signal classification model: An ALSTM-FCN is employed. The LSTM branch captures the time-series dependences of signals, while the FCN branch extracts the local spatial features of signals. The attention mechanism layer performs weighted fusion on the outputs of the two branches, and the Softmax classifier outputs the final class probability distribution.

2.1 Signal preprocessing and feature extraction

Raw sensor array signals usually contain noise, drift, and nontarget responses, requiring preprocessing to improve the signal-to-noise ratio. We adopt wavelet transform for signal denoising and baseline correction, with the following mathematical expression:

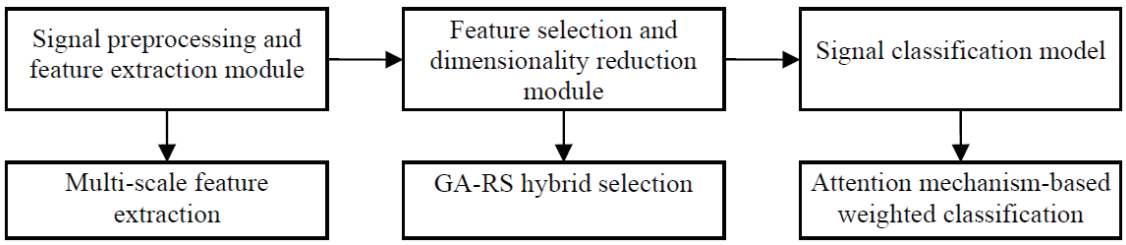

Fig. 1. Overall framework of the sensor array signal classification algorithm.

$$W_T(a,b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} x(t) \psi^* \left(\frac{t-b}{a} \right) dt, \tag{1}$$

where $x(t)$ is the original signal, $\psi(t)$ is the wavelet basis function, and a and b are the scale and translation parameters, respectively. The wavelet transform provides the local characteristic information of the signal in both time and frequency domains, making it suitable for nonstationary signal processing.

After preprocessing, features are extracted from three dimensions: time-domain features, including, for example, mean, variance, and crest factor; frequency-domain features, obtained via fast Fourier transform (FFT), such as power spectral density, and centroid frequency; and time–frequency domain features, extracted via wavelet transform, like scalogram features. These features constitute the initial feature set $F_{init} \in R^{N \times D}$, where N is the number of samples and D is the feature dimensionality. The types and descriptions of the extracted features are listed in Table 1.

2.2 Feature selection and dimensionality reduction

The initial feature set often contains a large number of redundant and irrelevant features that not only increase computational burden but also may reduce model generalization ability. We propose a hybrid feature selection method combining GA and RS, with the flow shown in Fig. 2.

The algorithm terminates when one of the following conditions is met: the maximum number of iterations is reached, the fitness function converges to a stable value, or a feature subset that satisfies predefined criteria is found. The algorithm exhibits the following characteristics: Global search capability: The GA can effectively explore the feature subset space and avoid falling into local optima. Attribute reduction: RS theory can eliminate redundant features while preserving classification ability. Adaptive parameters: The variable precision RS introduces a misclassification rate to enhance the algorithm's robustness. Efficiency balance: Computational complexity is significantly reduced and classifier performance is improved through feature reduction. By integrating the advantages of the two algorithms, this hybrid feature selection

Table 1
Types and descriptions of extracted feature.

Feature type	Specific features	Dimension	Description	Computational complexity
Time-domain features	Mean, variance, skewness, kurtosis, peak factor	12	Describe the statistical characteristics of signal amplitude	$O(N)$
Frequency-domain features	Spectral peak, center frequency, bandwidth, spectral entropy	15	Describe the frequency distribution characteristics of signals	$O(N\log N)$
Time–frequency features	Wavelet coefficient energy, scale map moment	20	Joint time–frequency analysis	$O(N^2)$
Nonlinear features	Fractal dimension, approximate entropy, Lyapunov exponent	10	Describe the nonlinear dynamic characteristics of signals	$O(N^2)$

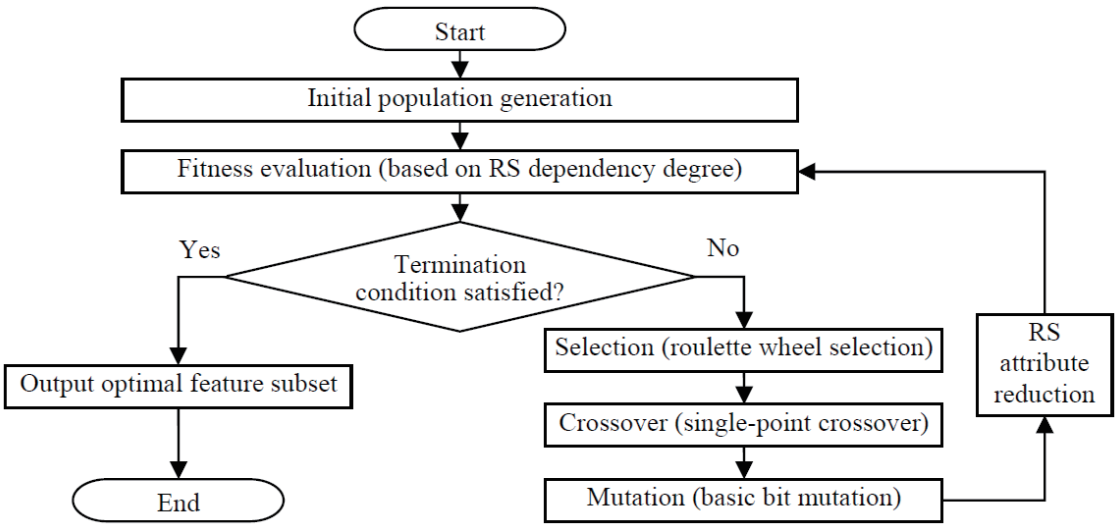

Fig. 2. GA-RS hybrid feature selection process.

method can substantially reduce feature dimensionality while maintaining classification accuracy, making it particularly suitable for high-dimensional sensor array signal processing scenarios.

Consider a decision system $S(U, A, V, f)$, where U is a nonempty finite set of objects (sensor data samples), $A = C \bigcup D$ is the attribute set, C represents condition attributes (features), and D represents the decision attribute (class label). The degree of dependence $\gamma_C(D)$ indicates the discriminative power of the condition attributes towards the decision attribute, defined as

$$\gamma_C(D) = \frac{|POS_C(D)|}{|U|}, \tag{2}$$

where $POS_C(D)$ is the positive region of U, representing the set of objects that can be unequivocally classified into equivalence classes of D on the basis of attributes C. A higher degree of dependence indicates a stronger classification capability of the feature subset.

The GA search is guided by a fitness function. We design the fitness function as

$$Fitness(S) = \alpha \cdot \gamma_C(D) + \beta \frac{|C| - |S|}{|C|}, \tag{3}$$

where S is the feature subset, $\gamma_C(D)$ is the degree of dependence of this subset on the decision attribute, $|S|$ is the size of the feature subset, $|C|$ is the total number of features, and α and β are weight coefficients. This function balances the classification capability and the size of the feature subset, avoiding the selection of too many redundant features.

2.3 LSTM-FCN classification model

We design an ALSTM-FCNclassification model, consisting of an LSTM branch, an FCN branch, and an attention mechanism, which can simultaneously capture the temporal dependences and local features of sensor signals. The model structure is shown in Fig. 3.

The "branch point" of the two branches lies in their different processing methods for input data. A unified input source is adopted—regardless of the branch, the input is the same preprocessed sensor signal tensor. The LSTM branch is focused on temporal flow. It captures the long-term dependences and dynamic change patterns of signals in the time dimension. The key operation is directly feeding the data into the LSTM layer. LSTM units process data sequentially by time steps, and their internal hidden state transfer mechanism enables them to memorize historical information. The subsequent attention mechanism assigns weights to different time steps to highlight critical time periods. The FCN branch is focused on local features. It extracts local patterns and spatial features from signals, similar to searching for discriminative "shapes" or "contours" in time series. The data usually first pass through a permute layer to adjust dimensions, allowing convolution kernels to slide effectively along a single dimension. Subsequently, a series of 1D convolutional layers capture multiscale local features through convolution kernels of different sizes.

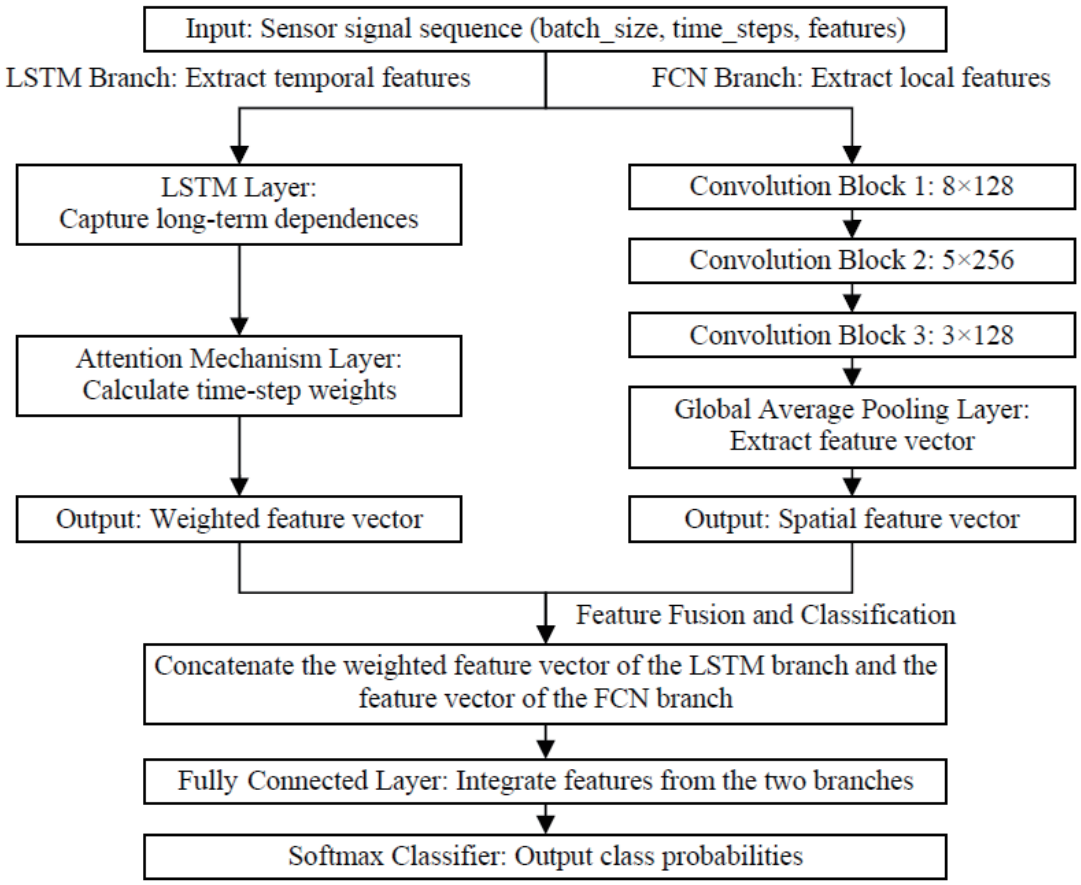

Fig. 3. ALSTM-FCN model structure.

The LSTM branch is responsible for processing the temporal dependences in the sensor signal sequence. Given an input sequence $X = (x_1, x_2, ..., x_T)$, the LSTM computation at time step t is

$$f_t = \sigma\left(W_f \cdot \left[h_{t-1}, x_t\right] + b_f\right), \tag{4}$$

$$i_t = \sigma\left(W_i \cdot \left[h_{t-1}, x_t\right] + b_i\right), \tag{5}$$

$$\tilde{C}_t = \tanh\left(W_C \cdot \left[h_{t-1}, x_t\right] + b_C\right), \tag{6}$$

$$C_t = f_t \cdot C_{t-1} + i_t \cdot \tilde{C}_t, \tag{7}$$

$$o_t = \sigma\left(W_o \cdot \left[h_{t-1}, x_t\right] + b_o\right), \tag{8}$$

$$h_t = o_t \cdot \tanh\left(C_t\right), \tag{9}$$

where f_t represents the forget gate, i_t the input gate, o_t the output gate, C_t the cell state, and h_t the hidden state.

The attention mechanism performs a weighted sum of the hidden states across all time steps of the LSTM, highlighting the contribution of key time points to the classification decision.

$$u_t = \tanh\left(W_W h_t + b_W\right), \tag{10}$$

$$a_t = \frac{\exp\left(u_t^T u_W\right)}{\sum_{t=1}^{T} \exp\left(u_t^T u_W\right)}, \tag{11}$$

$$v = \sum_{t=1}^{T} a_t h_t, \tag{12}$$

where u_t is the attention score, a_t is the normalized attention weight, and v is the weighted feature representation.

The FCN branch consists of three convolutional blocks. Each block contains a convolutional layer, a batch normalization layer, and a ReLU activation function for extracting local features of the signal. The output features from the two branches are fused and fed into a Softmax classifier to obtain the final classification result.

2.4 Summary of the mathematical model

The core mathematical model of the proposed algorithm can be summarized as the following optimization problem:

$$\min_{\theta} \frac{1}{N} \sum_{i=1}^{N} L\left(y_i, f\left(x_i;\theta\right)\right) + \lambda R\left(\theta\right), \tag{13}$$

where $L(\cdot)$ is the cross-entropy loss function, $f(x_i;\theta)$ is the predicted output of the ALSTM-FCN model, $R(\theta)$ is the regularization term, and λ is the regularization coefficient. The Adam optimization algorithm is used to solve the above problem, updating the model parameters through backpropagation. The hyperparameter settings for the ALSTM-FCN model are listed in Table 2.

3. Simulation Experiments

To verify the effectiveness of the proposed algorithm, comprehensive simulation experiments are designed to compare various traditional and modern sensor array signal classification methods.

3.1 Experimental setup and datasets

The experimental hardware environment consisted of an Intel i7-9700K CPU and an NVIDIA RTX 3080 GPU. The software environment was Python 3.8 with PyTorch 1.10. Three public sensor array datasets were used for performance evaluation: the University of California, Irvine (UCI) Gas Sensor Array dataset, containing response curves of multiple chemical gases under different concentrations and environmental conditions, used to evaluate gas recognition and concentration estimation algorithms; the Intelligent Biosensor Array dataset, containing responses of a biosensor array to various biological agents, used to study the performance of biological recognition, detection, and classification algorithms; and the Synthetic Aperture Sonar dataset, simulating target echo signals received by an underwater sensor array, used to assess algorithm performance in acoustic scenarios. Descriptions of the experimental datasets are shown in Table 3.

In this study, we select the UCI dataset because it is an authoritative public standard dataset in the sensing field with large sample size and complete annotations, suitable for verifying the generalization ability of the algorithm. The gas sensor is a MOS sensor, which can detect volatile organic gases such as formaldehyde, benzene, and ethanol, featuring fast response and high

Table 2
Hyperparameters of the ALSTM-FCN model.

Parameter category	Parameter name	Parameter value	Description
LSTMbranch	Number of hidden units	64	Control the dimensionality of temporal features
	Number of layers	2	Network depth
FCNbranch	Convolution kernel size	3, 5, 7	Multiscale receptive field
	Number of channels	128, 256, 128	Number of feature maps
Attention mechanism	Attention dimension	32	Attention weight calculation
Training parameters	Learning rate	0.001	Adam optimizer
	Batch size	64	Number of samples per training batch

Table 3
Descriptions of experimental datasets.

Dataset name	Sensor type	Number of classes	Number of samples	Feature dimensionality	Application scenario
UCI gas sensor	Metal oxide gas sensor	10	16000	128	Gas identification
Biosensor array	Biosensor	6	8400	96	Biological agent detection
Synthetic aperture sonar	Sonar sensor	8	12000	256	Underwater target classification

stability. The target variables of the biosensor array are the concentration of biological agents and specific responses; the synthetic aperture sonar sensor acquires underwater target echo amplitude, phase, and temporal signals.

3.2 Experimental results

To validate the performance of the proposed GA-RS-ALSTM-FCN algorithm, comparison algorithms were selected: the traditional signal processing algorithm MUSIC; classical machine learning algorithms SVM and RF; and deep learning algorithms CNN, LSTM, and CNN-LSTM.

The results of the classification accuracy for each algorithm on the three datasets are shown in Table 4. It can be observed that the GA-RS-ALSTM-FCN algorithm achieved the highest performance on all datasets.

The evaluation metric F1-score integrates precision and recall, ranging from 0 to 1. A value closer to 1 indicates a higher classification performance of the model. It balances two types of errors, is suitable for imbalanced sample scenarios, and comprehensively reflects the classification reliability of the model. Further analysis results of F1-score and computational complexity indicators are shown in Table 5. The proposed algorithm maintains a high F1-score while significantly reducing computational complexity through feature selection.

4. Analysis and Discussion

4.1 Impact of feature selection on algorithm performance

To verify the effectiveness of feature selection, the impact of different feature selection methods on classification accuracy is compared. The results are shown in Table 6. The proposed GA-RS hybrid feature selection method significantly reduces feature dimensionality while maintaining high classification accuracy.

4.2 Analysis of the role of attention mechanism

The attention mechanism enables the model to adaptively focus on the time segments within the sensor response sequence that are most critical for the classification decision, rather than treating all time steps equally. By introducing the attention mechanism, the model gains two

Table 4
Classification accuracies of different algorithms.

Algorithm	UCI gas sensor dataset	Biosensor array dataset	Synthetic aperture sonar dataset	Average accuracy
MUSIC	75.3	68.7	82.4	75.5
SVM	88.6	85.2	90.1	87.9
RF	91.2	89.7	92.3	91.1
CNN	93.5	90.3	94.7	92.8
LSTM	94.1	92.6	95.2	94.0
CNN-LSTM	95.8	94.1	96.5	95.5
Proposed algorithm	**98.3**	**96.7**	**98.9**	**98.0**

Table 5
F1-scores and computational complexities of different algorithms.

Algorithm	F1-score	Training time(s)	Inference time (ms)	Number of parameters (M)
MUSIC	0.72	—	12.5	—
SVM	0.86	45.3	5.2	—
RF	0.90	62.7	8.7	—
CNN	0.91	128.5	3.6	2.34
LSTM	0.93	156.8	4.9	3.12
CNN-LSTM	0.94	203.6	6.3	5.78
Proposed algorithm	**0.97**	**185.4**	**5.1**	**4.25**

Table 6
Performance characteristics of feature selection methods.

Feature selection method	Number of selected features	Classification accuracy (%)	Feature reduction rate (%)
No feature selection	128	95.8	0
Principal component analysis (PCA)	45	96.2	64.8
Random forest feature importance	38	96.5	70.3
RS attribute reduction	42	96.8	67.2
GA-RS (proposed)	**35**	**98.0**	**72.7**

main advantages: enhanced feature discriminability by focusing on the most distinctive signal segments, directly improving classification accuracy; improved model interpretability, where, by visualizing the distribution of attention weights, researchers can intuitively understand which parts of the signal the model bases its decisions on, which helps verify whether the model's decisions align with prior knowledge of physical or chemical processes. The visualization analysis of attention weights in the simulation experiments confirmed that the model indeed assigns high weights to key phases of the sensor response curve, such as transient changes and stable response periods, significantly improving the model's classification accuracy and robustness in complex signal environments.

4.3 Balancing model complexity and efficiency

High-accuracy models often imply a large number of parameters and computational overhead, limiting their deployment on resource-constrained edge devices or real-time systems.

The aim of the proposed GA-RS-ALSTM-FCN algorithm is the systematic optimization of this balance through a series of coordinated designs.

At the model input, the GA-RS hybrid feature selection module acts as a "prefilter," screening the most discriminative feature subset from the original high-dimensional features, reducing the feature dimensionality by approximately 72.7% on average. This not only directly reduces redundant data for the subsequent model input, lowering the computational load, but also improves the stability and convergence speed of model training by removing irrelevant and noisy features, indirectly enhancing computational efficiency.

In terms of model architecture, expressive efficiency was considered in the ALSTM-FCN design itself. Compared with a single deep CNN or LSTM, the dual-branch structure achieves functional division: the FCN branch efficiently extracts local spatial features through weight-sharing convolutions, while the LSTM branch focuses on modeling critical long-term temporal dependences. The introduced attention mechanism enhances efficiency from another dimension; it allows the model to avoid relying on deeper stacking or wider layers to enhance expressive power. Instead, it achieves higher accuracy gains through dynamic weighting focused on the most information-rich temporal segments, by implementing a more "intelligent" allocation of computation.

Although the proposed algorithm significantly leads in classification accuracy, the training and inference times do not increase proportionally, and the number of parameters remains within a reasonable range. Compared with the baseline CNN-LSTM model, the average accuracy of the proposed algorithm improves by 2.5 percentage points (from 95.5 to 98.0%) while the number of parameters is reduced by approximately 26.5%. This demonstrates that through carefully designed front-end feature engineering and efficient model architecture synergy, it is entirely possible to achieve a balance between model performance and computational cost that is highly favorable for practical applications, laying the foundation for the algorithm's deployment in edge computing environments. In the future, three lightweight technologies, model quantization, channel pruning, and knowledge distillation, will be adopted to greatly compress the number of parameters and computational complexity with controllable accuracy loss. Moreover, the feature selection process will be optimized to further reduce the front-end computational overhead, enabling the algorithm to achieve low-power and real-time inference on embedded edge sensor nodes to meet the practical deployment requirements of edge computing.In addition, the value of 95.5% here is the average classification accuracy of the benchmark CNN-LSTM algorithm on the three public datasets, calculated as the arithmetic mean of the measured results on each dataset. In the scenario of sensor array signal classification, this improvement is significant within the high-precision range, which can effectively reduce the misjudgment rate and enhance system reliability, with clear practical value for industrial monitoring, underwater detection, and other practical applications.

4.4 Limitations analysis

First, the model's performance highly depends on high-quality labeled data. As a data-driven deep learning method, this algorithm requires a large, balanced, and accurately labeled sample

set for training to fully utilize its advantages. However, in many practical industrial or scientific application scenarios, acquiring sufficient labeled data is often costly, time-consuming, and may even face safety and ethical restrictions. This may limit the model's generalization ability in data-scarce domains.

Second, there is a trade-off between the complexity of the model architecture and its interpretability. Although the attention mechanism is introduced to enhance interpretability to some extent, the overall model remains a "black box" (which means that the internal operation logic and decision-making process of deep learning models are difficult to understand intuitively and cannot be clearly traced by explicit physical or mathematical principles, with only inputs and outputs visible externally), and hence, its internal decision logic is difficult to fully trace and explain using clear physical or chemical principles. In safety-critical fields or domains requiring clear causal relationships, this lack of interpretability may affect user trust and the adoption of the model.

Furthermore, the algorithm's generalization ability faces the challenge of "domain shift." The model performs well under the training data distribution, but when significant changes occur in the sensor array type, working environment, or concentration range of the target substance, the performance may degrade. This often requires additional domain adaptation or recalibration in practical deployment, increasing the cost of use.

Finally, the overall computational cost of the model remains significantly higher than that of traditional signal processing methods. Achieving real-time processing on extremely resource-constrained micro-embedded sensor nodes still requires further optimization work such as model lightweighting, pruning, or knowledge distillation.

To address the above limitations, future improvements will be carried out in three aspects. First, metalearning, data augmentation, and transfer learning will be introduced to construct a few-shot learning framework, thus reducing the dependence on large-scale annotated data and improving the generalization ability in scarce sample scenarios. Second, attention visualization and signal mechanism analysis will be combined to associate model decisions with physicochemical properties, thereby enhancing interpretability to meet the requirements of high-safety fields. Third, lightweight technologies such as model pruning, quantization, and knowledge distillation will be adopted, together with the collaborative design of dedicated hardware, to further reduce computational overhead, realize the low-power real-time deployment of the algorithm on micro-embedded sensor nodes, and mitigate performance degradation caused by domain shift via domain adaptation methods.

5. Conclusions

In this study, we proposed a sensor array signal classification algorithm based on big data mining fusion. By combining traditional signal processing techniques with modern deep learning models the, high-precision classification of sensor array signals was achieved. Although the proposed algorithm is currently verified on public datasets, all the datasets selected in this study are derived from actual signals collected by real sensor arrays in industrial, biological, and underwater acoustic scenarios, with strong practical representativeness. Through feature

dimensionality reduction and lightweight design, the algorithm was fully adapted to the resource constraints of edge sensor nodes. In the future, hardware deployment and practical testing in real environments will be carried out, relying on the industrial monitoring and underwater detection platforms of cooperative institutions, transforming theoretical achievements into implementable engineering applications to ensure the practical deployment and utility value of the algorithm.

The main contributions include a GA-RS hybrid feature selection method designed to effectively reduce feature dimensionality and improve classification performance; an ALSTM-FCN classification model that enhances the ability to capture key signal features through an attention mechanism; and the validation of the algorithm's superior performance on multiple datasets through extensive experiments. Future research directions include exploring the application of few-shot learning techniques in scenarios with scarce labeled data, specifically through methods like metalearning, data augmentation, and transfer learning, enabling the model to adapt quickly with only a small number of target domain samples; enhancing model interpretability to provide physically meaningful explanations for classification decisions, mapping model decisions back to signal time–frequency characteristics or sensor response mechanisms, establishing a trustworthy human–machine collaborative analysis paradigm; and optimizing the model's edge deployment capability to improve its practicality on resource-constrained platforms, coordinating lightweight techniques with dedicated hardware design to achieve the low-power and real-time inference of the algorithm on embedded sensor nodes.

References

1 B. Sumanto, K. Triyana, and A. Kusumaatmaja: Instrum. Sci. Technol. **54** (2026) 118. https://doi.org/10.1080/10739149.2025.2504478
2 A. Tangirbergen, A. Tleubekova, G. Yergaliuly, A. Kurmanbayeva, B. Soltabayev, and A. Soltabayeva: LWT. Opt. **23** (2026) 1. https://doi.org/10.1016/J.LWT.2025.118926
3 O. Tayeng, P. Behera, and M. De: Nanoscale. **35** (2025) 1. https://doi.org/10.1039/D5NR02108A
4 N. Otrooshi, R. Morgenstern, B. Oripov, A. McCaughan, N. Nader, M. J. Collins, R. Mirin, A. E. Lita, and S. W. Nam: Opt. Express. **33** (2025) 51108. https://doi.org/10.1364/OE.567354
5 X. Zeng, D. Deng, H. Yang, Z. Yang, S. Ma, L. Yang, and Z. Wu: Aca. Aeronaut. Astronaut. Sinica **45** (2024) 100. https://doi.org/10.7527/S1000-6893.2024.30368
6 L. Tang, B. Han, Y. Huang, and Q. Zhou: Digit. Ocean Underw. Warfare **8** (2025) 613. https://doi.org/10.19838/j.issn.2096-5753.2025.05.011
7 G. E. Unzueta, and S. E. Romero: Fiber Integr. Opt. **44** (2025) 214. https://doi.org/10.18494/SAM.2012.764
8 A. S. Dyah, M. H. Tamimi, A. A. S. Pradhana, K. A. Alamsyah, H. Purnobasuki, M. Khasanah, Y. Susilo, K. Triyana, M. Kashif, and A. Syahrom: Biosens. Bioelectron. **9** (2021) 1. https://doi.org/10.1016/J.BIOSX.2021.100083
9 R. Laref, E. Losson, A. Sava, and M. Siadat: Chemom. Intell. Lab. Syst. **184** (2019) 22. https://doi.org/10.1016/j.chemolab.2018.11.011
10 J. Wang, X. Liu, and W. Ye: J. Beijing Norm. Univ. (Nat. Sci.). **53** (2017) 159. https://doi.org/10.16360/j.cnki.jbnuns.2017.02.008
11 N. Gupta, M. Ravisankar, Surjeet, and S. K. Dargar: Wirel. Netw. **30** (2025) 1. https://doi.org/10.1007/S11276-025-03977-5
12 Y. Y. Liu, R. Yuan, S. Y. Shao, and M. Chen: J. Mod. Def. Technol. **53** (2025) 215. https://doi.org/10.3969/j.issn.1009-086x.2025.05.022
13 J. Eisele, A. Gerlach, Y. Manz, M. Maeder, and S. Marburg: J. Acoust. Soc. Am. **157** (2025) 2556. https://doi.org/10.1121/10.0036385
14 S. Y. Heng, K. Z. Yap, W. Y. Lim, and N. Ramakrishnan: Sens. Imaging. **25** (2024) 51. https://doi.org/10.1007/S11220-024-00501-5
15 T. J. Tello, V. A. Guaman, and B. S. Ko: IEEE Sens. J. PP (2020) 1. https://doi.org/10.1109/jsen.2020.3007431

16 K. Ravi and S. Hardeep: Microw. Opt. Technol. Lett. **65** (2022) 859. https://doi.org/10.1002/MOP.33558
17 T. Li, Y. Zhang, H. Sun, Z. Zhang, C. Zhang, J. Sun, and H. Wang: ACS Sens. **10** (2025) 8809. https://doi.org/10.1021/ACSSENSORS.5C02740

About the Authors

Jiana Bi received her Ph.D. degree from Harbin Institute of Technology, China, in 2009. From 2010 to 2021, she was a professor at Bohai University, China. Since 2021, she has been a professor at Changzhou College of Information Technology, China. Her research interests are in network security, big data technology, and artificial intelligence.(bijiana@czcit.edu.cn)

ChunCheng Wei received his M.S. degree from Cheng Kung University, Taiwan, China, in 1996. From 1999 to the present, he has been a lecturer at both Tajen University and Changzhou College of Information Technology. His research interests are in information security, cloud computing, hardware design, and artificial intelligence.(ccwei@czcit.edu.cn)

Ning He received her Ph.D. degree from Waseda University, Japan, in 2013. She is currently a faculty member at the School of Software and Big Data, Changzhou College of Information Technology, China. She has led multiple research projects including the Natural Science Foundation of Jiangsu Higher Education Institutions and the Longcheng Talent Program in Changzhou. Her research interests are in big data technology and artificial intelligence. (hening@czcit.edu.cn)