Hydrometeor Classification for Dual Polarization Radar Based on Multi-Sample Fusion SVM

Zehu LUO, Xudong WANG, Ling WANG, Guiguang XU, Yongxing GAO

Dept. of the Key Laboratory of Radar Imaging and Microwave Photonics of the Ministry of Education, Nanjing University of Aeronautics and Astronautics, 29 Yudao Street, Nanjing, China

{luozehu1264, xudong, tulip_wling, xuguiguang}@nuaa.edu.cn, gyx_nuaa@163.com

Submitted September 1, 2022 / Accepted December 23, 2022 / Online first March 6, 2023

Abstract. *In order to enhance the accuracy of dual polarization radar in hydrometeor classification, a hydrometeor classification algorithm based on multi-sample fusion Support Vector Machine (SVM) is proposed in this paper after considering that traditional fuzzy logic algorithm has the defect of over relying on expert experience to set parameters. The data of four polarization parameters (horizontal reflectivity factor, differential reflectivity, correlation coefficient and differential propagation phase constant) detected by the KOHX radar were taken as the feature information of hydrometeors. The dataset was collected, and the model was trained. According to the classification results of SVM model and combined with the distribution characteristics of target particles in the rainfall area, a classification system that can effectively identify four types of particles (dry snow, moderate rain, big drops and hail possibly with rain) was established. This model greatly reduced the misidentification of dry snow (DS) and moderate rain (RA)) in the precipitation area, and significantly improved the overall classification effect of hydrometeors in the area. The 0.5° elevation scanning data of the radar at a certain time were tested, and the classification accuracy of system model was up to 97.21%. The average accuracy of other elevation scanning data was approximately 97%, which showed strong robustness.*

Keywords

Dual polarization radar, fuzzy logic, feature dimension, hydrometeors classification, Support Vector Machine (SVM)

1. Introduction

Traditional Doppler weather radar sends electromagnetic waves in a single direction. It only obtains such parameters as reflectivity factor, Doppler velocity and spectral width. However, dual polarization radar strengthens the function of detection by emitting electromagnetic waves in the horizontal and vertical directions. This makes it impossible to obtain more information inside the cloud cluster, including differential reflectivity, differential propagation phase constant and correlation coefficient. These parameters help to obtain the shape, size, orientation and other important information of hydrometeors and greatly improve the capability of weather radar in hydrometeor classification [1–3].

Fuzzy logic theory was an important means to solve classification problems and achieved remarkable results in the field of hydrometeor classification. Many researchers concentrated on the optimization of the membership function and robustness of the algorithm. However, one of the drawbacks of this algorithm was that the parameters of membership function and the weight of polarization parameters depended on the experience of experts. Meanwhile, the shape of membership function was predetermined, such as beta function [4], asymmetric T function [5] and Gaussian function [6]. This would exert some influence on the classification result.

It can be seen from the process of the fuzzy logic algorithm that the selection of its membership function and the determination of its parameters require expert experience to determine manually, which significantly affects the algorithm's objectivity and portability. Therefore many scholars have researched hydrometeor classification methods based on data-driven models. Such models can learn from a large number of data to obtain the best parameters of the model without manual determination. In 2015, Grazioli et al. created an unsupervised clustering algorithm and partitioned the temperature. The clustering results were obtained based on the similarities of polarization data in each region, which effectively avoided the subjective choice of hydrometeor types [7]. In 2020, Han et al. proposed a deep learning method using Convolution Neural Network (CNN) for prediction. A cross channel three-dimensional convolution in the first layer of CNN was used to fuse the original data, and it enhanced the ability of the detection and real-time prediction of convective storms [8]. In 2017, Wang et al. proposed a fuzzy clustering recognition algorithm based on deep learning. CNN was used for initial clustering to eliminate the influence caused by measurement

DOI: 10.13164/re.2023.0151

errors of polarization parameters [9]. In 2019, Lu et al. proposed a multi-classification model based on residual convolutional neural network (ResNet). The polarization parameters were superimposed through data channels and processed in blocks to produce the dataset of the algorithm and improve the classification accuracy of some hydrometeor [10].

As shown in references [7–10], it is proposed to use the neural network model in the depth learning algorithm to classify hydrometeors, which solves the problem that the model parameters need to be determined by human factors. However, the neural network model has a relatively high computing cost for hardware and computing time. Therefore, based on the characteristics of PPI scanning data of dual-polarization weather radar and the distribution of hydrometeors in the precipitation area, a hydrometeor classification method based on multi-sample fusion SVM is proposed in this paper. This model is also a data-driven model. It not only overcomes the objectivity and portability problems of fuzzy logic algorithm, but also has much less computation compared with neural network model, so it has a specific application value. It is worth noting that this method solves the problem of high false recognition rate of DS and RA particles in the area. It dramatically improves the overall classification accuracy of target hydrometeors.

The main structure of this paper is as follows: Section 2 mainly describes the principle of SVM classification and how to use dual-polarization radar data to construct an SVM multi-classifier to classify hydrometeors. Section 3 testes the specific training performance of the SVM model on the data, including three steps. The first step discusses the single-sample SVM model's high false recognition rate of DS and RA particles. In the second step, aiming at the problems in the first step, a multi-sample fusion method is proposed to improve the dimension of hydrometeor feature information to solve the problem of false recognition. In the third step, combined with the classification of the first two steps and the distribution characteristics of hydrometeors in the precipitation area, a system model based on multi-sample fusion SVM is establised. Section 4 verifies the system model built in Sec. 3 by using the measured data of different elevations collected by KOHX radar at 00:49 on February 24, 2019. Section 5 is the overall analysis and comprehensive evaluation of the model built in this article and included the future work plans.

2. SVM Classification Model

SVM model can realize the binary classification of hyperplane classification. Binary classification problem is one of the main forms of pattern recognition. Taking the multidimensional data of the model as the location variables of the hyperplane, the samples in the model can be transformed into several points in the hyperplane. The optimal division is achieved by finding an optimal hyperplane [11–13].

$(\mathbf{x}_i, y_i)$ is a linear sample that needs to be dichotomized. Specifically, $i = 1, 2, \cdots, n$ are the different dimensions of the feature information. $\mathbf{x}_i$ is a feature vector, belonging to the real number field in n-dimensional space. y_i is the classification mark, and its value can be +1 or −1. In the classification of hydrometeors, feature vector $\mathbf{x}_i$ refers to different polarization parameters. Their values are different from each other in dimension and order of magnitude. If they are directly used for classification, the data of some dimensions will amplify or reduce their influence on classification results. This will result in a non-objective classification result. Therefore, the original data should be initially normalized, as shown in (1):

$$\hat{\mathbf{x}}_i = \frac{\mathbf{x}_i - \min\{\mathbf{x}_i\}}{\max\{\mathbf{x}_i\} - \min\{\mathbf{x}_i\}}. \tag{1}$$

The initialized feature dimension data is converted into a value between 0 and 1. $\hat{\mathbf{x}}_i$ is the normalized result of polarization parameters of a hydrometeor samples. $\max\{\mathbf{x}_i\}$ is the maximum value of all polarization parameter values. $\min\{\mathbf{x}_i\}$ is the minimum value of all polarization parameter values.

The hyperplane $y(\mathbf{x}) = \mathbf{w}\cdot\mathbf{x}+b$ is defined. If the sample $(\mathbf{x}_i, y_i)$ that makes $y(\mathbf{x}) = \mathbf{w}\cdot\mathbf{x}+b > 0$, then the sample is classified as +1. Otherwise, it is classified as −1. The label value ±1 of y_i is guaranteed to be output by scaling the values of parameters $\mathbf{w}$ and b . The classification interval is $2/\|\mathbf{w}\|$. In order to make this hyperplane more robust, a maximum interval hyperplane needs to be defined [14], [15]. The minimum value $d_{\min}$ is obtained by satisfying $\|\mathbf{w}\|$. The minimum classification interval is as follows:

$$d_{\min} = \min\left(\|\mathbf{w}\|/2\right) \tag{2}$$

where the hyperplane with linear separability between samples should meet the following condition:

$$y_i \cdot (\mathbf{w}\cdot\mathbf{x}_i + b) > 0, \quad i = 1, 2, \cdots, n. \tag{3}$$

The model in (2) can only process ideal linearly separable data, and the actual hydrometeor samples are not completely linearly separable data. Therefore, slack variable ξ and penalty coefficients C need to be introduced to process these special samples. The overall classification robustness of the model can be improved by weakening model constraints [16]. As shown in Fig. 1, it is a two-dimensional linear non-separable SVM model for slack variables. The multidimensional SVM model is obtained through (4):

$$d_{\min} = \min\left(\|\mathbf{w}/2\| + C\sum_i \xi_i\right), \quad C > 0 \tag{4}$$

where the hyperplane with slack variable should meet the following condition:

$$y_i \cdot (\mathbf{w}\cdot\mathbf{x}_i + b) > 1 - \xi_i, \quad \xi_i \geq 0. \tag{5}$$

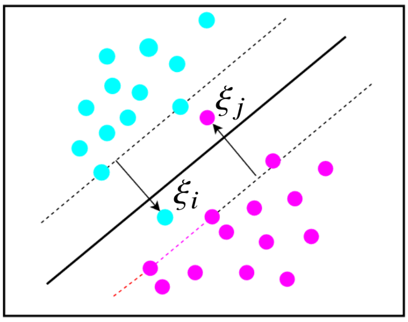


Fig. 1. Two-dimension SVM model with slack variables.

In this paper, the classification of hydrometeors is a multi-classification model. Compared with binary classification SVM, we need to train every SVM model to build a multi-classification model pool. If an unknown sample needs to be classified, we will count the number of votes obtained by each label in the output result from the pool. The label corresponding to the highest number of votes is the category of the input sample.

3. Training of Models

In this paper, there are two steps to build the system model. The first step is the single-sample step. The SVM model is optimized and saved in the dataset composed of four-dimensional polarization parameters. The second step is the multi-sample fusion step. Due to the existence of the melting layer, the distribution of some particles in the actual precipitation region has obvious boundaries. The one-dimensional fusion method of multiple samples can be used to improve the dimension of feature information. The best multi-sample fusion SVM model can be obtained in the second step. Then the final system model is established by combining with the classification model in the first step.

The hardware platform of this model training is the CPU Intel Core i7-10700K. The model uses the SVC structure in the Sklearn (Scikit-Learn) framework to carry out the essential development of the binary classifiers of SVM. Sklearn is a machine learning tool based on the Python language. We use the One Versus One (OVO) method to establish the multi-classification model of SVM. The data and source code are available on github: https://github.com/lzehu/Hydrometeor-classification

3.1 Training of Single-Sample SVM

We firstly collected the 0.5° elevation scanning data of the KOHX radar in 2018 and then selected four types of target particle samples including horizontal reflectivity, differential reflectivity, correlation coefficient and differential propagation phase constant. The label values of DS, RA, BD and HA were $0, 1, 2, 3$ respectively. These label values were taken as the fifth dimension of the vector. In this way, the dataset for the single-sample SVM model training was prepared. As shown in Fig. 2, the training models with different sizes of train sets were used to establish the relationship between the accuracy of validation and the size of train set. The minimum size for each binary classification model was determined when the accuracy tended to be stable. Within the first 2,000 samples, the accuracy fluctuated and rose when sample size was increased. When the sample size was greater than 2,000 or so, it gradually became stable. Finally, the minimum training data size of single-sample SVM model was determined to be 2,000 samples of every type of hydrometeors. In the experiment, we adjusted the optimal parameters of the model by the grid search method.

As shown in Tab. 1, they were the optimal model parameters of each binary classifier, and the column model name such as DS_RA was a binary single-sample SVM model. Among them, the Kernel function was selected as nonlinear radial basis function (RBF), Gamma was mainly used to determine the mapping dimension of the Kernel function, and C was used to determine the penalty coefficient of error item [17], [18].

SVM is a binary classifier. There are six optimal binary classification SVM models established through the optimal parameters in Tab. 1. When dealing with multiple classification problems, it is necessary to construct appropriate multiple classifiers. In this paper, OVO is used to build a multi-classification training pool. For a test sample, the output labels of each binary model are counted. The label with the highest number of votes is the category of the hydrometeor. Based on a such voting mechanism, this OVO-SVM multi-classification model is established.

Model name	Kernel	C	Gamma	Accuracy
DS_RA	RBF	910	0.7	0.7031
DS_BD	RBF	150	0.1	0.9552
DS_HA	RBF	95	0.4	0.9933
RA_BD	RBF	90	0.3	0.9625
RA_HA	RBF	120	0.4	0.9999
BD_HA	RBF	105	0.2	0.9980

Tab. 1. The best parameters of every single-sample SVM.

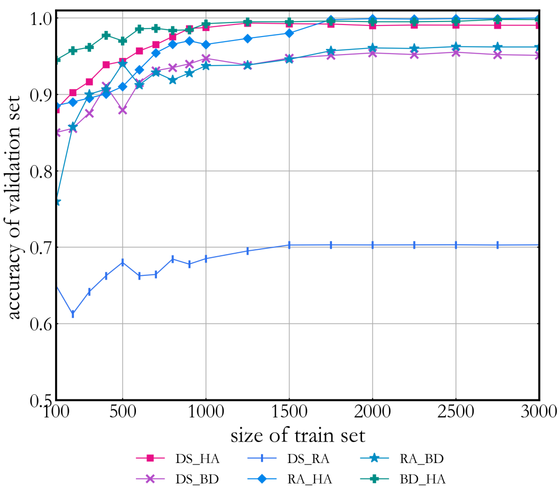


Fig. 2. Variation curve of accuracy with training amount on validation set in the first step.

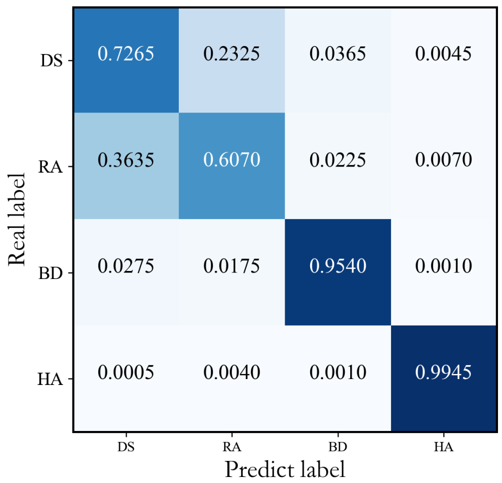


Fig. 3. Confusion matrix of classification results on the test set in the first step.

As shown in Fig. 3, it is the confusion matrix of test set. For the single-sample SVM model, the recognition accuracy of HA and BD is above 95%. The mutual misjudgment between RA and DS reduces the overall recognition rate. Therefore, the classification of DS and RA based on the SVM model needs to be further processed.

3.2 Training of Multi-Sample SVM

As only four-dimensional feature information is used to obtain classification results, DS and RA have the problem of mutual misidentification. Considering the characteristics of PPI scanning data of dual-polarization radar and the distribution of hydrometeors in the precipitation area, a hydrometeor classification method based on multi-sample fusion SVM is proposed, which improves the classification ability of DS and RA.

The method of integrating multiple samples to expand the dimension of feature information was proposed to enhance the discriminability of different categories. Its essence was to use the mapping principle of the SVM kernel function, which mapped linearly inseparable low-dimensional feature vectors to high-dimensional features. Then, they were linearly separable and trained in SVM [19], [20]. The Lagrangian multiplier method was used to transform the solution process of the maximum classification interval with constraints into a function without constraints to find the extremum [21]:

$$\max_{\alpha}\left(L(\mathbf{w},b,\alpha)\right) = \frac{1}{2}\|\mathbf{w}\|^2+\sum_{i=1}^{n}\alpha_i\left(1-y_i\left(\mathbf{w}\cdot\Phi(\mathbf{x})+b\right)\right) \quad (6)$$

where the Lagrange multiplier was $\alpha_i > 0$. We took the derivatives of $\mathbf{w}$ and b respectively to obtain the extreme value of L, as shown in (7):

$$\begin{aligned}\frac{\partial L}{\partial \mathbf{w}} &= 0 \Rightarrow \mathbf{w}=\sum_{i=1}^{n}\alpha_i y_i \Phi(\mathbf{x}_i),\\ \frac{\partial L}{\partial b} &= 0 \Rightarrow 0=\sum_{i=1}^{n}\alpha_i y_i.\end{aligned} \quad (7)$$

Then, we brought the two conditions obtained into (6) and obtained its minimum value by the derivation of α by L in (8):

$$\min_{\alpha}\left(L(\mathbf{w},b,\alpha)\right) = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\alpha_j y_i y_j \Phi^{\mathrm{T}}(\mathbf{x}_i)\,\Phi(\mathbf{x}_j)-\sum_{i=1}^{n}\alpha_i. \quad (8)$$

As shown in (9), we obtained the $\mathbf{w}$ and b of the hyperplane satisfying the condition:

$$\begin{aligned}\mathbf{w} &= \sum_{i=1}^{n}\alpha_i y_i \Phi(\mathbf{x}_i),\\ b &= y_i-\sum_{i=1}^{n}\alpha_i y_i \Phi^{\mathrm{T}}(\mathbf{x}_i)\,\Phi(\mathbf{x}_j).\end{aligned} \quad (9)$$

Theoretically, a suitable mapping can be found for any hydrometeor samples, and these samples that cannot be divided in the low dimensional space can be divided linearly after being fused into the high dimensional space. As shown in (8), $\Phi^{\mathrm{T}}(\mathbf{x}_i)\,\Phi(\mathbf{x}_j)$ is the two inner products mapped to high dimensions that need to be solved. If there is such a function $K(\mathbf{x}_i,\mathbf{x}_j)=\Phi^{\mathrm{T}}(\mathbf{x}_i)\,\Phi(\mathbf{x}_j)=\langle\Phi(\mathbf{x}_i),\Phi(\mathbf{x}_j)\rangle$, the inner products of $\mathbf{x}_i$ and $\mathbf{x}_j$ in the corresponding high-dimensional space are equal to the results of their calculation in the original space through the K function. Such K function is a corresponding kernel function. The commonly used kernel functions include linear kernel, Gaussian kernel, polynomial kernel, etc.

In order to get the influence of sample fusion number on classification accuracy, we have only made data sets in the number of 2–15, the fusion number of the multi-sample fusion SVM model and corresponding accuracy are shown in Fig. 4. When the fusion number starts from 10, the accuracy of the test set is basically maintained at about 91%. In order to ensure that the classification accuracy met the requirements, other noise particles were mixed into the fusion samples as few as possible. In this paper, 10 samples in the same type were randomly fused in the azimuth direction. Compared with two-dimensional fusion performed in range direction and azimuth direction, this fusion method could effectively avoid such a defect that the fusion samples were mixed with other one-dimensional feature information. This made the fusion samples purer. As shown in Fig. 5, the PPI scanning mode of dual polarization radar and the comparison of the two fusion methods in this scanning mode of the radar.

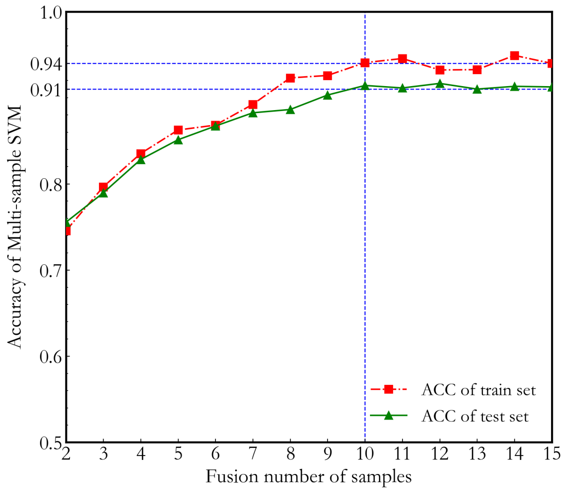


Fig. 4. Corresponding relationship between sample fusion number and model accuracy.

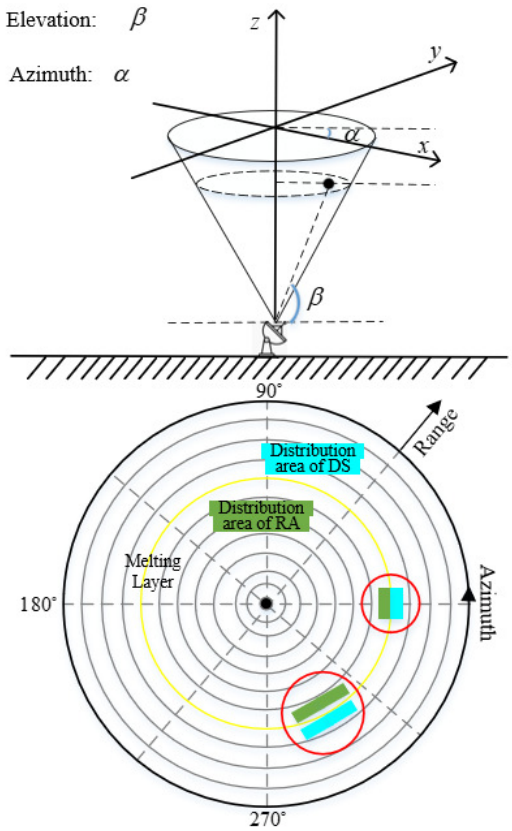


Fig. 5. PPI scanning mode of dual polarization radar and the different fusion methods in this scanning mode.

After fusing, the dimension of feature vector was increased from the original four dimensions to forty dimensions. The label value of every type of sample was still added to the last dimension of the vector. The dataset was built for the multi-sample fusion SVM model in this way. As shown in Fig. 6, the relationship between size of train set and accuracy of validation set was established. It could be concluded that the classification accuracy of hydrometeors was improved by the multi-sample SVM model.

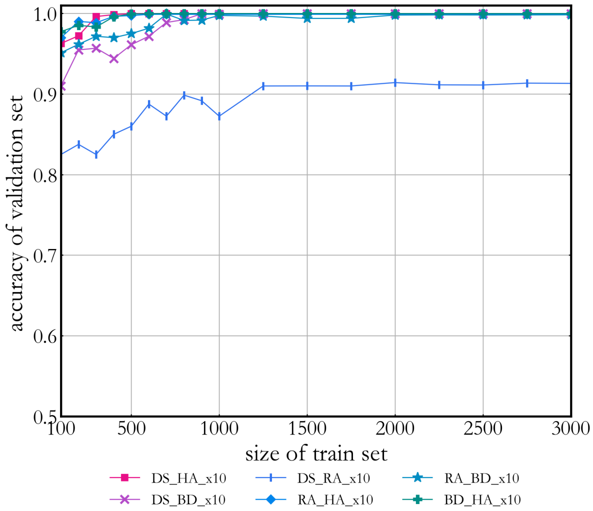


Fig. 6. Variation curve of accuracy with training amount on validation set in the second step.

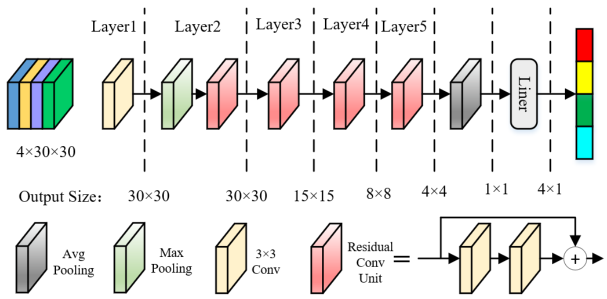


Fig. 7. The structure of ResNet18 in [10].

When the size of train set reached about 1,250, the accuracy of each SVM model tended to be stable. Therefore, the minimum size of train set was recommended to be 1,250. In order to conduct horizontal comparison experiment with the classification effect of the single-sample SVM model, the same size was adopted in this step like the first step. That was, 2,000 samples per category.

This reflects the advantages of SVM classification model in training with a small size of dataset. According to the algorithm complexity calculation equation $T_{\mathrm{SVM}} = O\left(N^2 m^2\right)$, N is sample number, and M is the dimension of feature vector. It can be concluded that the algorithm complexity with one sample of the multi-sample fusion SVM model is $m^2 = 10^2$. Convolutional neural network algorithms are based on data-driven models such as [9] and [10]. However, its complexity is determined by the complexity of each convolution layer. It can be calculated by $T_{\mathrm{CONV}} = O\left(M^2 K^2 C_{\mathrm{in}} C_{\mathrm{out}}\right)$. M is the size of the input feature graph, K is the size of the convolution kernel, C_{in} is the number of input channels, and C_{out} is the number of output channels. Taking ResNet18 network structure in [10] as an example, its network structure is shown in Fig. 7, and the algorithm complexity of a sample can be calculated. Its order of magnitude is 10^7.

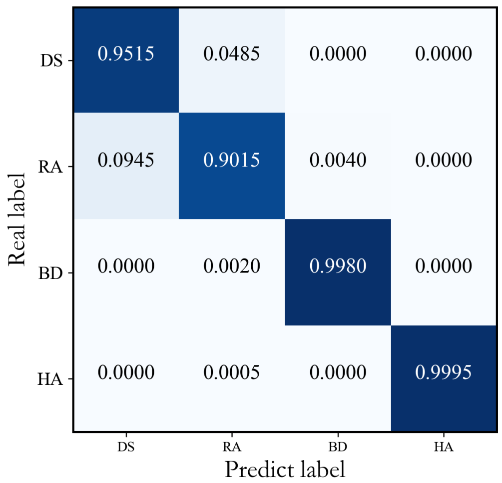


Fig. 8. Confusion matrix of classification results on the test set in the second step.

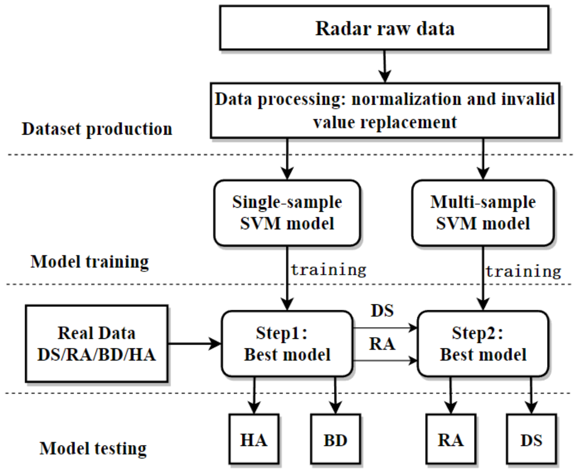


Fig. 9. The structure of four kinds of target hydrometeor classification system model.

Model name	Kernel	C	Gamma	Accuracy
DS_RA_x10	POLY	500	3	0.9142
DS_BD_x10	POLY	100	5	0.9996
DS_HA_x10	POLY	300	4	0.9999
RA_BD_x10	POLY	100	2	0.9993
RA_HA_x10	POLY	100	6	0.9989
BD_HA_x10	POLY	100	5	0.9999

Tab. 2. The best parameters of every multi-sample SVM.

The optimal parameter selection problem was studied through the grid search method. As shown in Tab. 2, they were the optimal parameters of each binary classifier, and the column model name such as DS_RA_x10 was a binary multi-sample fusion SVM model. The kernel function was Polynomial kernel (POLY).

The model pool in the second step was established in the same way as the first step. The classification result of test data was obtained through the voting mechanism. As shown in Fig. 8 it was the confusion matrix of test set.

3.3 Construction of System Model

Considering the distribution characteristics of particles in the precipitation region, the multi-classification SVM model was adopted.

As shown in Fig. 9, a classification system model of four types of target particles was established. Firstly, the initial prediction of the four types of target particles came from single-sample SVM. The BD and HA samples in the prediction results were directly outputted. The RA and DS samples with high misidentification rate were further predicted by the multi-sample fusion SVM model after feature information fusion processing. Finally, the classification results of the two steps were integrated to obtain the final classification of all samples.

4. Real Data Verification

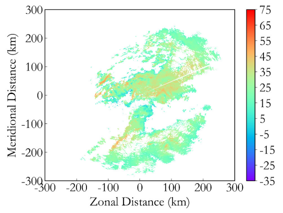
As the mainstream algorithm of hydrometeor recognition, fuzzy logic hydrometer classification (FHC) had the advantages of easy implementation and high efficiency [22], [23]. The analytical solution of the problem was not discussed in this paper. In order to compare the classification accuracy of the FHC model with the T-type membership function, the single sample SVM model and the system model were established in this paper. We verified the 0.5° elevation scanning data of the KOHX radar at 00:49 on February 24, 2019. We also obtained the classification accuracy of the target particles in the elevation scanning data by using the methods in [9] and [10]. In addition, we collected and processed the radar scanning data of 0.9°, 1.5°, 1.8°, 2.4° elevation and then verified each model in turn.

4.1 Result of 0.5° Elevation

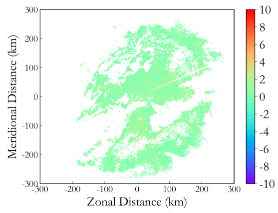
In the 0.5° elevation PPI scanning data from the KOHX radar at 00:49 on February 24, 2019, the visualization results of the four polarization parameters were shown in Fig. 10.

The data of the KOHX radar at a time contained $360 \times 1,200$ samples. At that time, we collected 37,243 DS samples, 126,307 RA samples, 20,261 BD samples and 544 HA samples. Figure 11(a)–(d) showed the NOAA reference results and the classification results of each model. According to the results in the black box in Fig. 11(b), FHC model mistakenly recognized DS particles as HA particles, which affected the classification accuracy. The overall accuracy was only 85.47%. As shown in the black box in Fig. 11(c), the single-sample SVM model was poor in distinguishing DS and RA. According to the PPI scanning characteristics of the KOHX radar, the DS particles only appeared above the melting layer. There was an obvious distribution boundary between them. However, many RA particles in the central area of Fig. 11(c) were mistakenly classified as DS particles, and RA particles that should not appear were seen in the periphery. The accuracy of single-sample SVM model was only 72.22%. 21.25% of DS samples were wrongly identified as RA samples, and 35.47% of RA samples were wrongly identified as DS samples. This dramatically affected the overall recognition effect of hydrometeors in the precipitation area. Through the multi-sample fusion of DS and RA samples, the feature information dimension was improved and the distinguishability was enhanced. The false recognition rates decreased to 5.27% and 10.12% respectively. As shown in Fig. 11(d), the identification of DS and RA particles was significantly improved. The accuracy of the system model was greatly improved, and it was 97.21%.

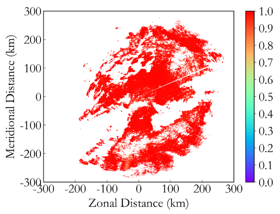
The algorithms proposed in [9] and [10] are also important research on the classification of hydrometeors. Table 3 shows the classification accuracy of the models in each reference for 0.5° elevation scanning data of the KOHX radar. It indicates that the system model has the highest accuracy for target particles.



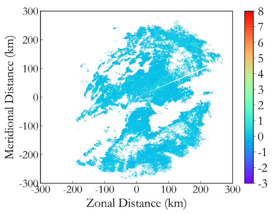
(a) Horizontal reflectivity



(b) Differential reflectivity



(c) Correlation coefficient



(d) Differential propagation phase constant

Fig. 10. Polarization parameter PPI image of the KOHX radar.

4.2 Result of Other Elevation

The training and verification data in the experiment were all from 0.5° elevation scanning data of the KOHX radar. In order to verify the robustness of the system model in this paper, we tested the model by KOHX radar data of different scanning elevations at the same time. As shown in Fig. 12, they were the classification performance of the system model at 0.9°, 1.5°, 1.8° and 2.4° elevation.

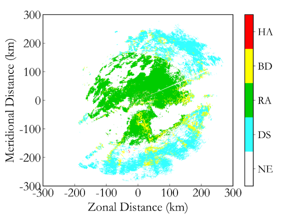
As shown in Fig. 12, as radar scanning elevation increased, central rainfall area and overall meteorological echo data would decrease. This was consistent with the actual situation. Meanwhile, the system model was compared with the fuzzy logic and single-sample SVM model. As shown in Tab. 4, the performance of the system model based on multi-sample fusion SVM was better than other models at different elevation angles.

Model	All samples	DS	RA	BD	HA
Model [9]	0.88	0.86	0.92	0.89	0.94
Model [10]	0.93	0.94	0.90	0.92	0.99
System model	0.97	0.96	0.98	0.93	1.0

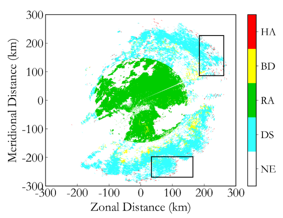
Tab. 3. Test accuracy of different models for the measured data by the KOHX radar.

Elevation	FHC	Single-sample SVM	System model
0.5°	85.47%	72.22%	97.21%
0.9°	88.38%	73.16%	97.47%
1.5°	98.29%	73.65%	97.51%
1.8°	89.51%	70.41%	97.49%
2.4°	90.12%	72.84%	96.93%

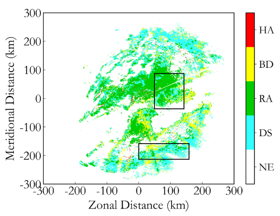
Tab. 4. Classification accuracy of different elevation angles of each model.



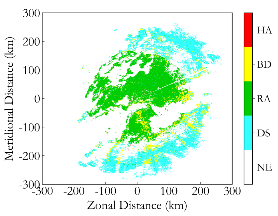
(a) NOAA



(b) FHC

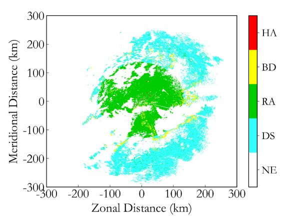


(c) Single-sample SVM

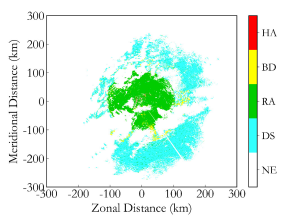


(d) System model

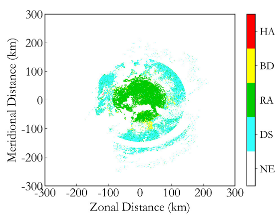
Fig. 11. Classification results of different model.



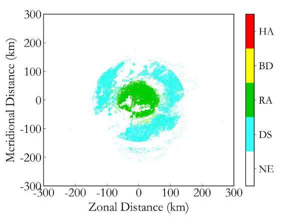
(a) 0.9°



(b) 1.5°



(c) 1.8°



(d) 2.4°

Fig. 12. Classification performance of the system model on different elevation.

5. Conclusion

In this paper, the hydrometeor classification algorithm based on multi-sample fusion SVM is proposed, and it classifies four types of target particles in the precipitation region. Aiming at such a disadvantage that fuzzy logic relies on expert experience to determine parameters, the SVM algorithm was used to prepare the dataset and train model through the four polarization parameters data of the KOHX radar. This data-driven model obtained parameters from samples. Compared with fuzzy logic algorithm, it has higher portability. According to the characteristics of particle distribution in the precipitation region, the dimension of feature information is improved by multi-sample fusion. The classification capability of the single-sample SVM model for DS and RA particles is improved. We tested the KOHX radar data of five different scanning elevations. The results showed that the average accuracy of system model based on the multi-sample fusion SVM was 97.32%, which was significantly higher than the single-sample SVM model 72.46% and the fuzzy logic algorithm model 88.55% and showed strong robustness.

The multi-sample fusion method in the azimuth direction was used to solve the problem of low classification accuracy of DS and RA particles in the precipitation region. The interference from a few of other noise hydrometeors on the fused samples was not considered during multi-sample fusion. In the future research, we will focus on how to solve the interference of noise hydrometeors in the fused samples to improve further the classification accuracy of hydrometeors in the precipitation area.

Acknowledgments

This study was thankfully supported in part by the Civil Aircraft Project of the Ministry of Industry and Information Technology, in the field of application of the Airborne Weather Radar project (No. MJ-2018-s-28).

References

[1] BYRD, A. D., PALMER, R. D., FULTON, C. J., et al. Doppler velocity bias mitigation through sidelobe whitening for multistatic weather radar. *IEEE Transactions on Geoscience and Remote Sensing*, 2021, vol. 59, no. 2, p. 1130–1142. DOI: 10.1109/TGRS.2020.2997882

[2] KIKUCHI, H., SUEZAWA, T., USHIO, T., et al. Initial observations for precipitation cores with X-band dual polarized phased array weather radar. *IEEE Transactions on Geoscience and Remote Sensing*, 2020, vol. 58, no. 5, p. 3657–3666. DOI: 10.1109/TGRS.2019.2959628

[3] PURNAMASARI, R., SUKSMONO, A. B., ZAKIA, I., et al. Compressive sampling of polarimetric Doppler weather radar processing via inverse fast Fourier transform. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2021, vol. 14, p. 5269–5284. DOI: 10.1109/JSTARS.2021.3081265

[4] LIM, S., CHANDRASEKAR, V., BRINGI, V. N. Hydrometeor classification system using dual-polarization radar measurements: Model improvements and in situ verification. *IEEE Transactions on Geoscience and Remote Sensing*, 2002, vol. 43, no. 4, p. 792–801. DOI: 10.1109/TGRS.2004.843077

[5] MARZANO, F. S., SCARANARI, D., VULPIANI, G. Supervised fuzzy-logic classification of hydrometeors using C-band weather radars. *IEEE Transactions on Geoscience and Remote Sensing*, 2007, vol. 45, no. 11, p. 3784–3799. DOI: 10.1109/TGRS.2007.903399

[6] PARK, H. S., RYZHKOY, A. V., ZRNIC, D. S., et al. The hydrometeor classification algorithm for the polarimetric WSR-88D: Description and application to an MCS. *Weather and Forecasting*, 2009, vol. 24, no. 3, p. 730–748. DOI: 10.1175/2008WAF2222205.1

[7] GRAZIOLI, J., TUTA, D., BERNE, A. Hydrometeor classification from polarimetric radar measurements: A clustering approach. *Atmospheric Measurement Techniques*, 2015, vol. 8, no. 1, p. 149–170. DOI: 10.5194/amt-8-149-2015

[8] HAN, L., SUN, J., ZHANG, W. Convolutional neural network for convective storm nowcasting using 3-D Doppler weather radar data. *IEEE Transactions on Geoscience and Remote Sensing*, 2020, vol. 58, no. 2, p. 1487–1495. DOI: 10.1109/TGRS.2019.2948070

[9] WANG, H., RAN, Y., DENG, Y., et al. Study on deep-learning-based identification of hydrometeors observed by dual polarization Doppler weather radars. *EURASIP Journal on Wireless Communications and Networking*, 2017, vol. 2017, no. 1, p. 173. DOI: 10.1186/s13638-017-0965-5

[10] LU, Y., KUMAR, J. Convolutional neural networks for hydrometeor classification using dual polarization doppler radars. In *International Conference on Data Mining Workshops (ICDMW)*. Beijing (China), 2019, p. 288–295. DOI: 10.1109/ICDMW.2019.00050

[11] LIANG, H., YANG, Z., SHI, F., et al. Energy and width features-based SVM for vehicles classification using low power consumption radar. In *IEEE 3rd International Conference on Electronic Information and Communication Technology (ICEICT)*. Shenzhen (China), 2020, p. 129–134. DOI: 10.1109/ICEICT51264.2020.9334245

[12] ZHAO, C., ZHAO, H., WANG, G., et al. Improvement SVM classification performance of hyperspectral image using chaotic sequences in artificial bee colony. *IEEE Access*, 2020, vol. 8, p. 73947–73956. DOI: 10.1109/ACCESS.2020.2987865

[13] ZHU, G., BLUMBERG, D. G. Classification using ASTER data and SVM algorithms. *Remote Sensing of Environment*, 2002, vol. 80, no. 2, p. 233–240. DOI: 10.1016/S0034-4257(01)00305-4

[14] LIN, Y., LEE, Y., WAHBA, G. Support vector machines for classification in nonstandard situations. *Machine Learning*, 2002, vol. 46, no. 1, p. 191–202. DOI: 10.1023/A:1012406528296

[15] FUANGKHON, P. Multiclass contour-preserving classification with support vector machine (SVM). *Intelligent Systems*, 2017, vol. 26, no. 2, p. 323–334. DOI: 10.1515/jisys-2015-0087

[16] LIU, X., HE, B., PU, K., et al. Classification of precipitation particles types using images from precipitation microphysical characteristics sensor. In *IEEE 5th International Conference on Computer and Communications (ICCC)*. Chengdu (China), 2019, p. 576–580. DOI: 10.1109/ICCC47050.2019.9064276

[17] ZHAO, Z., SONG, Y., CUI, F., et al. Point cloud features-based kernel SVM for human-vehicle classification in millimeter wave radar. *IEEE Access*, 2020, vol. 8, p. 26012–26021. DOI: 10.1109/ACCESS.2020.2970533

[18] AN, Y., DING, S., SHI, S., et al. Discrete space reinforcement learning algorithm based on support vector machine classification. *Pattern Recognition Letters*, 2018, vol. 111, p. 30–35. DOI: 10.1016/j.patrec.2018.04.012

[19] ADHIKARY, K., BHUSHAN, S., KUMAR, S., et al. Evaluating the performance of various SVM kernel functions based on basic features extracted from KDDCUP'99 dataset by random forest method for detecting DDoS attacks. *Wireless Personal Communications*, 2022, vol. 123, no. 4, p. 3127–3145. DOI: 10.1007/s11277-021-09280-8

[20] BENHAMMOUD, R., KACHA, A. Automatic classification of disordered voices based on a hybrid HMM-SVM model. *Communications Technology and Electronics*, 2021, vol. 66, no. 2, p. S139–S148. DOI: 10.1134/S1064226921140023

[21] KHAN, M., REZA, M. Q., SALHAN, A. K., et al. Classification of oils by ECOC based multi-class SVM using spectral analysis of acoustic signals. *Applied Acoustics*, 2021, vol. 183, p. 1–10. DOI: 0.1016/j.apacoust.2021.108273

[22] ROBERTO, M., ADIROSI, A., BALDINI, L., et al. Hydrometeor classification for X-band dual polarization radar on-board civil aircrafts. In *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*. Milan (Italy), 2015, p. 2319–2322 DOI: 10.1109/IGARSS.2015.7326272

[23] SCHIMIDT, M., TROMEL, S., RYZHKOV, A. V., et al. Severe hail detection: Hydrometeor classification for polarimetric C-band radars using fuzzy-logic and T-matrix scattering simulations. In *19th International Radar Symposium (IRS)*. Bonn (Germany), 2018, p. 1–7. DOI: 10.23919/IRS.2018.8448037

About the Authors . . .

Zehu LUO was born in Hefei, Anhui Province, China in 1997. He received the B.S. degree in the Communication Engineering from Anhui University of Technology, Maanshan, China in 2019. He is currently studying the M.S. degree at the College of Electronic and Information Engineering, Nanjing University of Aeronautics and Astronautics (NUAA), Nanjing, China. His current research interests include radar signal processing and machine learning.

Xudong WANG (corresponding author) was born in Huainan, Anhui Province, China in 1978. He received the B.S., M.S. and Ph.D. degrees in the Communication Engineering from Nanjing University of Aeronautics and Astronautics (NUAA), Nanjing in 2001, 2004, 2008, respectively. Currently, he is an Associate Professor at the College of Electronic and Information Engineering, Nanjing University of Aeronautics and Astronautics. His research interests include radar signal processing, FPGA hardware implementation.

Ling WANG was born in Luoyang, Henan Province, China in 1977. She received the B.S., M.S. and Ph.D. degrees in the Communication Engineering from the Nanjing University of Aeronautics and Astronautics (NUAA), Nanjing in 2000, 2003, 2006, respectively. Currently, she is a Professor at the College of Electronic and Information Engineering, Nanjing University of Aeronautics and Astronautics. Her research interests include radar signal processing, inverse scattering, microwave imaging technology.

Guiguang XU was born in Guilin, Guangxi Province, China in 1996. He received the B.S. degree in the Electronic Information Engineering from East China University of Technology, Jiangxi, China in 2018. Currently, he is stuying the M.S. degree at the College of Electronic and Information Engineering, Nanjing University of Aeronautics and Astronautics (NUAA), Nanjing. His current research interests include radar signal recognition and machine learning.

Yongxing GAO was born in Dingxi, Gansu Province, China in 1996. He received the B.S. and M.S. degrees in the Communication Engineering from Nanjing University of Aeronautics and Astronautics (NUAA), Nanjing, China in 2019 and 2022 respectively. His research interests include radar signal processing and machine learning.