Bagged Extended Nearest Neighbors Classification for Anomalous Propagation Echo Detection

Hansoo Lee

School of Electrical and Computer Engineering
Pusan National University
Busan, Republic of Korea, 46241
Email: hansoo@pusan.ac.kr

Hye-Young Han

Weather Radar Center
Korea Meteorological Administration
Seoul, Republic of Korea, 07062
Email: hyhan98@gmail.com

Sungshin Kim

School of Electrical and Computer Engineering
Pusan National University
Busan, Republic of Korea, 46241
Email: sskim@pusan.ac.kr

***Abstract*—Radar is one of essential and popular devices in weather prediction process because of its wide array of advantages. Unfortunately, the observation results contains lots of unwanted radar signals and they disrupt forecasting process. The representative non-precipitation echoes are permanent, spurious, and anomalous propagation echoes. Among them, the anomalous propagation echo can be a source of severely negative influences in a quantitative precipitation estimation. Therefore, a reliable automatic systems for identifying the anomalous propagation echo is needed. In this paper, we suggest a novel k-nearest neighbors algorithm, by combining the Hamamoto's bootstrap II method and the extended nearest neighbors for improving performance of the classifier. Using the actual appearance cases of the anomalous propagation echo, it is confirmed that the suggested method is better than the k-nearest neighbors and the extended nearest neighbors.**

Keywords–Extended nearest neighbors; Hamamoto's bootstrap II; Anomalous propagation echo; Weather prediction; Classification.

I. Introduction

Weather radar is an essential device in weather forecasting process because of its wide array of advantages. For example, the weather radar is capable of near-real time observation with high resolution monitoring over a wide area. Also, the radar can observe development, movement of precipitation areas, and calculate rainfall intensity [1]. By virtue of its advantages, the weather radars are installed in many places of the world and actively involved in various kinds of weather-related fields such as estimating precipitation, disaster management, and so on. Unfortunately, the weather radar has no function to make meteorological observation selectively. Namely, the observation results contains lots of unwanted radar signals inevitably, which disrupt weather prediction process and make low prediction accuracy. Therefore, a quality control process is an indispensable part to remove these unwanted radar signals, so-called non-precipitation echoes [2].

The representative non-precipitation echoes are permanent, spurious, and anomalous propagation echoes. The permanent echoes are caused by mountains, skyscrapers, or other kinds of surface obstacles blocking the radar beam inside the observation area [3]. The spurious echoes are caused by various reasons such as chaff in use of military exercises, jamming by other radars, and so on [4] [5]. And the anomalous propagation echoes are caused by refracted radar beam. It appears in certain conditions of non-standard refraction in the atmosphere when the radar beam passes through air of varying density. The resultant echo represents reflection of the ground or not a meteorological target, and it can be misinterpreted as a heavy precipitation [6].

Considering that the anomalous propagation echo can be a source of significantly negative influences in a quantitative precipitation estimation, a reliable automatic systems for identifying the anomalous propagation echo is needed. Unless, there is a chance to make erroneous calculations of quantitative precipitation estimation or other types of mislead forecasting results.

To classify the anomalous propagation echo in the radar data automatically, several researches using data mining techniques have been studied: fuzzy logic [7] [8]; Bayesian approaches [9] [10]; artificial neural networks [11] [12]; support vector machine [13]; and so on. According to these researches, two important things can be derived. First, the previous researches consider selecting the most efficient classifier for implementing the automated anomalous propagation echo identification system with serious consideration. Second, these researches are focused on a single classification methods.

There are various types of classification methods in machine learning, and used to solve a variety of practical problems. Among them, the k-nearest neighbors [14] algorithm has been a successful choice under many circumstances because of its advantages, such as easy implementation and a good performance without requiring knowledge of a probability distribution function. This decision rule provides a simple nonparametric procedure for the assignment of a class label to the input pattern based on the class labels represented by the k-closest neighbors of the vector [15].

However, the k-nearest neighbors algorithm has some drawbacks. One of representative drawbacks is that k-nearest neighbors algorithm is sensitive to the scale or variance of the distributions of the pre-defined class data. In other words, the nearest neighbors of an unknown sample will tend to be dominated by the class with the highest density [16] [17]. Fortunately, the novel kind of k-nearest neighbors algorithm is suggested, called as the extended nearest neighbors that uses the generalized class-wise statistics [18].

Furthermore, we consider a bagging method in order to improve performance of the extended nearest neighbors. By

 ISBN: 978-1-61208-518-0

generating an artificial training samples from the original training samples and obtaining classification results from majority vote, it is possible to improve performance of the extended nearest neighbors algorithm. However, taking into account that small changes in the training sample generated by sampling with replacement do not lead to significantly different classification results of k-nearest neighbors algorithm due to its stable characteristics [19], we consider Hamamoto II bootstrap method [20], which generates a new training sample by resampling and locally transforming.

Consequently, we suggest a novel type of nearest neighbors algorithm by combining Hamamoto's bootstrap II method and extended nearest neighbors in this paper. The rest of the paper is organized as follow. Section 2 explains the bagged extended nearest neighbors with its essential components, extended nearest neighbors and bagging method. And in Section 3, the anomalous propagation echo is briefly elucidated. After that, the experimental results with actual radar observation data are described in Section 4. Finally, the conclusion and future works are showed in Section 5.

II. METHODS

To illustrate the principles of the bagged extended nearest neighbors, fundamental algorithms should be described. This section explains extended nearest neighbors, bagging and Hamamoto's bootstrap II, and the suggested bagged extended nearest neighbors.

A. Extended Nearest Neighbors

k-nearest neighbors algorithm is a popular nonparametric method used for both classification and regression [21]. The input consists of the k closest training samples measured by distance in feature space, and the output indicates a class. An object is classified by a majority vote of its neighbors, with the object being assigned to the class most common among its nearest neighbors.

k-nearest neighbors classifier has remarkable advantages, such as easy implementation, competitive performance, independent of the underlying data distribution, and so on. However, it also has some disadvantages. One of typical weaknesses is that k-nearest neighbors method is sensitive to the scale or variance of distributions of the pre-defined classes. In other words, the nearest neighbors of an unknown object will tend to be dominated by the class with the highest density. This has been a long-standing limitation of the classic k-NN method [16] [17].

In order to solve the problem, a novel nearest neighbors algorithm is suggested, namely extended nearest neighbors. The extended nearest neighbors makes a prediction in a "two-way communication" style using the generalized class-wise statistics T_i^j: it considers not only who are the nearest neighbors of the test sample, but also who consider the test sample as their nearest neighbors [18].

The entire process of the extended nearest neighbors is described in Fig. 1, which considers a two-class problem. The first step of the extended nearest neighbors is applying k-nearest neighbors to the training samples. Let's assume S is an entire training data set, $\mathrm{S} = \mathrm{S}_1 \cup \mathrm{S}_2$, S_1 and S_2 indicate the samples in class 1 and class 2, respectively. Each training sample saves its k nearest neighbors and distances. The second step is getting one sample $\mathbf{z}$ from testing data $\mathbf{Z}$, $\mathbf{z} \in \mathbf{Z}$.

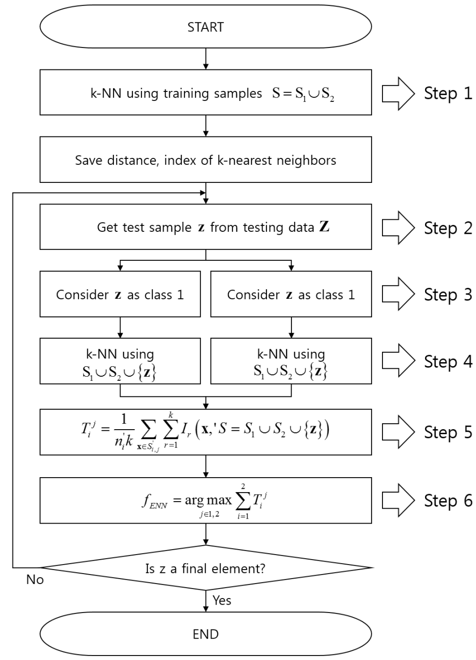

Figure 1. Principles of extended nearest neighbors

The third step is a core step. The obtained testing sample $\mathbf{z}$ is considered as class 1 and class 2, simultaneously and individually. And the fourth step is applying k-nearest neighbors again to union set of the training data set and the testing sample, $\mathrm{S} = \mathrm{S}_1 \cup \mathrm{S}_2 \cup \{\mathbf{z}\}$. In the fifth step, the generalized class-wise statistics is applied to estimate the influences of given $\mathbf{z}$ using (1).

$$T_i^j = \frac{1}{n_i' k} \sum_{\mathbf{x} \in \mathrm{S}'_{i,j}} \sum_{r=1}^{k} I_r \left(\mathbf{x}, \mathrm{S}' = \mathrm{S}_1 \cup \mathrm{S}_2 \cup \{\mathbf{Z}\} \right) \qquad i, j = 1, 2 \tag{1}$$

where $\mathbf{x}$ denotes one of samples in $\mathrm{S}_1 \cup \mathrm{S}_2 \cup \{\mathbf{z}\}$. And k is the user-defined parameter of the number of the nearest neighbors. n_i' is the size of $\mathrm{S}'_{i,j}$ and $\mathrm{S}'_{i,j}$ is defined as

$$\mathrm{S}'_{i,j} = \begin{cases} \mathrm{S}_i \cup \{\mathbf{Z}\}, & \text{when } j = i \\ \mathrm{S}_i, & \text{when } j \neq i \end{cases} \tag{2}$$

The indicator function indicates whether both the sample $\mathbf{x}$ and its r-th nearest neighbor belong to the same class as shown in (3)

$$I_r(\mathbf{x}, \mathrm{S}) = \begin{cases} 1, & \text{if } \mathbf{x} \in \mathrm{S}_i \text{ and } \mathrm{NN}_r(\mathbf{x}, \mathrm{S}) \in \mathrm{S}_i \\ 0, & \text{otherwise} \end{cases} \tag{3}$$

where $\mathrm{NN}_r(\mathbf{x}, \mathrm{S})$ denotes the r-th nearest neighbor of $\mathbf{x}$ in S. This equation means for either class, if both the sample $\mathbf{x}$ and its r-th nearest neighbor in the pool of S belong to the same class, then the outcome of the indicator function $I_r(\mathbf{x}, \mathrm{S})$ equals 1; otherwise, it equals 0.

In sixth step, the generalized class-wise statistics are derived. Given two-class classification problem, we have four generalized class-wise statistics: T_1^1, T_2^1, T_1^2 and T_2^2. The extended nearest neighbors classifier predicts its class membership according to the following target function

$$f_{\text{ENN}} = \arg\max_{j \in 1,2} \sum_{i=1}^{2} T_i^j \tag{4}$$

Using (4), the class of unknown sample $\mathbf{z}$ is defined. And it is repeated until all the testing elements are went through the processes, from the second to sixth step.

B. Bagging

Bagging (Bootstrap aggragating) is a type of ensemble method, which uses bootstrap to improve the performance of the classifier [19]. With bootstrap, many new training samples are generated from the original training set. Then, for each bootstrap training set, the test object is classified using k-nearest neighbors. As a result of this process, a series of classification results for each object are obtained. The test object is finally assigned to the class where it was classified by majority vote.

There are several possible setups for bootstrap [19] [22] [20]. The classical bootstrapping uses random sampling with replacement. This was already used with k-nearest neighbors but without satisfactory results due to the "stability" of the k-nearest neighbors [19]. k-nearest neighbors is "stable" because small changes in the training data do not lead to significantly different classification results.

However, Hamamoto's bootstrap method [20] is considerable because all the objects in the original training set participate in creating the bootstrap training set using locally weighted sum as shown in (5). Fig. 2 explains the principles of Hamamoto's bootstrap II method when $k = 3$ in a two-class problem. The given data is separated by class and applied k-nearest neighbors individually including selected sample itself. The generated class data is derived using locally weighted sum, and the process is repeated until all the data is processed. The entire process is shown below.

1) Select one sample $\mathbf{x}_i$ from $\mathbf{X}_c$.
2) Using Euclidean distance, find the r nearest neighbors $\mathbf{x}_{i,1}, \mathbf{x}_{i,2}, \cdots, \mathbf{x}_{i,r}$ from $\mathbf{X}_c$.
3) Compute a new bootstrap sample $\mathbf{x}_i^b$ as a weighted average of r nearest neighbors, including the selected object i itself ($\mathbf{x}_{i,0}$):

$$\begin{aligned} \mathbf{x}_i^b &= \sum_{j=0}^{r} \omega_j \mathbf{x}_{i,j} \\ &= \omega_0 \mathbf{x}_{i,0} + \omega_1 \mathbf{x}_{i,1} + \cdots + \omega_r \mathbf{x}_{i,r} \end{aligned} \tag{5}$$

   The weight ω_j is given by

$$\omega_j = \frac{\Delta_j}{\sum_{c=0}^{r} \Delta_c}, \qquad 0 \le j \le r \tag{6}$$

   where Δ_j is chosen from a uniform distribution on $[0, 1]$ and $\sum_{j=0}^{r} \omega_j = 1$.
4) Step 1) to 3) are run for all the objects $i = 1, \cdots, l_c$ of $\mathbf{x}_c$, thus obtaining a new matrix $\mathbf{X}_c^b$ for class $c = 1$.
5) Step 1) to 4) are repeated for the other classes $c = 2, \cdots, C$.
6) The bootstrap matrices $\mathbf{X}_c^b$ generated for all the classes are then adjoined to obtain the bootstrap training set $\mathbf{X}^b$ and $\mathbf{X}^b$ is used to classify the test object.
7) Step 1) to 5) are repeated B times and the results are finally combined.

C. Bagged Extended Nearest Neighbors

Combining the Hamamoto's II bootstrap method and the extended nearest neighbors, we suggest the bagged extended nearest neighbor as shown in Fig. 3. The operating principle is as follow. First, the training data is divided into r number of data by Hamamoto's II bootstrap method. The samples inside the divided data is not identical to the original training data, because it is derived by (5). Second, each generated data is applied to extended nearest neighbors classifier respectively. Third, the testing data is applied each trained extended nearest neighbors. Fourth, the results are gathered for voting using (7).

$$f_{\text{Bagged\_ENN}}(\mathbf{X}) = \arg\max_{i} \sum_{j=1}^{r} I(f_{ENN_j}(\mathbf{x}_j) = i) \tag{7}$$

where $I(f_{ENN_j}(\mathbf{x}_j) = i)$ is an indicator function, which derives 1 when they are matched, 0 otherwise.

III. Anomalous Propagation Echo

For ground-based radar propagation at quasi-horizontal beam elevation, the sensitive terms are the vertical gradient of temperature distribution and water vapor. The quantity used to describe the radar beam propagation is the refractivity N, a particular form of the refractive index n used because n is close to unity for the atmosphere [23]. The refractivity can be approximated with the simplified expression in (8)

$$(n-1) \times 10^6 = N = \frac{0.776p}{T} + \frac{3730e}{T^2} \tag{8}$$

where p is the total atmospheric pressure, e is the water vapor partial pressure, and T is the temperature [24].

Let's assume α is the angle of the radar ray with the surfaces of constant N, and let's consider an arc ∂s along a radar ray. And assume that $\partial\alpha$ is the corresponding variation of the angle of the tangent to this ray. The curvature of the ray is C and the radius of curvature ρ with $C = 1/\rho = d\alpha/ds$. From geometrical consideration, the radius of curvature is related to the vertical gradient of refractivity $\partial N/\partial z$ where z is the vertical coordinate, as shown in (9)

$$\frac{1}{\rho} = -\frac{1}{n}\frac{\partial N}{\partial z}\cos\alpha \times 10^6 \tag{9}$$

where ρ in meters if z is in meters. For an elevation close to zero, it can be re-written as shown in (10)

$$\frac{1}{\rho} \approx -\frac{\partial N}{\partial z} \times 10^6 \tag{10}$$

There are four types of propagation: subrefraction, normal refraction, superrefraction, and ducting as follows. [25].

- Subrefraction
  - The radar beam bends less than usual

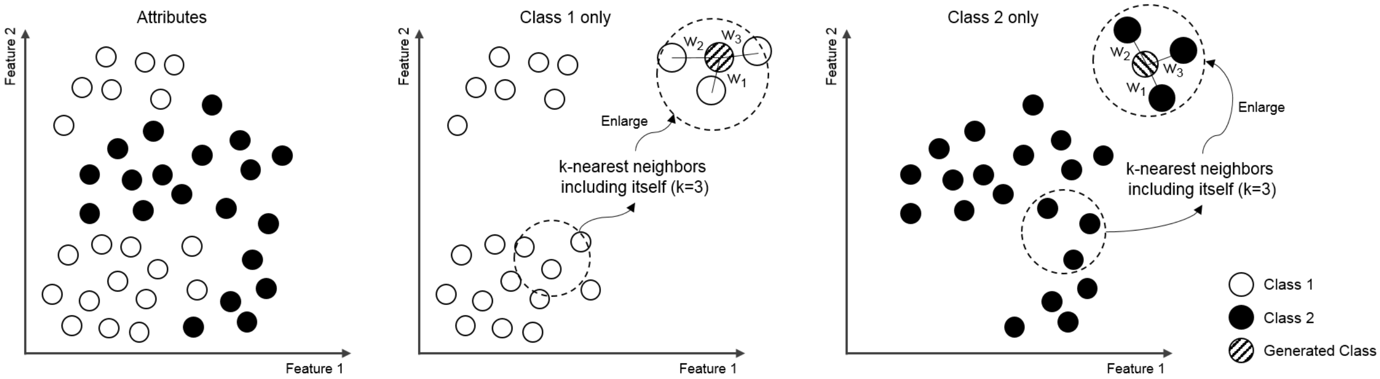

Figure 2. Principles of Hamamoto's bootstrap II method

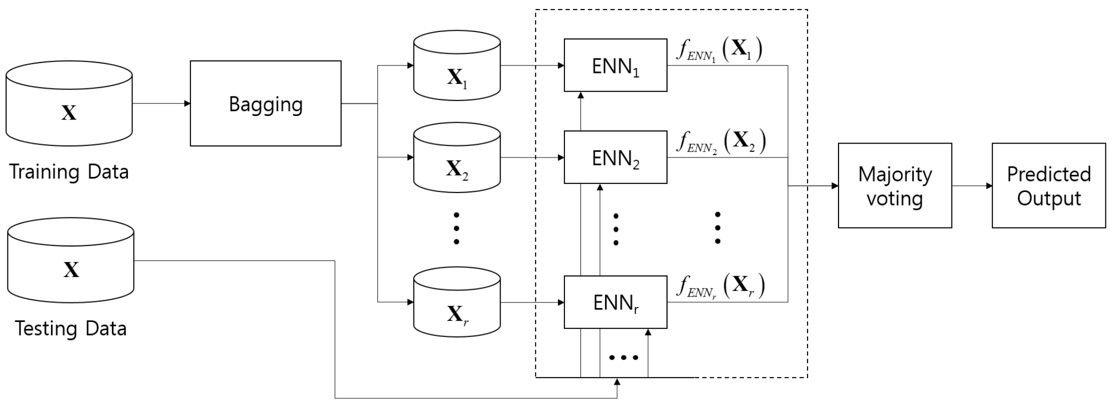

Figure 3. Overall structure of suggested method

- $\frac{\partial N}{\partial z} > 0$
- Normal refraction
  - Considered as standard radar beam trajectory
  - Corresponding to rays bending downward with $\rho \geq \rho_e$
  - $\rho_e \approx 6371km$: the radius of curvature of the Earth's surface
  - $\frac{\partial N}{\partial z} = 0$
- Superrefraction
  - The radar beam bends more towards the ground surface
  - $-0.157 \leq \frac{\partial N}{\partial z} \leq -0.0787m^{-2}$
- Ducting
  - Extreme case of superrefraction
  - The ground surface can be observed as objects in the atmosphere
  - $\frac{\partial N}{\partial z} \leq -0.157m^{-2}$

The subrefraction, superrefraction, and ducting are categorized as the anomalous propagation echoes. The echoes can be lead to erroneous calculations of quantitative rainfall estimation. Therefore, reliable automatic detection and removal of anomalous propagation echoes is one of the essential problems in this area. In the weather forecasting process, there are some complicated expert's knowledge for removing the anomalous propagation echo in the radar data as shown below.

1) The echo moves with near zero Doppler velocity $\approx 0m/s$
2) The maximum altitude of the echo is low
3) The reflectivity distribution is discontinuous in vertical and horizontal way

IV. EXPERIMENTAL RESULTS

In order to evaluate and compare the nearest neighbors classifiers, this paper selected actual appearance cases of the anomalous propagation echo. According to the expert's knowledge described in previous section, it is confirmed that Doppler velocity, reflectivity, and altitude are essential input variables for classification. Therefore, we use five features as inputs in this paper: centroid altitude of the cluster, average reflectivity, maximum reflectivity, average Doppler velocity, and minimum Doppler velocity.

Considering that the suggested system is a type of binary classifier, we applied accuracy, sensitivity and specificity as verifications of each classifier performance as shown in (11), (12), and (13).

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{11}$$

$$\text{Sensitivity} = \frac{TP}{TP + FN} \tag{12}$$

$$\text{Specificity} = \frac{TN}{TN + FP} \tag{13}$$

TABLE I. PERFORMANCE COMPARISON OF k-NN, ENN, AND BAGGED ENN

		Accuracy		Sensitivity		Specificity	
		Average	StDev	Average	StDev	Average	StDev
k=3	k-NN	0.8075	0.0237	0.8335	0.0325	0.8431	0.0381
	ENN	0.8055	0.0195	0.8258	0.0318	0.8325	0.0397
	BENN	**0.8794**	**0.0075**	**0.8690**	**0.0114**	**0.8638**	**0.0137**
k=5	k-NN	0.8022	0.0183	0.8151	0.0245	0.8195	0.0313
	ENN	0.8029	0.0145	0.8375	0.0355	0.8496	0.0417
	BENN	**0.8593**	**0.0079**	**0.8598**	**0.0106**	**0.8585**	**0.0122**
k=7	k-NN	0.8000	0.0148	0.8205	0.0158	0.8297	0.0195
	ENN	0.8063	0.0238	0.8399	0.0360	0.8520	0.0411
	BENN	**0.8516**	**0.0078**	**0.8560**	**0.0132**	**0.8561**	**0.0156**
k=9	k-NN	0.8051	0.0187	0.8343	0.0302	0.8455	0.0350
	ENN	0.8059	0.0248	**0.8481**	0.0409	**0.8536**	0.0591
	BENN	**0.8399**	**0.0050**	0.8422	**0.0091**	0.8415	**0.0113**

where TP is true positive, TN is true negative, FP is false positive, and FN is false negative. Also, in this paper, the true means the anomalous propagation echo, and the false indicates the non-anomalous propagation echo, respectively.

As shown in Table I, we compared the suggested method, BENN which is a written abbreviation for bagged extended nearest neighbors, to other nearest neighbors classifiers, the k-nearest neighbors and the extended nearest neighbors. To avoid a tie vote, we selected the number of nearest neighbors as odd numbers under 10. In bagged extended nearest neighbors, the number of k for bagging is set to 5. The experiments are conducted 30 times in each case. The average and standard deviation values of accuracy, sensitivity, specificity are shown in Table I.

The bagged extended nearest neighbors shows the best accuracy regardless of the number of k. And it shows the best sensitivity and specificity in most of cases. In $k = 9$ case, the sensitivity and specificity of the bagged extended nearest neighbors are slightly lower than the extended nearest neighbors. However, considering that its standard deviations of those factors are small, it seems more stable than the extended nearest neighbors.

Fig. 4 shows the performances of nearest neighbors classifiers in a form of boxplot: the first, fourth, seventh, and tenth indicates the k-nearest neighbors; the second, fifth, eighth, and eleventh indicates the extended nearest neighbors; and the third, sixth, ninth, and twelfth indicates the bagged extended nearest neighbors, respectively. Fig. 4 (a) describes that the suggested method, bagged extended nearest neighbors, shows impressive accuracy distribution than others when $k = 3$. From Fig. 4 (b) to (d), even though the accuracy of the bagged extended nearest neighbors is gradually decreased, it shows better result than other results. Consequently, it is confirmed that the bagged extended nearest neighbors classifier has the best performance in most cases.

Fig. 5 shows one of graphically described experiment results using the bagged extended nearest neighbors. Fig. 5 (a) indicates a mixed case of precipitation echo and anomalous propagation echo that the upper area is represented as the anomalous propagation echo. Fig. 5 (b) describes the identified anomalous propagation echo, and Fig. 5 (c) shows the radar image without anomalous propagation echo. As a result, it is also confirmed that the bagged extended nearest neighbors can detect the anomalous propagation echo successfully.

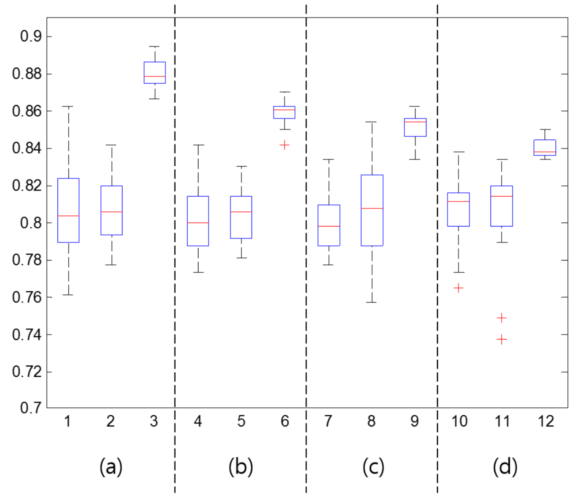

Figure 4. Accuracy comparison of k-NN, ENN, and Bagged ENN methods: (a) k=3, (b) k=5, (c) k=7, (d) k=9

V. CONCLUSION

The anomalous propagation echo occurs frequently and has similar characteristics to precipitation echoes. And it should be removed because it has a serious effect on the quantitative precipitation estimation. Therefore, we suggest a novel nearest neighbors classifier by combining bagging method and extended nearest neighbors for identifying anomalous propagation echo in radar data. Using the actual appearance cases of the anomalous propagation echo, it is confirmed that the suggested method is better than other nearest neighbors classifiers.

In the future work, we will continue to study not only for enhancing classification performance using parameter optimization but also for applying to other representative non-precipitation echoes such as chaff and sea clutter. Furthermore, based on the fact that the classification technique is one of the most important of the data mining method, the proposed method in this paper is expected to be able to perform an important role in various fields.

ACKNOWLEDGMENT

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2014R1A1A2056958).

REFERENCES

[1] S. Jebson, “Fact sheet number 15: Weather radar,” 2007.

[2] S. Moszkowicz, G. J. Ciach and W. F. Krajewski, “Statistical detection of anomalous propagation in radar reflectivity patterns,” Journal of Atmospheric and Oceanic Technology, vol. 11, no. 4, pp. 1026-1034, 1994.

[3] U. Germann, G. Galli, M. Boscacci and M. Bolliger, “Radar precipitation measurement in a mountainous region,” Quarterly Journal of the Royal Meteorological Society, vol. 132, no. 618, pp. 1669-1692, 2006.

[4] Y. H. Kim, S. Kim, H.-Y. Han, B.-H. Heo and C.-H. You, “Real-time detection and filtering of chaff clutter from single-polarization doppler radar data,” Journal of Atmospheric and Oceanic Technology, vol. 30, no. 5, pp. 873-895, 2013.

[5] J. Sugier, J. P. du Chatelet, P. Roquain and A. Smith, “Detection and removal of clutter and anaprop in radar data using a statistical scheme based on echo fluctuation,” Proceedings of ERAD (2002), pp. 17-24, 2002.

[6] J. Pamment and B. Conway, “Objective identification of echoes due to anomalous propagation in weather radar data,” Journal of Atmospheric and Oceanic Technology, vol. 15, no. 1, pp. 98-113, 1998.

[7] Y.-H. Cho, G. W. Lee, K.-E. Kim and I. Zawadzki, “Identification and removal of ground echoes and anomalous propagation using the characteristics of radar echoes,” Journal of Atmospheric and Oceanic Technology, vol. 23, no. 9, pp. 1206-1222, 2006.

[8] M. Berenguer, D. Sempere-Torres, C. Corral and R. Sánchez-Diezma, “A fuzzy logic technique for identifying nonprecipitating echoes in radar scans,” Journal of Atmospheric and Oceanic Technology, vol. 23, no. 9, pp. 1157-1180, 2006.

[9] J. R. Peter, A. Seed and P. J. Steinle, “Application of a Bayesian classifier of anomalous propagation to single-polarization radar reflectivity data,” Journal of Atmospheric and Oceanic Technology, vol. 30, no. 9, pp. 1985-2005, 2013.

[10] S. Rennie, M. Curtis, J. Peter, A. Seed, P. Steinle and G. Wen, “Bayesian Echo Classification for Australian Single-Polarization Weather Radar with Application to Assimilation of Radial Velocity Observations,” Journal of Atmospheric and Oceanic Technology, vol. 32, no. 7, pp. 1341-1355, 2015.

[11] R. B. Da Silveria and A. R. Holt, “An automatic identification of clutter and anomalous propagation in polarization-diversity weather radar data using neural networks,” IEEE Transactions on Geoscience and Remote Sensing, vol. 39, no. 8, pp. 1777-1788, 2001.

[12] M. Grecu and W. F. Krajewski, “An efficient methodology for detection of anomalous propagation echoes in radar reflectivity data using neural networks,” vol. 17, no. 2, pp. 121-129, 2000.

[13] H. Lee, E. K. Kim and S. Kim, “Anomalous Propagation Echo Classification of Imbalanced Radar Data with Support Vector Machine,” Advances in Meteorology, vol 2016, pp. 1-13, 2016.

[14] T. Cover and P. Hart, “Nearest neighbor pattern classification,” IEEE transactions on information theory, vol. 13, no. 1, pp. 21-27, 1967.

[15] J. M. Keller, M. R. Gray and J. A. Givens, “A fuzzy k-nearest neighbor algorithm,” IEEE transactions on systems, man, and cybernetics, no. 4, pp. 580-585, 1985.

[16] S. Har-Pelec, P. Indyk and R. Motwani, “Approximate nearest neighbor: Towards removing the curse of dimensionality,” Theory of computing, vol. 8, no. 1, pp. 321-350, 2012.

[17] J. H. Friedman, S. Steppel and J. Tukey, A nonparametric procedure for comparing multivariate point sets, Stanford Linear Accelerator Center Computation Research Group Technical Memo, no. 153, 1973.

[18] B. Tang and H. He, “ENN: Extended nearest neighbor method for pattern recognition [research frontier],” IEEE Computational Intelligence Magazine, vol. 10, no. 3, pp. 52-60, 2015.

[19] L. Breiman, “Bagging predictors,” Machine learning, vol. 24, no. 2, pp. 123-140, 1996.

[20] Y. Hamamoto, S. Uchimura and S. Tomita, “A bootstrap technique for nearest neighbor classifier design,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 19, no. 1, pp. 73-79, 1997.

[21] J. MacQueen, “Some methods for classification and analysis of multivariate observations,” in Proceedings of the fifth Berkeley symposium on mathematical statistics and probability, vol. 1, no. 14, pp. 281-297, 1967.

[22] R. Wehrens, H. Putter and L. M. Buydens, “The bootstrap: a tutorial,” Chemometrics and intelligent laboratory systems, vol. 54, no. 1, pp. 35-52, 2000.

[23] F. Mesnard and H. Sauvageot, “Climatology of anomalous propagation radar echoes in a coastal area,” Journal of Applied Meteorology and Climatology, vol. 49, no. 11, pp. 2285-2300, 2010.

[24] B. R. Bean and E. Dutton, Radio meteorology, Dover Publications, 1966.

[25] P. Lopez, “A 5-yr 40-km-resolution global climatology of superrefraction for ground-based weather radars,” Journal of applied meteorology and climatology, vol. 48, no. 1, pp. 89-110, 2009.

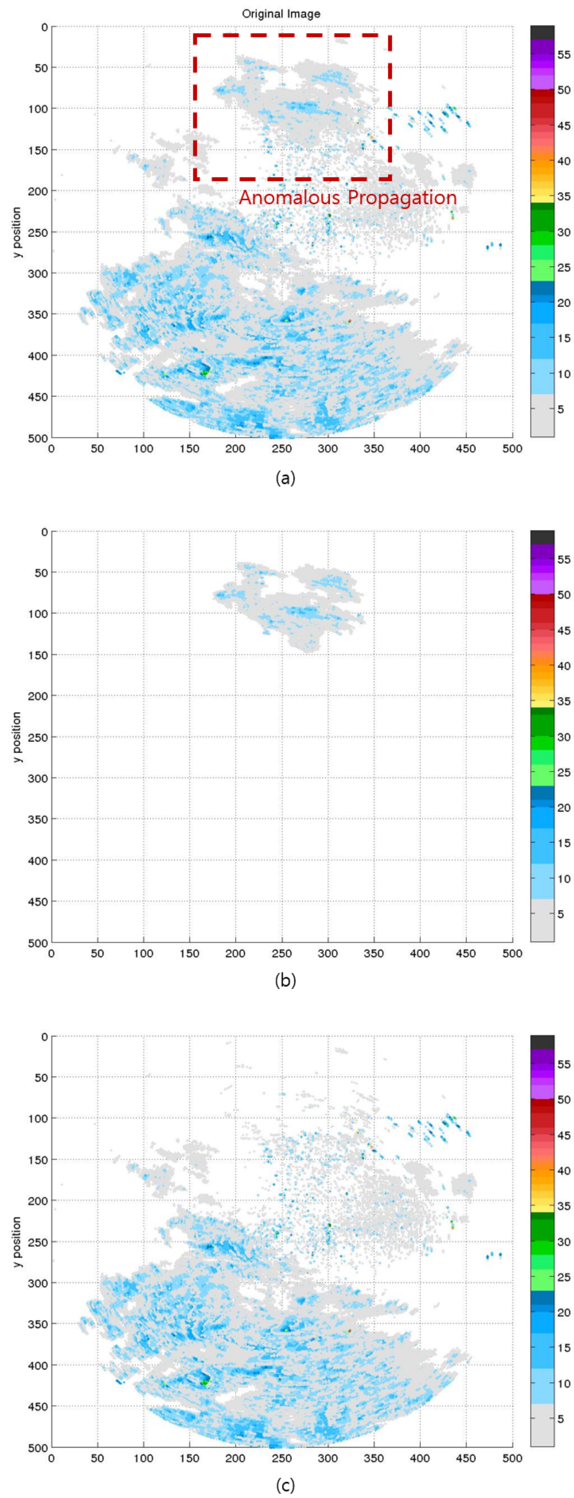

Figure 5. Experimental result: (a) Original radar image, (b) Identified anomalous propagation echo image, (c) Modified radar image