A Machine Learning Approach for Classifying Bird and Insect Radar Echoes with S-Band Polarimetric Weather Radar🖉

Precious Jatau,[a,b,c] Valery Melnikov,[b,d] and Tian-You Yu[a,c,e]

[a] Advanced Radar Research Center, University of Oklahoma, Norman, Oklahoma
[b] Cooperative Institute for Mesoscale Meteorological Studies, University of Oklahoma, Norman, Oklahoma
[c] School of Electrical and Computer Engineering, University of Oklahoma, Norman, Oklahoma
[d] National Severe Storms Laboratory, Norman, Oklahoma
[e] School of Meteorology, University of Oklahoma, Norman, Oklahoma

ABSTRACT: The S-band WSR-88D is sensitive enough to observe biological scatterers like birds and insects. However, their nonspherical shapes and frequent collocation in the radar resolution volume create challenges in identifying their echoes. We propose a method of extracting bird (insect) features by coherently averaging dual-polarization measurements from multiple radar scans containing bird (insect) migration. Additional features are also computed to capture aspect and range dependence and the variation of these echoes over local regions. Next, ridge classifier and decision tree machine learning algorithms are trained, first only with the averaged dual-polarization inputs and then different combinations of the remaining features are added. The performance of all models for both methods are analyzed using metrics computed from the test data. Further studies on different patterns of birds/insects, including roosting birds, bird migration, and insect migration cases, are used to further investigate the generality of our models. Overall, the ridge classifier using only dual-polarization variables was found to perform consistently well across all these tests. Our recommendation is that this classifier can be used operationally on the U.S. Next Generation Weather Radars (NEXRAD), as a first step in classifying biological echoes. It would be used in conjunction with the existing hydrometeor classification algorithm (HCA), where the HCA would first separate biological from nonbiological echoes, then our algorithm would be applied to further separate biological echoes into birds and insects. To the best of our knowledge, this study is the first to train a machine learning classifier that is capable of detecting diverse patterns of bird and insect echoes, based on dual-polarization variables at each range gate.

KEYWORDS: Algorithms; Data processing; Data quality control; Remote sensing; Weather radar signal processing; Artificial intelligence; Classification; Data science; Decision trees; Machine learning; Model interpretation and visualization; Neural networks; Pattern recognition

1. Introduction

The U.S. national network of S-band polarimetric Doppler weather radars (WSR-88D) consists of 160+ systems deployed across the continental United States, Alaska, Hawaii, and Puerto Rico. Each WSR-88D measures six variables composed of three legacy single-polarization variables and three dual-polarization variables. The legacy variables are the radar reflectivity factor (Z) which is proportional to the power of the received signal, Doppler velocity (V_r) which is determined from the power-weighted mean Doppler frequency shift of targets within the radar sampling volume, and spectrum width (σ_V) which measures the variability of Doppler velocities within the sampling volume (Kumjian 2013; Doviak and Zrnić 1993; Rinehart 2004). The dual-polarization variables include differential reflectivity (Z_{DR}), the logarithmic ratio of the reflectivity factors at horizontal (H) and vertical (V) polarized waves, differential propagation phase shift (Φ_{DP}), the difference in phase shift between H and V polarizations, and

cross correlation coefficient (ρ_{HV}) which is a measure of the diversity in type, shape, and/or orientation of scatterers in the sampling volume (Kumjian 2013).

Though primarily designed to monitor severe weather and precipitation, the WSR-88D can detect biological scatterers such as birds, bats, and insects (Gauthreaux et al. 2008). Thus, the WSR-88D can be exploited to better understand the broadscale behavior of biological targets. For example, the long-distance transport of migrants has frequently been associated with outbreaks of agricultural pests and insect-borne plant or animal diseases (Drake and Farrow 1988; Pedgley 1982). Therefore, the large-scale radar monitoring can improve our understanding of the spread of avian diseases by allowing a detailed mapping of migratory flyways (Peterson and Williams 2008; Dokter et al. 2011). Additionally, bird strikes are a serious aviation hazard for low-level flights (Dokter et al. 2011). Strong potential exists for the use of WSR-88D as a continental bird surveillance system to improve aviation safety (Dokter et al. 2011; Jatau and Melnikov 2019). Another application is the identification of wind tracers. Insects have been found to be passive flyers, driven by the wind while birds are active fliers and can contaminate the wind derived from radar (Bachmann and Zrnić 2007). Zhang et al. (2005) and Liu et al. (2005) used the features of reflectivity and Doppler velocity fields to obtain velocities contaminated by birds. Therefore, proper differentiation

of insects from birds in the atmosphere can improve the quality of radar wind products.

Though many advances have been made in characterizing hydrometeor types (Zrnić and Ryzhkov 1999; Park et al. 2009; Chandrasekar et al. 2013), the classification of biological taxa is still an active research field (Jatau and Melnikov 2019; Gauthreaux and Diehl 2020; Radhakrishna et al. 2019; Kilambi et al. 2018). The hydrometeor classification algorithm (HCA) on NEXRAD has eight classes for weather compared to one for biological echoes (Park et al. 2009). One major challenge to distinguishing biological echoes is that the shapes of birds and insects are strongly nonspherical (Gauthreaux et al. 2008). Moreover, polarimetric measurements have a strong dependence on their size, shape, and orientation (Gauthreaux et al. 2008; Bridge et al. 2011). Thus, even in the case of single-species ensembles, polarimetric quantities can vary significantly depending on the azimuthal orientation (Stepanian and Horton 2015). This leads to frequently similar features for bird and insect echoes, making it difficult to differentiate them. For example, the Z_{DR} of purple martin colonies have been found to range from -4 to $6\,dB$ (Van Den Broeke 2013) while insects have been found to have Z_{DR} between 2 and $9\,dB$ (Zrnić and Ryzhkov 1998).

Various methods have been explored to detect biological echoes with radars (Drake and Reynolds 2012; Hardy and Katz 1969; Lang et al. 2004; Contreras and Frasier 2008; Melnikov et al. 2012; Gauthreaux and Diehl 2020; Radhakrishna et al. 2019; Kilambi et al. 2018). Much less work has been done in distinguishing bird and insect radar echoes though. Nonpolarimetric radar was used in Martin and Shapiro (2007) to discriminate the echoes by measuring radar cross sections of individual targets observed at close ranges from radar. However, only two cases were examined with this approach. A fuzzy logic algorithm implemented in Jatau and Melnikov (2019) can separate birds and insects for some cases. But the use of Z as an input complicates the resolution of densely aggregated insects and sparse groups of large birds. As a result, postprocessing is required to adjust the classification output.

More recently, machine learning models have been developed for detecting roosting birds in weather radar echoes by leveraging their distinct toroidal shape. Convolutional neural networks were used in Chilson et al. (2019) to detect whether an individual radar image contained at least one purple martin or tree swallow roost with correct determinations made about 90% of the time. Another machine learning system was developed in Cheng et al. (2020) that locates roosts within images and tracks them across frames. Although these methods are useful, they can only detect one orientation of birds while using the entire radar image as an input. They are not able to operate on a gate-by-gate basis and cannot be used in situations where birds are not roosting.

In this study, we propose a machine learning model that is capable of classifying diverse orientations of bird and insect echoes, using information from a single radar range gate instead of an entire radar scan. We start by investigating two machine learning methods: ridge classifier and decision tree. We first collected dual-polarization radar scans containing separate large-scale bird and insect migration (section 2). Next,

blob coloring was used to extract the migration echoes and then their textures were computed (section 3b). Velocity–azimuth display (VAD) is applied to obtain a new azimuth relative to the aspect of birds (insects) in section 3c and multiple bird (insect) scans are averaged to reduce contamination by insects (birds) in section 3d. The averaged scans are used to derive training inputs for the classifiers. The final feature set is described in section 3e. The following sections summarize machine learning methods used (section 4) and the metrics for evaluation (section 5). Both machine learning methods are trained, first on only dual-polarization variables and then on different combinations of the remaining features (section 6) and their performances evaluated using metrics computed on test data (sections 7). Further case studies (section 8) on migration and bird roosting cases are used to analyze performance on new scenarios and to investigate the effect of using different input features. Finally, our recommendations for operational implementation of the selected model are presented in section 9. To the best of our knowledge, this is the first study that applies machine learning to distinguish bird and insect radar echoes.

2. Data collection

a. Selection of bird and insect scans

Finding the ground truth for biological radar data at each individual range gate can be challenging. First, a radar resolution volume is large compared to biological targets. As a result, one volume could contain multiple birds, insects, and even other types of echoes at the same time. Second, biota echoes can cover a large area (hundreds of kilometers), and so it is impossible to verify the exact composition of scatterers within each radar range gate. Our approach was to collect multiple scans of mass bird (insect) migration in clear air (this term is used for radar observations free from precipitation), where we expect to obtain the highest possible number of range gates containing birds (insects). A substantial part of nocturnal echoes in spring and fall are contributed by migrating birds (Zrnić and Ryzhkov 1998). Such migration is characterized by a large area of echoes centered around the radar site with velocities moving in the same direction (highly aligned). Further analysis of birds (purple martins) in Stepanian et al. (2016) has shown that they have modes around 0 dB for Z_{DR} and 110° for Φ_{DP}. On the other hand, insects are routinely observed in clear air during warm seasons in Oklahoma, reaching peak intensity in the late afternoon (Zrnić and Ryzhkov 1998). Analyses have also shown that insects have modes around 6 dB for Z_{DR}, with most echoes saturating at the 8 dB limit of the WSR-88D and Φ_{DP} modes around 70° (Stepanian et al. 2016). It should be noted that both set of echoes show azimuthal dependence for most cases. Furthermore, bird migration echoes typically occupy a larger area compared to insect migration echoes at the 0.5° elevation scan. We used these properties to select 90 clear-air scans in which 45 PPIs are dominated by migrating birds and the other 45 scans are dominated by insects. All scans were collected from KTLX (located near Oklahoma City, Oklahoma) at the 0.5° elevation. We determined a scan to be clear air if few or no precipitation echoes were identified.
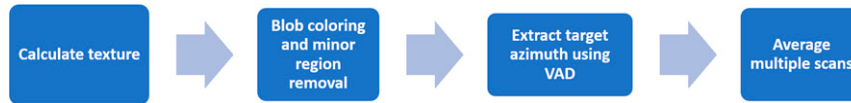
FIG. 1. Flowchart of feature processing.

Each scan was also chosen such that the majority of echoes were due to biological migration activity. This is important for the subsequent extraction of migration echoes by blob coloring. Finally, all gates with range less than 10 km were excluded to reduce contamination by ground clutter.

b. Selection of radar variables for machine learning algorithms

In general, the size of birds is larger than insects. Additionally, since Z depends on the size of targets within the radar resolution volume, it could be used to differentiate bird and insect echoes. However, Z also depends on the number of targets. In other words, a large Z value could be contributed by a large bird or a dense aggregation of insects. Because of this ambiguity in interpretation, Z was excluded from the inputs to the model. Similarly, birds typically move faster than insects. However, biological targets leverage the underlying wind field to aid their flight. As a result, passively flying insects on a windy day could migrate with larger velocities than actively flying birds in a mild wind field. Furthermore, radial projection and aliasing complicates the interpretation of V_r. Thus, V_r is excluded as an input to the model though it is used to recover measurements from the target's aspect. Signal-to-noise ratio from biological scatterers are frequently low for the reliable measurement of spectrum width (σ_V). Due to this high noise contamination, σ_V is also excluded. None of the single-polarization variables are used in training our models.

Dual-polarization variables have been used for the identification of biological echoes (Zrnić and Ryzhkov 1998; Stepanian et al. 2016; Jatau and Melnikov 2019; Gauthreaux and Diehl 2020; Radhakrishna et al. 2019; Kilambi et al. 2018). In this work, they are used in training our model to distinguish between birds and insects. Another factor that motivated their exclusive selection is the volume coverage pattern (VCP) of the WSR-88D. Biological echoes are predominant at the lowest antenna elevation scan of 0.5° in "clear" air. At this elevation, the WSR-88D completes two sweeps, about 30 s apart. The first (surveillance) sweep measures the dual-polarization variables and Z. The second (Doppler) sweep measures the legacy single-polarization variables. Using variables from both sweeps could introduce errors. Assuming a target flies along a radial at a speed of 10 m s^{-1}, it would have migrated 300 m (more than the length of one range gate) between both sweeps. The use of only dual-polarization measurements ensures temporal coherence.

3. Feature processing to prepare inputs

In this section, feature processing (Fig. 1) is performed on the collected bird (insect) scans to prepare inputs for training and evaluating the classifiers. All scans in the dataset are for highly aligned migration cases. First, variables in addition to dual-polarization variables called textures are computed for each scan. Next, blob coloring and minor region removal are used to extract only range gates containing migration echoes in radar scans, followed by VAD analysis to find the direction of migration. The next step is critical. Ideally, we would desire a bird migration scan to be purely composed of bird echoes. However, they usually also contain few insect echoes. Similar scenarios often occur for insect migration scans as well. We propose a way of coherently averaging multiple scans, along range and the target's aspect to improve the purity of our training data.

a. Texture

Many image operations are performed on a local section defined by a window. Such windows are usually described with respect to a reference pixel, where the result of any computation is output. In our case the reference pixel is the middle one. Textures are the result of one of such operations, that characterizes the spatial variation of radar variables in the two-dimensional fields i.e., azimuth and range directions (Chandrasekar et al. 2013; Stepanian et al. 2016). We calculated texture using an eight-connected window, which comprises pixels in the top, left, right, down of the reference pixel, and the four other pixels in the diagonals. Mathematically, at a given radar gate with range r and azimuth angle ϕ, the mean absolute deviation of a variable x from its neighbor gates within the window is calculated as

$$\Delta x_{r,\phi} = \frac{1}{N-1} \sum_{i=-1}^{1} \sum_{j=-1}^{1} |x_{r,\phi} - x_{r+i,\phi+j}|, \tag{1}$$

where i is the range gate offset, j is the azimuth offset, and N is the window size. Calculations were performed only at gates that contained values in all neighbors.

b. Blob coloring and minor region removal to extract migration echoes

Blob coloring is an image processing method used to identify connected groups of pixels with the same value (Bovik 2009; Stepanian et al. 2014). It is applied to create a mask indicating migration echoes. Let us define some relevant terms before describing the algorithm. All definitions are with respect to a binary image where a pixel either contains a target (pixel value = 1) or background (pixel value = 0). A region (or blob) is a group of contiguous pixels with the same value. Two types of windows were applied in this study. The first is a four-connected neighborhood which refers to the reference pixel and pixels above, left, right and below this reference. The other window is the eight-connected neighborhood already described in the previous section. Another operation performed is dilation, which involves iterating a window over an image and setting the pixels of interest at each step as the
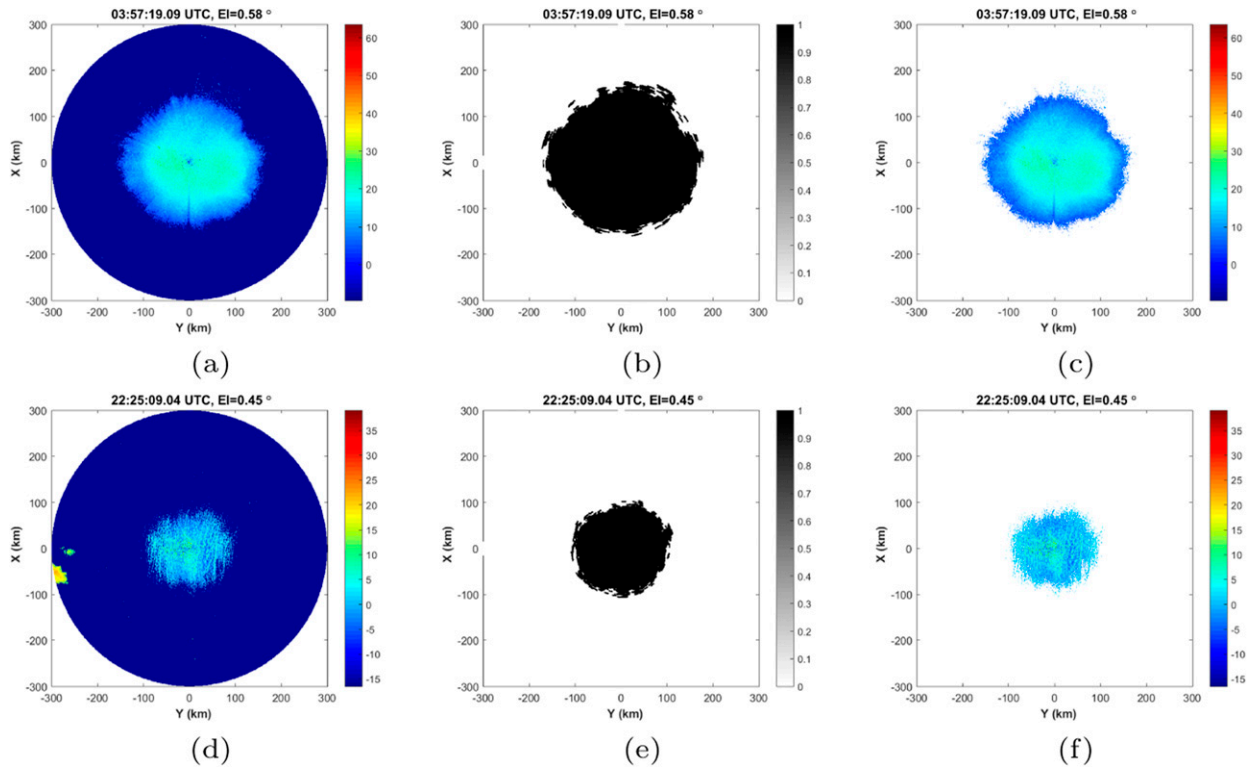
FIG. 2. Blob coloring with minor region removal to extract large-scale migration echoes. (a) Reflectivity (in dBZ) of bird migration echoes, (b) mask of bird migration echoes, (c) gates containing birds extracted using mask, (d) reflectivity (in dBZ) of insect echoes, (e) mask of insect echoes, and (f) gates containing insects extracted using mask.

OR of all pixels within the window. The result of this operation is that regions of target pixels are expanded.

Data for bird and insect migration were collected for clear-air days, which are characterized by a large area of biological echoes centered around the radar and usually against a clear background. An example for bird migration Z is shown in Fig. 2a within a maximum range of approximately 150 km. Note that Z is only chosen for demonstration, it is not used at any other point of this study. The radar scan can be considered as an image I, where rows correspond to ranges, columns correspond to azimuth angles, and range gates correspond to pixels. The blob coloring with minor region removal algorithm is implemented as follows. First, the radar image I is binarized by setting all gates containing echoes to 1 while the remaining gates are set to 0. Usually, migration echoes farthest from the radar are disconnected from other echoes close to the radar. The second step involves dilating the binarized image twice to connect the fringe echoes to the main blob. The dilated image J is given as

$$J = (I \oplus B) \oplus B, \qquad (2)$$

where $\oplus$ is the dilation operator and B is the eight-connected window. In the third step, a region labeling algorithm (Bovik 2009) is applied to identify the different target regions in J. The size of all identified regions are then computed as the number of pixels they contain. Next, minor region removal (Bovik 2009) is applied, where the largest target region is kept and all other target regions are set to background. The result is a single

major region of target echoes. Oftentimes, this major region contains few isolated holes of background pixels. These holes are plugged, by complementing the image, repeating the blob coloring with minor region removal algorithm and recomplementing the image (Bovik 2009). The resulting mask M is a binary image containing one solid target blob (shown in Fig. 2b) corresponding to the region containing migration echoes. The final image K (Fig. 2c) is extracted by the element wise multiplication of the map M and the original image I. This is expressed as

$$K = I \odot M, \qquad (3)$$

where $\odot$ represents the multiplication operation. This image would contain the migrating birds. All echoes within 10 km are excluded to reduce ground clutter contamination. The remaining echoes are collected as bird samples. The same procedure is repeated for insect cases. Figure 2d shows Z for insects with a minor precipitation region west of the radar. The generated map excludes this minor region (Fig. 2e). The final extracted echoes would contain insects (Fig. 2f) and those beyond 10 km are collected as insect samples.

c. Reference with respect to target azimuth

We defined aspect as the orientation of the target's body with respect to the radar beam. Because of the nonspherical shape of biological targets, their radar returns would depend on the aspect observed. As such, methods for identifying biological echoes will have to account for this dependence. Cases of wide spread

alignment can leverage traditional VAD (Doviak and Zrnić 1993; Browning and Wexler 1968; Gauthreaux and Belser 1998) or azimuthal patterns in the correlation coefficient (Stepanian and Horton 2015) to recover aspect information. We used VAD to rotate the variables, so they become a function of their aspect azimuth (ϕ_{aspect}). First, a sinusoid model is fit to V_r at every range,

$$\widehat{V}_r(\phi) = |V| \cos(2\pi f\phi + \delta), \qquad (4)$$

where $\widehat{V}_r$ is the fitted radial velocity, ϕ is the radar's azimuth, $|V|$ is the magnitude of velocity along the migration direction, f is frequency, and δ is a phase offset (in degrees). It is assumed that the wind field is uniform at every height so $f \approx 1/360$ cycles per degree. The migration direction is defined as where targets move toward. It is obtained as the radar azimuth that maximizes $\widehat{V}_r$,

$$\phi_{migration} = \arg\max_\phi \widehat{V}_r(\phi). \qquad (5)$$

This direction captures measurements from the tail aspect. Scattering from other azimuthal aspects can be deduced by the lag from $\phi_{migration}$ as

$$\phi_{aspect} = \phi - \phi_{migration} \qquad (6)$$

such that $\phi_{aspect} = 0°$ represents the tail region of biota, $\phi_{aspect} = 90°$ represents the left wing region, and $\phi_{aspect} = 180°$ represents the head region.

An example for this procedure is shown in Fig. 3. Figure 3a shows the VAD at range 70 km for one of the scans in the training set. The blue line is the filtered velocity obtained by applying a tenth-order one-dimensional median filter on V_r. The green line is the fitted $\widehat{V}_r$. Migration was found to be toward 13.73°. The radial velocity with respect to the target $V_r(\phi_{aspect})$, shown in Fig. 3b, is obtained by shifting $V_r(\phi)$ to the left by 13.73°. Migration would be toward ϕ_{aspect} of 0°. This process is applied at every range ring to find the migration direction and rotate all dual-polarization measurements and their textures. All measurements are now relative to the aspects of the targets.

d. Averaging bird and insect cases

To reduce the contamination of bird migration cases by insect echoes and vice versa, multiple scans are averaged. Following blob coloring and rotation of the collected scans and their textures, they are grouped into three batches. Let us call them batches A, B, and C. Each batch contains 15 randomly selected scans per class (a total of 30 scans per batch). We will focus on A though all steps discussed equally apply to B. Each scan will have different azimuths, so we created a new range and aspect azimuth grid both starting at 0 and with resolutions of 250 m and 0.5°, respectively. All scans were interpolated to this common grid. The new 15 scans for birds (insects) are then averaged. In the last step, all range gates in the resulting averaged scans from A and B are combined to form the training set, containing 1 711 624 samples: 57% bird and 43% insect cases. Batch C is used as the test set. It is not averaged so that it represents the kind of measurements we expect from the WSR-88Ds. The test set contained 9 402 821 range gates with 60% bird and 40% insect cases.

Some visualizations of the averaged training cases are shown in Fig. 4. The blue curve is for birds and the red for insects. Each plot
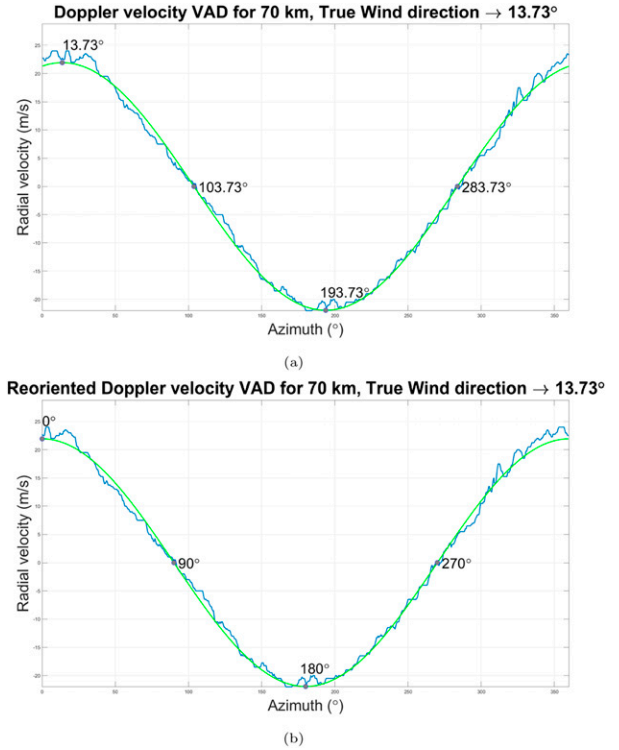

FIG. 3. VAD analysis to reorient velocity to be relative to the target aspect for 70-km range gates. The blue line shows filtered V_r while the green line is the sine fit. (a) Initial VAD finds targets to be migrating toward 13.73° and (b) reoriented fit shifts measurements so migration is toward 0°.

in Fig. 4a–c is for a dual polarization variable against the target aspect at specific ranges. From top to bottom, rows correspond to measurements 15, 30, and 45 km from the radar. From left to right, columns correspond to Z_{DR}, Φ_{DP}, and ρ_{HV}, respectively. The averaging procedure shows that dual-polarization variables have a strong dependence on ϕ_{aspect} and exposes clear delineations between birds and insects. The distribution of the whole training set is also shown in Figs. 4d–f for Z_{DR}, Φ_{DP}, and ρ_{HV}, respectively. These results are consistent with previous literature. Analysis in Van Den Broeke (2013) found that echoes attributed to birds (purple martins) had Z_{DR} between −4 and 6 dB. In our case, the averaged Z_{DR} (shown in Figs. 4a,d) for birds is generally low, between −2 and 4 dB. The highest value is around 230° (between the head and right wing) and the lowest around 75° (between the tail and left wing). Insects were found to have high Z_{DR} (up to 10 dB) in Zrnić and Ryzhkov (1998). Our averaged insect Z_{DR} is also generally higher with most gates between 3 and 8 dB. Interestingly, the values dip below the bird Z_{DR} values between ϕ_{aspect} of 230° and 300°. The value of Φ_{DP} (Fig. 4b) for birds are generally higher than for insects, with peaks around 50° and 300°. This is consistent with the observed symmetry of Φ_{DP} about the direction of migration (Zrnić and Ryzhkov 1998). Insects have lower Φ_{DP} values (Fig. 4e). ρ_{HV} for bird migration have been observed to have low values corresponding to tail-on viewing angles and high values for head-on angles (Stepanian and Horton 2015; Van Den Broeke 2013). This can be seen in the sinusoid-like
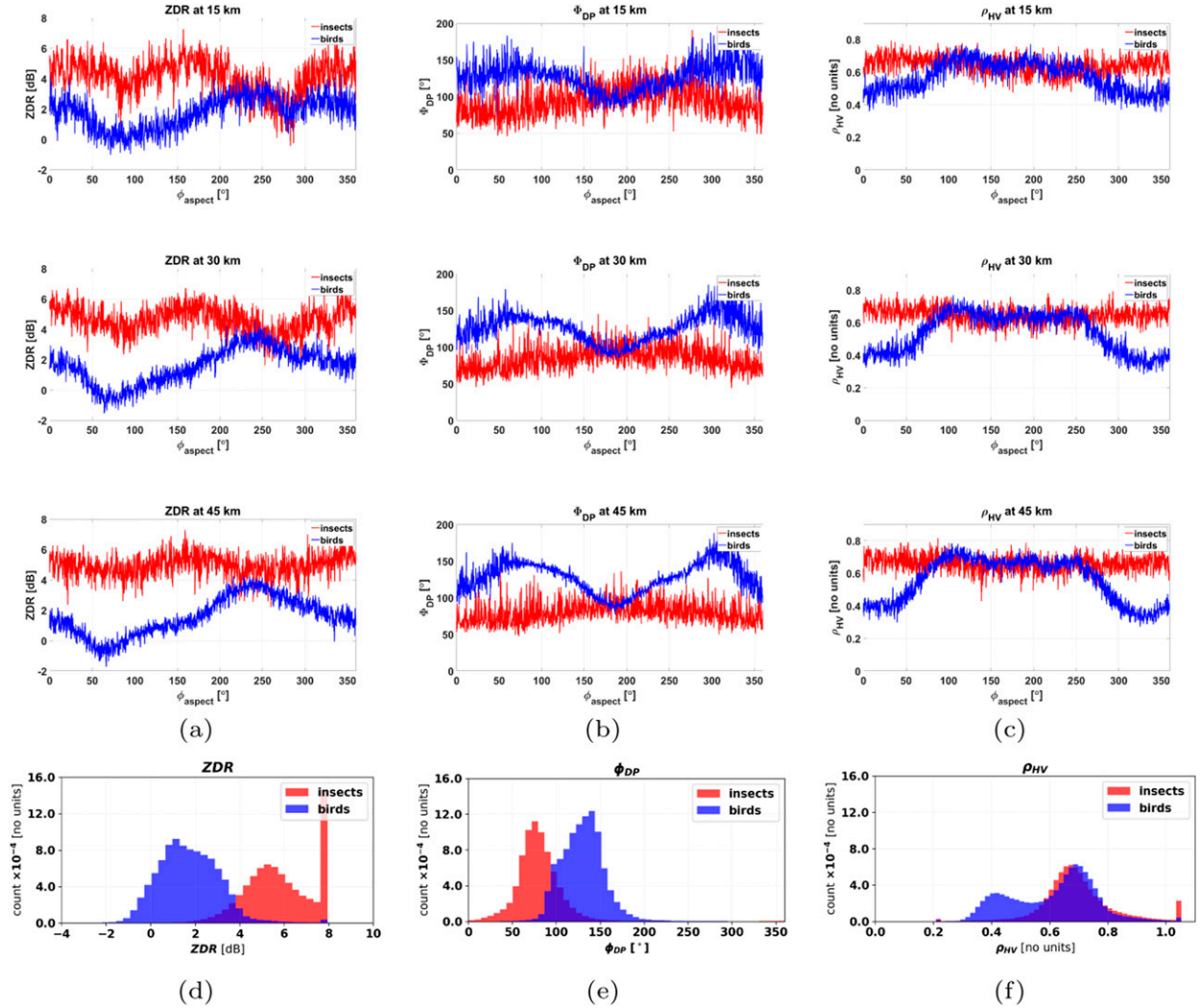
FIG. 4. Averaged dual-polarization variables as a function of ϕ_{aspect} for the training set. (a) Z_{DR} vs ϕ_{aspect}, (b) Φ_{DP} vs ϕ_{aspect}, and (c) ρ_{HV} vs ϕ_{aspect}. Birds are in blue and insects are in red. Rows represent ranges of 15, 30, and 45 km from the radar. Training set distributions of averaged (d) Z_{DR}, (e) Φ_{DP}, and (f) ρ_{HV} for birds and insects.

pattern in Fig. 4c, with high values (around 0.7) between 60° and 250° and low values (around 0.4) otherwise. The bimodal ρ_{HV} distribution for birds (Fig. 4f) is also due to this sinusoidal pattern. Insects generally have slightly higher ρ_{HV} than birds with a mean value around 0.7.

After the averaging procedure, both the training and test datasets are normalized. The mean and standard deviation for each variable was computed from the 60 scans in batches A and B. They are used to normalize each variable by mean centering and scaling by their standard deviation. This ensures that all variables are on the same scale. The same procedure was applied to normalize their textures.

e. Input features for classifiers

The normalized dual-polarization variables and their textures, are used as input features for the classifiers. Additionally, inspection of the data revealed that variables varied gradually with range and ϕ_{az}. Thus, two new discrete features were created to capture this variation. The first is range interval, which refers to 10 km bins. The second is sector, which refers to 20° sectors computed from ϕ_{az}. So, for example, a range gate at 25 km and with ϕ_{az} of 85° would belong to the third range interval and the fifth sector. Both features are one hot encoded to convert them to indicator vectors, as shown below:

$$\text{range interval}_i = \begin{cases} 1, & \text{if} \quad i = \lfloor \text{range}/10 \rfloor \\ 0, & \text{otherwise} \end{cases} \quad (7)$$

$$\text{sector}_j = \begin{cases} 1, & \text{if} \quad j = \lfloor \phi_{az}/20 \rfloor \\ 0, & \text{otherwise}. \end{cases} \quad (8)$$

All echoes collected in this study was from 10 to 230 km, so range interval would contain 22 elements. The first element is

ignored since we did not consider ranges below 10 km. Sector would contain 18 elements.

4. Machine learning methods

Our goal was to train an algorithm for distinguishing bird from insect echoes, that could be implemented operationally on NEXRAD. Traditionally, fuzzy logic has been used for classifying weather radar echoes. However, we opted for a supervised machine learning (ML) approach because models trained this way can easily be updated as new data are collected. These ML methods learn a model that minimizes prediction errors on the training data. As such, they generally improve as more data are provided. This is not necessarily the case with fuzzy logic where the classification rules and membership functions are manually defined. Consequently, an established ML pipeline would allow future research efforts to be focused on much needed data collection and verification. Furthermore, the ML methods provide probabilities of a range gate being the target class.

More complex neural networks have been successfully applied to detect (Chilson et al. 2019) and track roosting birds (Cheng et al. 2020); however, there are some drawbacks. First, they were not designed to make classification on a single radar range gate and rather use a rendered image of a full radar scan as input. Second, they are trained to specifically detect birds engaging in roosting behavior. As such, these networks cannot be generalized to other patterns of bird activity or types of biological echoes. A more robust algorithm for differentiating biological echoes is needed.

In this study, we investigate two supervised ML algorithms for distinguishing birds and insects that are (i) capable of using inputs from a single range gate and as such, adaptive to different orientations of biota; (ii) able to provide a probability that a range gate contains birds (or insects); and (iii) easy to retrain as more data are collected. These algorithms are the ridge classifier and decision tree. Furthermore, they form the basis of more sophisticated neural networks and random forests. Therefore, with understanding of their performance, complexity can be added as required. This way we keep the models as useful and as simple as possible.

The ridge classifier learns a linear function of input variables that achieves the best separation between classes in the output. We used the SGDClassifier (Pedregosa et al. 2011) in scikit-learn. Let us consider a model with only dual-polarization variables as inputs (could also have range interval and sector inputs). For a single range gate, the function is given as

$$f(w,b) = \begin{vmatrix} w_{Z_{DR}} & w_{\Phi_{DP}} & w_{\rho_{HV}} \end{vmatrix} \begin{vmatrix} Z_{DR} \\ \Phi_{DP} \\ \rho_{HV} \end{vmatrix} + b, \qquad (9)$$

where $w_{Z_{DR}}$, $w_{\Phi_{DP}}$, and $w_{\rho_{HV}}$ are the weights for Z_{DR}, Φ_{DP}, and ρ_{HV}, respectively, and b is a bias term. It can also be expressed as

$$f(\mathbf{x}) = \mathbf{w}^T \mathbf{x} + b, \qquad (10)$$

where $\mathbf{w}$ is the weight vector and $\mathbf{x}$ is the input vector. The goal is to find parameters that minimizes the log loss error given by

$$L[y_i, f(x_i)] = \log\{1 + \exp[-y_i f(x_i)]\}, \qquad (11)$$
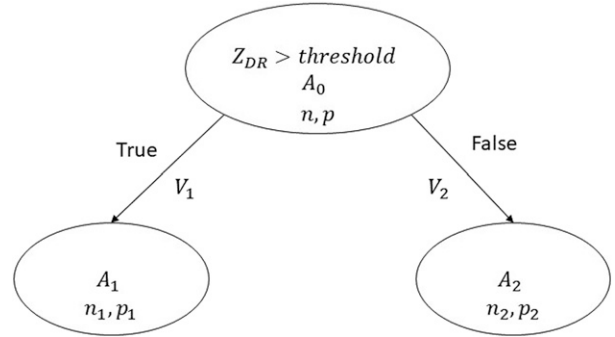

FIG. 5. Part of decision tree, containing three attributes.

where y_i is the label for each training example. A scaled L2 norm of the weights is also added to the above loss. It helps to stabilize learning by penalizing explosion of the weights. The final loss function is given by

$$C(w,b,\alpha) = \frac{1}{n}\sum_{i=1}^{n} L[y_i, f(x_i)] + \alpha||w||_2^2, \qquad (12)$$

where n is the number of training examples and α controls the effect of the weight penalty. The learning process uses stochastic gradient descent (Bottou 2010) on w and b, and a search on α to find values that minimize $C(w,b,\alpha)$.

Our second technique, decision trees, learn rules to recursively partition data so that samples with the same labels are grouped together. We used the DecisionTreeClassifier (Pedregosa et al. 2011) from scikit-learn. A part of a decision tree is shown in Fig. 5. Let us define some terms before discussing how decision trees learn. An attribute A is a question asked about the data; e.g., is $Z_{DR} > $ threshold? Answers to this question, like true or false, are called values V_k and are used to partition the dataset. There are also two classes c_j containing p positive examples (birds) and n negative examples (insects). The entropy of an attribute measures its homogeneity. It is defined as

$$E[p(c_j),\dots,p(c_m)] = \sum_{j=1}^{m} -p(c_j)\log_2 p(c_j), \qquad (13)$$

where $p(c_j)$ is proportion of the jth class. High entropy indicates a uniform distribution over classes while low entropy indicates the dominance of some classes. Information gain measures the reduction in entropy for a given split. It is defined as

$$G(A) = E[p(c_j),\dots,p(c_m)] - \sum_{k=1}^{l} \frac{n_k + p_k}{n+p} E[p(c_j),\dots,p(c_m)|V_k], \qquad (14)$$

where n_k and p_k are the number of positive and negative examples, respectively, in the kth split. In other words, $G(A)$ is the difference between the entropy before a split and the mean entropy after the split. Decision trees learn by finding attributes that maximizes $G(A)$. They can also be thought of as a graph, comprising nodes and edges. Nodes in a tree can be any/both of two types: parent nodes which are attributes, and children nodes, the offspring of parent nodes. Terminal nodes are exclusively children nodes. We applied minimal cost complexity pruning

(Breiman et al. 1984) which regularizes the tree by penalizing the number of terminal nodes.

5. Metrics

We used four metrics to assess the quality of our classifiers. They are accuracy (ACC), true positive rate (TPR), true negative rate (TNR), and area under curve (AUC). Table 1 below shows the confusion matrix for our classification problem. Birds are used as the positive class, so true positives (TP) are birds that are correctly classified as birds, false positives (FP) are birds classified as insects, false negatives (FN) are insects classified as birds, and true negatives (TN) are insects correctly classified as insects. Each instance corresponds to a range gate.

ACC is perhaps the most popular metric for evaluating classifiers. It is the proportion of the total correct predictions. However, ACC can be biased by the distribution of classes. Assuming a dataset contains 90% positive and 10% negative examples, a naive classifier that simply predicts all positives will have an ACC of 90%, though it would not be useful for making any future predictions. Other metrics are needed to better evaluate classifier performance. TPR is the proportion of correct predictions for the positive class and TNR the proportion of correct predictions for the negative class. Our naive classifier would have a TPR of 100% but a TNR of 0% (bad). Ideally all three metrics should be 100%. They are calculated as shown in Eqs. (13)–(15) below. Binary classifiers usually predict a probability (or score) for the positive class and then a threshold is applied to obtain the final class. The receiver operating characteristics (ROC) curve plots TPR against the false positive rate (FPR) for varying probability thresholds (Fawcett 2006). FPR is 1 − TPR. At a threshold of 0, TPR is 100% (good) and FPR is 100% (bad). At a threshold of 1, both are 0%. The goal of the ROC curve is to find an intermediate threshold that maximizes TPR and minimizes TNR. The AUC metric summarizes the area under the ROC curve (Fawcett 2006). Good classifiers should have an AUC close to 100%:

$$ACC = \frac{TP + TN}{TP + FN + FP + FN}, \tag{15}$$

$$TPR = \frac{TP}{TP + FN}, \tag{16}$$

$$TNR = \frac{TN}{FP + TN}. \tag{17}$$

6. Model training and validation

Classifiers are sensitive to the class distribution of the training set. Thus, we applied class weights (Pedregosa et al. 2011) to samples to balance their effect on the loss function. They are computed as

$$w(c_j) = \frac{\sum_{j}^{m} N(c_j)}{m \times N(c_j)}, \tag{18}$$

where $N(c_j)$ is the number of examples in class c_j and m is the total number of classes. For each machine learning method, 8

TABLE 1. Confusion matrix for bird detection.

Classifier output	True labels	
	Birds	Insects
Birds	TP	FP
Insects	FN	TN

models are trained using different combinations of inputs. First, a base model is trained on only dual-polarization variables and then different combinations of the remaining features are added to investigate their effect on performance. It should also be noted that not all inputs features can always be obtained from the radar scan. For example, sector is calculated using a sinusoid fit to the velocity of migration echoes. These echoes are mostly composed of a single species moving in a particular direction. In cases with diverse species moving in different directions, the sinusoid fit will not be possible and the sector cannot be recovered. Velocity aliasing could also prevent the recovery of sector. Similarly, the calculation of texture for a given gate depends on the presence of targets in all neighboring gates. In an alternative scenario, this calculation will not be possible. In such cases, however, the base model can always be used.

K-fold cross validation (Kohavi 1995) was used to tune the model hyperparameters. In this method, the dataset is divided into K folds, model training is performed on K − 2 folds, validation on 1 fold and testing on 1 fold. Since we already held out a test set (batch C), training was performed on K − 1 folds and validation on 1 fold. The whole process is repeated K times where each fold is used as training and validation once. A total of five folds were used. After cross validation, the hyperparameters that have the best performance are chosen for each model. Final training is performed using the selected hyperparameters and the full training set. An ROC curve is then generated and a critical threshold found, such that it maximizes TPR and TNR. This threshold would be used to convert predictions into classes. The training process is stochastic, so each run produces slightly different results. To have a robust assessment of performance, 30 independent training runs are repeated for each model. All the trained models are then evaluated on the test data. Confidence intervals for each metric is calculated using the bootstrapping percentile method where each metric is computed from an iteratively chosen random sample of the test data (Chilson et al. 2019; Efron and Tibshirani 1986). We computed each metric for 100 iterations based on 1000 randomly chosen samples. The 100 metrics for 30 repeated runs form a total of 3000 estimates. The confidence interval is found as the 2.5% and the 97.5% points of the distribution (Chilson et al. 2019; Efron and Tibshirani 1986).

7. Performance

The 95% confidence interval for the model metrics are shown in Table 2. All the ridge classifiers are predictive with ACC > 0.81, TPR > 0.82, TNR > 0.77, and AUC > 0.86. Sector is expected to greatly improve results; however, its addition to ridge classifiers cause marginal changes to performance. It slightly improves TNR, slightly reduces TPR and

TABLE 2. The 95% confidence interval for the ridge classifiers' and decision trees' metrics on migration data. The possible inputs are dual polarization (DP), their textures (ΔDP), sector (sect), and range interval (RI).

	ACC	TPR	TNR	AUC
Ridge classifiers				
DP	0.814–0.858	0.832–0.892	0.778–0.844	0.868–0.911
DP + ΔDP	0.849–0.891	0.874–0.924	0.808–0.874	0.909–0.943
DP + sect	0.812–0.858	0.822–0.886	0.784–0.85	0.869–0.911
DP + RI	0.815–0.86	0.836–0.894	0.774–0.844	0.869–0.912
DP + ΔDP + sect	0.849–0.891	0.87–0.924	0.81–0.874	0.91–0.943
DP + sect + RI	0.813–0.858	0.826–0.884	0.78–0.848	0.869–0.912
DP + ΔDP + RI	0.849–0.891	0.872–0.926	0.81–0.872	0.91–0.944
DP + ΔDP + sect + RI	0.85–0.891	0.87–0.922	0.812–0.876	0.91–0.944
Decision trees				
DP	0.8–0.856	0.786–0.892	0.762–0.852	0.872–0.92
DP + ΔDP	0.778–0.852	0.786–0.902	0.718–0.846	0.866–0.925
DP + sect	0.751–0.815	0.658–0.8	0.79–0.88	0.831–0.884
DP + RI	0.75–0.82	0.668–0.794	0.79–0.882	0.792–0.868
DP + ΔDP + sect	0.713–0.804	0.628–0.79	0.768–0.866	0.793–0.873
DP + sect + RI	0.684–0.762	0.518–0.686	0.814–0.894	0.711–0.796
DP + ΔDP + RI	0.714–0.813	0.7–0.828	0.676–0.844	0.751–0.871
DP + ΔDP + sect + RI	0.669–0.796	0.572–0.732	0.71–0.87	0.702–0.832

does not seem to have a noticeable effect on AUC. To understand the role of sector, let us consider classification using Z_{DR} from a range gate 30 km from the radar (Fig. 4). If this gate has $Z_{DR} > 4$ dB, it should be classified as insects. If $Z_{DR} < 0.5$ dB the echoes should classified as birds. However, Z_{DR} values within 0.5–4 dB can belong to any class. In this scenario sector can further differentiate between the classes. If the undetermined gate falls in the 80°–100° aspect (or sector 5), then it should be classified as insects. But if it falls in the 220°–240° aspect (sector 12), Z_{DR} cannot accurately detect the correct class because the values of both classes overlap. Observing the 15 km range gates in Fig. 4, bird and insect echoes can easily be separated by at least one variable across all aspects except around 200°–250° where all variables overlap. Similar to the previous example, sector would not improve classification in this region. This could be the reason for the marginal results due to sector. However, it should be noted that the clear delineation between both classes at most aspects, so that classification can be effectively performed without sector, is due to the previous averaging along aspect angles. Addition of range interval marginally changes performance, improving ACC, TPR, and AUC and reducing TNR. The addition of texture generally improves the model metrics.

The decision tree models are also predictive for TPR, TNR, and AUC but perform significantly worse on TPR with some models having values around 0.5. This seems to coincide with models using range interval and/or sector. A possible cause could be that its binary decision making tends to prefer classifying whole range intervals (or sectors) as one class in contrast to ridge classifiers that only learns a probability adjustment. Using smaller range intervals and sectors might mitigate this problem. Similar to the ridge classifiers, incorporation of texture generally improves the model metrics. However, this might not generalize to nonmigration cases. Recall that labels were provided based on the dominant migrating taxa. Textures have the effect of averaging measurements derived over a 3 × 3 neighborhood, so

would emphasize the dominant class leading to better metrics for migration cases. However, for nonmigration cases with a heterogeneous mix of scatterers, texture could lead to misclassifications.

An independent two-sample t test is conducted to find the better performing machine learning method. Each metric distribution for equivalent versions of ridge classifiers and decision trees were tested with a significance level of 0.05. The null hypothesis is that both metric distributions have the same mean. All tests were significant ($p < 0.05$). The final results of our t test is shown in Table 3. The cells, labeled R (for ridge classifiers) and D (for decision trees) indicate which model was found to be better for a metric. Our criteria for a method to be better was that it should have higher means on ≥ 3 metrics. The ridge classifiers satisfy this criteria for all input combinations and was chosen as the better method. The best ridge classifier uses dual-polarization variables, texture, sector, and range interval as inputs with AUC > 0.91. It is possible though, that the improvement caused by the added features could be the classifiers overfitting to migration cases. Thus, additional studies of

TABLE 3. Model comparison results. Each cell is labeled R (ridge) or D (decision tree) depending on which model performs better.

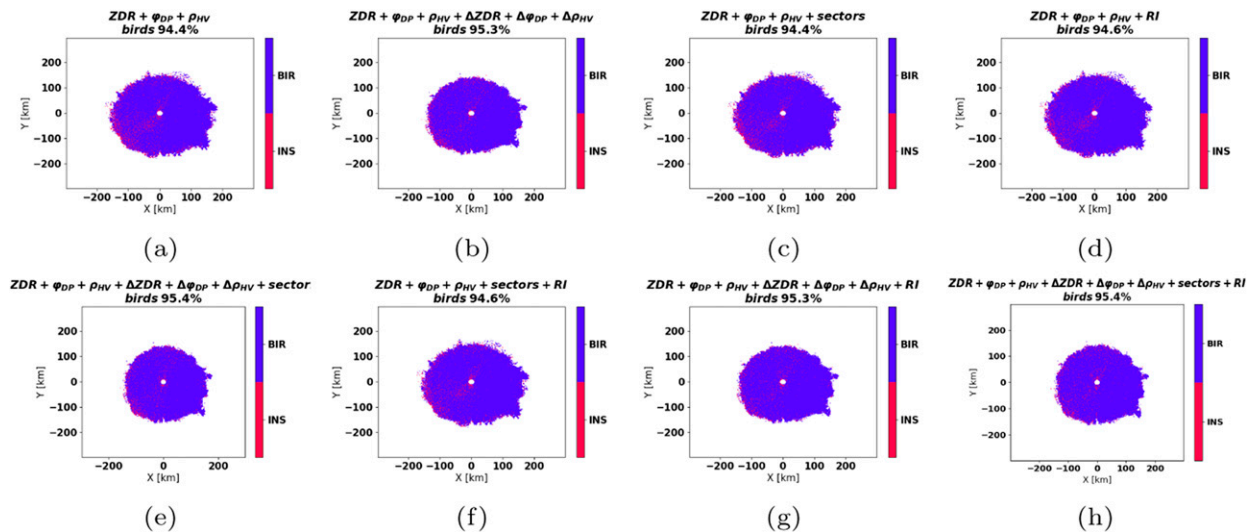|	ACC	TPR	TNR	AUC
DP	R	R	R	D
DP + ΔDP	R	R	R	R
DP + sect	R	R	D	R
DP + RI	R	R	D	R
DP + ΔDP + sect	R	R	R	R
DP + sect + RI	R	R	D	R
DP + ΔDP + RI	R	R	R	R
DP + ΔDP + sect + RI	R	R	R	R

FIG. 6. Ridge classification results for bird migration observed with KTLX radar at 0413 UTC 2 May 2015. BIR represents birds and INS represents insects.

ridge classifiers on a diverse variety of cases (presented in the next section) are required to understand the effect of these features on performance. Similar discussion on decision trees can be found in the online supplementary materials.

8. Case studies and discussion

In this section, the performance of the ridge classifiers are further tested on four cases. The aim of these analysis is to verify that the classifier's detections are consistent with the available ground truth and to compare the effect of different features on the classifier's performance. For operational use, we recommend that the ridge classifiers be used in conjunction with the HCA. So the HCA is applied first and if the HCA classifies a range gate as biological, the ridge classifiers are subsequently applied to classify the gate as containing either bird or insect echoes. This operational configuration is applied to cases in this section, where echoes had not already been identified as biological.

The first case study is for bird and insect migration scans from the test set. This case was chosen to understand the output of the models when all variables are recovered. The second case explores a new bird migration scan where velocities were heavily aliased, so sector could not be recovered. It is also important to investigate the ability of our models to generalize to new situations. One of such scenarios involves the mass exodus of birds from their nighttime roosting locations usually observed as expanding rings of echoes. In the third case, a scan from the KHTX radar (located at Huntsville, Alabama) containing two groups of roosting birds, insects and precipitation echoes was analyzed. This case is especially important because our training data were collected from a different radar (KTLX) and contained no bird roosts. As a result, the ability to detect these roosts and insect echoes will be strong evidence that it can be extended to new patterns of biological activity and to other radars in NEXRAD. It also demonstrates how precipitation

echoes would be handled. The final case also investigates the performance on roosting birds from the KTLX radar.

a. Bird/insect migration test cases

The first case was collected from KTLX at 0413 UTC 2 May 2015 composed of migrating birds. It is one of the PPIs in the test set. Since blob coloring with minor region removal had been applied to extract migration echoes, the HCA is not used in this case. Sector could be recovered because of the homogeneity and alignment of scatterers. The classifier outputs a probability of each range gate containing bird echoes. The critical threshold is applied to binarize these probabilities to 0 (insects) and 1 (birds), obtaining the output class. The ridge classifier outputs are shown in Fig. 6. Birds are colored blue while insects are colored red. Across the 8 ridge classifiers, birds are detected in 94.4%–95.4% of classified gates. This is consistent with our hypothesis of migrating birds. Overall, the performance of the base model barely changes as the remaining features (texture, sector, and range interval) are added, with at most a 1% difference in proportion of birds detected.

The second case is also a test scan collected from KTLX at 1708 UTC 11 July 2019 and composed of migrating insects. Sector was recovered and the HCA was not applied because migration echoes had already been extracted by blob coloring. The ridge classifier outputs are shown in Fig. 7. Across the classifiers, insects are detected in 92.3%–92.7% of classified gates, consistent with our hypothesis of echoes from insects. Like the bird migration case, there is a bare difference of 0.4% in the proportion of insects detected between the base model and those with added features.

b. Bird migration with aliased velocities

The third case was collected from KTLX at 0441 UTC 3 March 2018, composed of nocturnal bird migration during spring. The characteristic symmetry pattern about the migration direction was observed in Φ_{DP} (not shown here) with
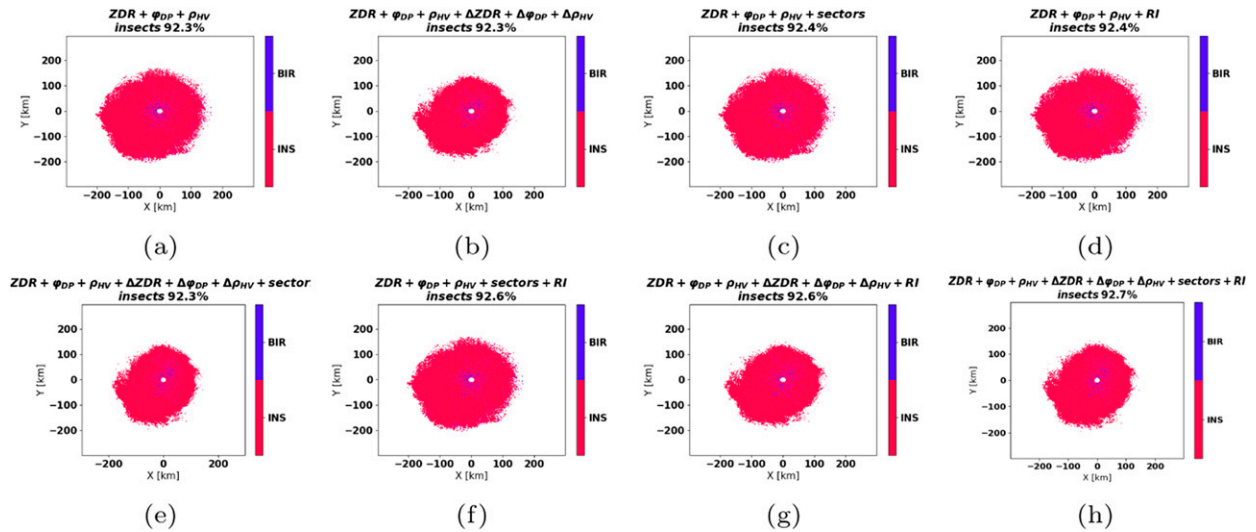
FIG. 7. Ridge classification results for insect migration observed with KTLX radar at 1708 UTC 11 Jul 2019.

values generally larger than 110°. The radial velocities (shown in Fig. 8b) were also highly aligned, with large magnitudes causing aliasing. Furthermore, the bird/insect fuzzy logic classifier in Jatau and Melnikov (2019) detected birds as the dominant taxa for this case. The HCA output is shown in Fig. 8e. All weather classes are condensed into one resulting in

five possible classes which include weather (WEA), anomalous propagation/ground clutter (GC), unknown (UK), biological (BI), and range folded (RH). Most of the echoes are classified as biological with a few gates classified as unknown. We could not recover sector because of heavy velocity aliasing. Our ridge classifiers detects birds as the dominant taxa found in
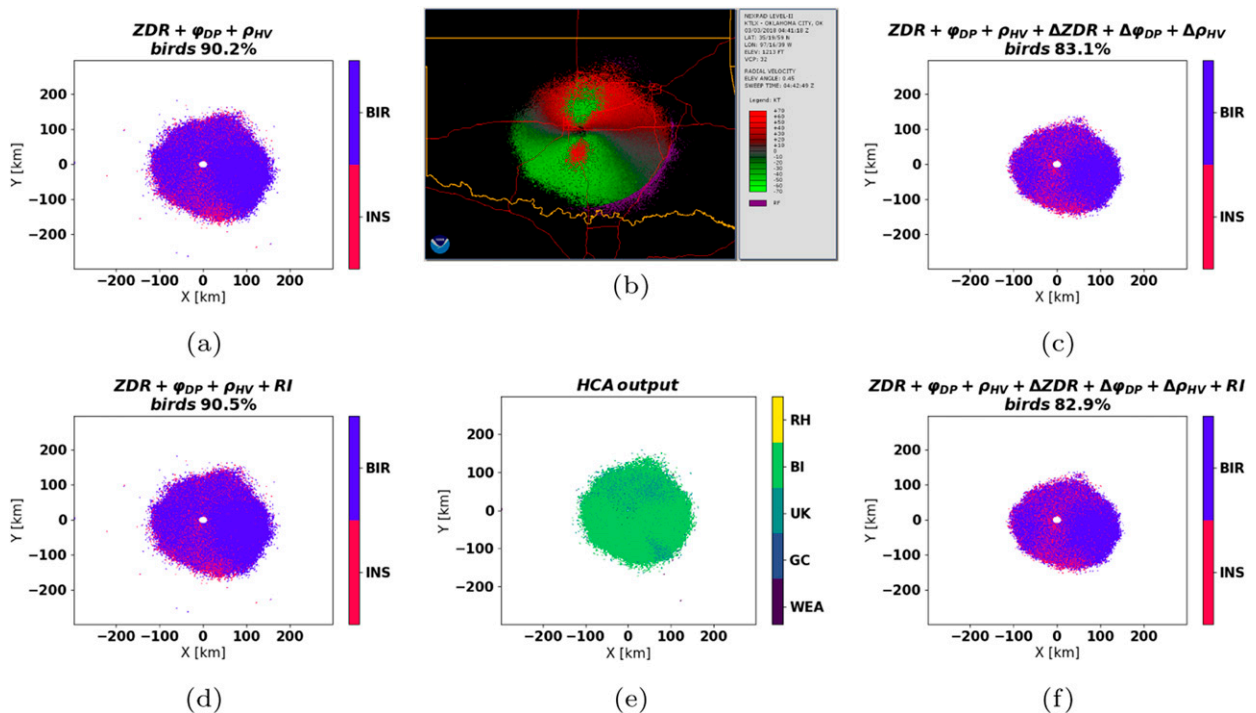

FIG. 8. Bird migration with velocity aliasing observed with KTLX at 0441 UTC 3 Mar 2018. (a) Ridge classifier (RC) using dual-polarization variables, (b) 0.5° V_r shows heavy aliasing north and south of radar, (c) RC using dual-polarization variables and their textures, (d) RC using dual-polarization variables and range interval, (e) hydrometeor classification algorithm (HCA), and (f) RC using dual-polarization variables, their textures, and range interval.
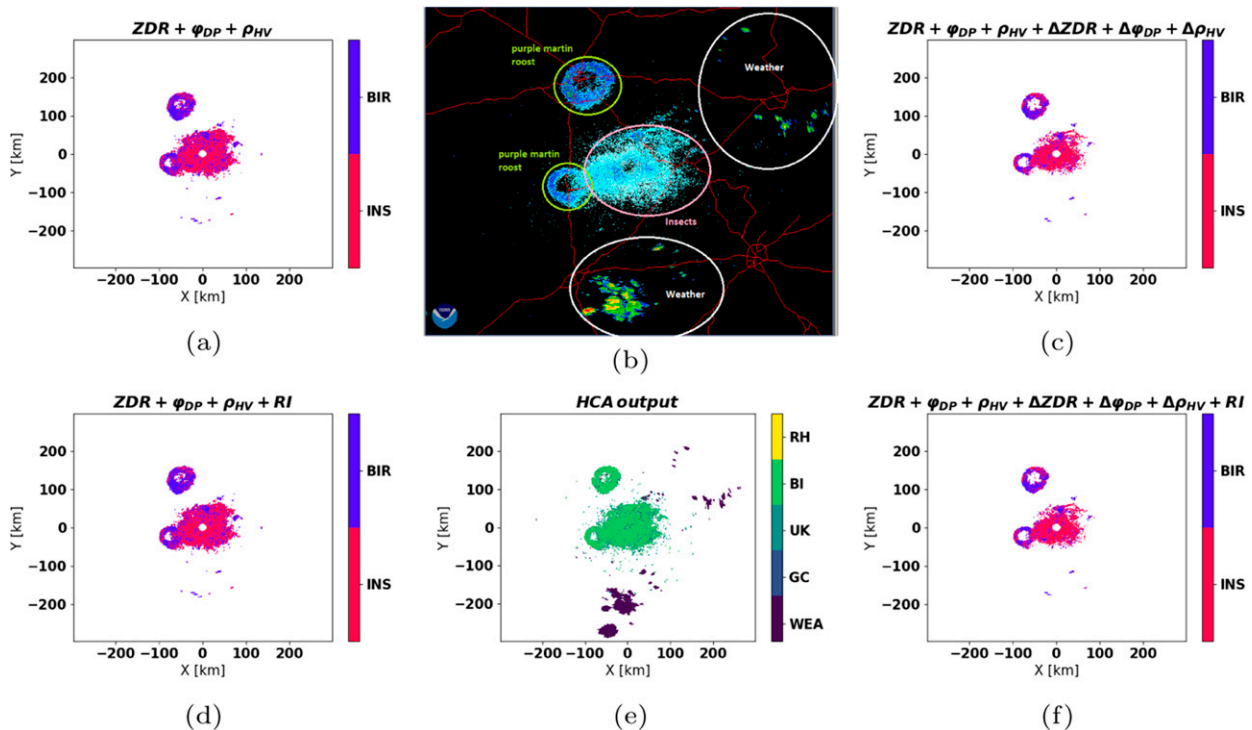
FIG. 9. Bird roosts observed with KHTX at 1115 UTC 11 Aug 2015. (a) RC using dual-polarization variables, (b) 0.5° Z scan showing bird roosts (west and northwest), insects (around the radar), and weather echoes (northeast and south), (c) RC using dual-polarization variables and their textures, (d) RC using dual-polarization variables and range interval, (e) HCA, and (f) RC using dual-polarization variables, their textures, and range interval.

82.9%–90.5% of classified gates. Addition of texture to the base classifier decreases the proportion of birds by about 7%. The addition of range interval barely changes performance, with an increase of 0.29% in the proportion of birds detected.

c. KHTX bird roosts

The third case was collected from KHTX radar (located at Huntsville) at 1115 UTC 11 August 2015. This case demonstrates the performance of the ridge classifiers on bird roosts from a different WSR-88D. The Z scan is shown in Fig. 9b with three groups of echoes, labeled based on analysis conducted in Stepanian et al. (2016). The first group contains two colonies of purple martins engaging in their morning roosts, verified by ground observers from the Purple Martin Conservation Society (Stepanian et al. 2016). The roosts are located northwest and west of the radar. The second group contains insects around the radar, identified by their comparatively low mean airspeed of $1.8\,\mathrm{m\,s^{-1}}$ and concentration at low heights (Stepanian et al. 2016). Airspeed was calculated by vector subtraction of wind speeds (obtained from balloon sounding) from ground speeds obtained from radial velocity (Stepanian and Horton 2015). The third group contains weather echoes identified by their near $0\,\mathrm{dB}$ Z_{DR}, near 1 ρ_{HV} and Φ_{DP} near the calibration offset of $60°$ (Stepanian et al. 2016).

The classifiers are applied as we recommend for operational use. First, the HCA is used to classify range gates into the five condensed classes. The results (shown in Fig. 9e) matches the labels provided in

Stepanian et al. (2016) with the bird roosts and insects identified as biological echoes and the precipitation echoes also correctly identified (seen farther south and northeast of the radar in Fig. 9e). Next, the ridge classifiers are applied on only the biological class. Sector could not be recovered here because of the presence of diverse scatterers with different velocities. The base classifier (Fig. 9a) and the classifier with range interval (Fig. 9d) identify the roosts as bird dominated and the insect region as insect dominated. These results are consistent with the bird and insect groups observed in Stepanian et al. (2016). The models with texture (Figs. 9c,f) similarly identifies the insect region but misclassify large parts of the roosting birds as insects, most noticeable where the western roost intersect with insect echoes.

Overall, the base classifier's detections matches the labels provided demonstrating the potential of this classifier to be applied to new cases of biological activity and different radars within NEXRAD. The addition of range interval does not have a noticeably effect on performance. The addition of texture seems to degrade performance on fine and hollow features like bird roosts.

d. KTLX bird roosts

The final case was collected from KTLX at 1147 UTC 8 August 2017. We identified four bird roosts against an insect background (Fig. 10b) by tracking their expansion across multiple scans. Furthermore, the bird/insect fuzzy logic classifier in Jatau and Melnikov (2019) detected the western and the two southeastern roosts as bird dominated. It was not
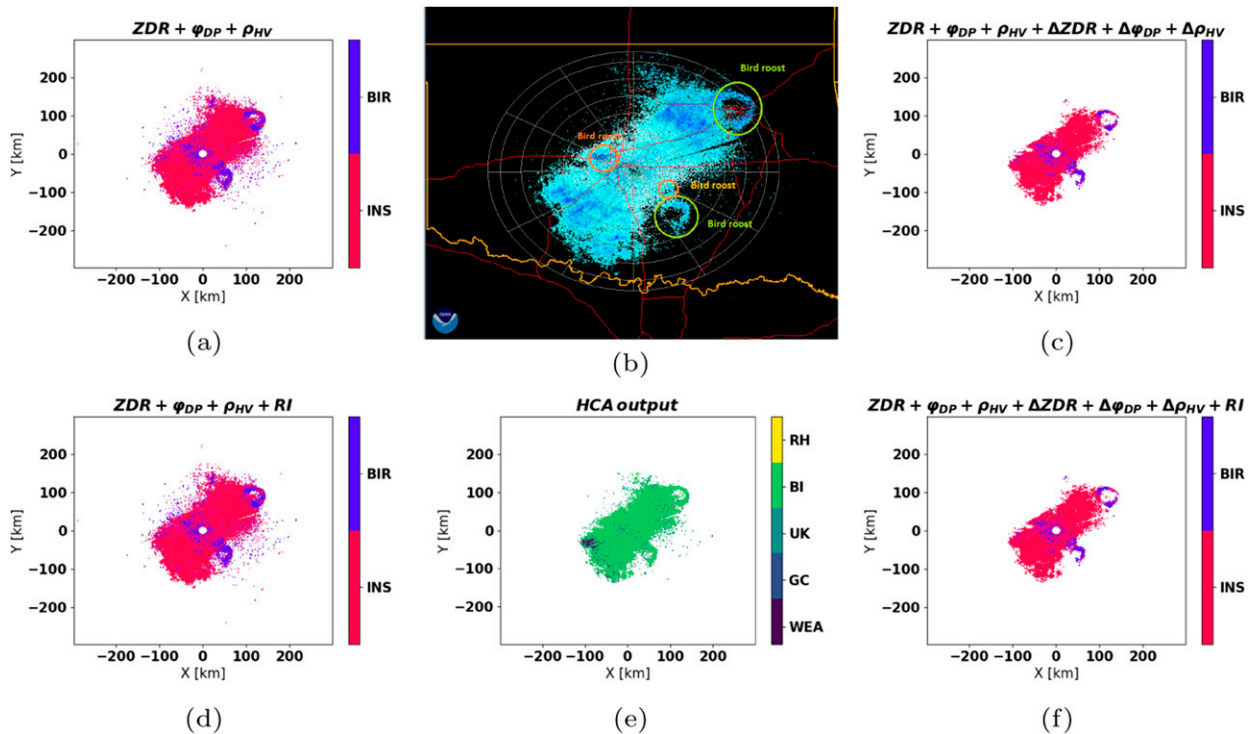
FIG. 10. As in Fig. 9, but for roosts observed with KTLX at 1147 UTC 8 Aug 2017. Four bird roosts were identified northwest, northeast, and southeast of the radar.

tested on the northeastern roost because it has a range limit of 100 km. A major issue with this fuzzy logic classifier is that it has a tendency to classify gates containing dense aggregation of insects as birds because it uses Z as one of the inputs. A correction was made using thresholds on Z_{DR} and ρ_{HV}. However, this also led to other complications due to the variation of biological measurements with their viewing angle.

Since our ridge classifiers are based on dual-polarization variables, they are robust to the concentration of scatterers in a range gate. The HCA (Fig. 10e) detects most gates as composed of biological echoes and the classifiers are applied to these echoes. Sector could not be recovered because of the presence of diverse scatterers. The base classifier (Fig. 10a) and the classifier with range interval (Fig. 10d) detects all the roosts as bird dominated. Addition of texture (Figs. 10c,f) leads to more parts of the roosts being classified as insects. Similar to the previous case, the base classifier's output matches the provided labels. Their performance on bird roosts are also consistent with the results obtained from the bird/insect fuzzy logic classifier in Jatau and Melnikov (2019). Overall, the addition of range interval does not have a noticeably effect on performance while the addition of texture degrades the detection of roosts.

9. Recommendations for operational implementation

Our performance analysis on the test data revealed that the classifier with the full set of features improved each metric from the base classifier by about 4%. However, because the test data were collected from cases of migration, it is possible

that the former is overfitting to one particular scenario and might not extend to new scenarios effectively. Additional case studies in the previous section provide strong evidence that this might be the case, with the base classifier having consistently strong performance across all scenarios. The addition of range interval and sector seems to have a benign effect, probably because they have already been applied to improve the purity of the training data during the averaging step. Texture, on the other hand, seems particularly suited to broadscale migration but has worse performance on finer and hollow features like bird roosts.

Additionally, the base classifier was demonstrated to be effective on new orientations of biological echoes and a different WSR-88D in the KTLX and KHTX bird roost cases. As such, we recommend that the base classifier can be used operationally on NEXRAD. Here, the HCA would be applied to a range gate, and if it is classified as biological echoes, then the base classifier is used to further differentiate into either birds or insects. All inputs for the classifier are available from a single range gate, so no further feature processing steps would be required.

Based on previous literature and the results illustrated in Fig. 4, our intuition is that birds generally have a lower Z_{DR}, higher ϕ_{DP}, and lower ρ_{HV} compared to insects. The weights [refer to Eq. (9)] learned by the base classifier confirms this. They are shown in Fig. 11. For a given input, a positive weight implies that the probability of being a bird increases as the input increases. Negative weights imply that the probability of being an insect increases as the input increases. The weight
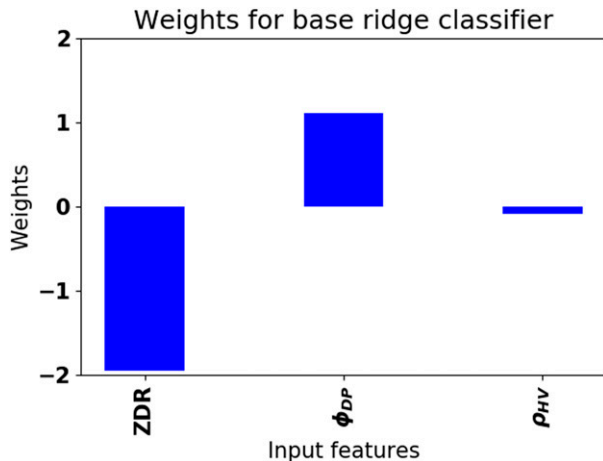
FIG. 11. Weights of base ridge classifier. Positive (negative) weights for a variable imply that the probability of being a bird is directly (inversely) proportional to the variable.

magnitude corresponds to its importance on the classification outcome. Z_{DR} has a weight of -2, meaning that the higher its value, the higher the probability of being an insect. Higher ϕ_{DP} values would increase the probability of being a bird, since it has a weight of $+1$. ρ_{HV} is more subtle with a weight of -0.1, and so higher values would slightly nudge the classifier toward insects. Also notice how the magnitude of the weights are proportional to the degree of separation between both classes in Fig. 4. Z_{DR} seems to be the most important with about twice the effect of ϕ_{DP} and 20 times the effect of ρ_{HV}.

10. Conclusions

The ability of NEXRAD to detect birds and insects offers much promise for meteorological and biological studies. In this work, we developed a classifier for distinguishing bird and insect radar echoes based on dual-polarization variables. Unique challenges were presented by the complex scattering off their nonspherical bodies. We addressed this by leveraging cases of large-scale highly aligned migration. Measurements relative to the body aspect of biota were extracted as the lag off the mean flight direction, obtained using VAD. Another issue is the reliability of ground truth for training datasets because of the frequent collocation of birds and insects in the radar sampling volume, leading to the contamination of scans otherwise dominated by a single species. We addressed this by averaging 15 alignment-calibrated bird (insect) migration scans to reduce the effect of the less dominant class.

Overall, our data preparation for the training and test set involves the following steps. First, 45 migration scans were collected for each class. Blob coloring with minor region removal is then applied to extract migration echoes and textures computed. Extracted migration echoes are rotated to obtain measurements relative to the target's aspect. Next, rotated scans are grouped into 3 batches, each containing 15 scans per class. Within each class in the first two batches, all scans are averaged. Gates from the 4 resulting averaged scans are used

as the training samples. The last batch is used as the test set. All gates are grouped into 10-km range intervals and 20° target azimuth sectors. The final inputs considered are dual-polarization variables, their textures, range intervals, and sectors.

Two machine learning methods were explored: ridge classifier and decision tree. Eight models were trained for each method, starting with a base model using only dual-polarization variables and then adding other features (range interval, texture, and sector). Models were evaluated based on accuracy (ACC), true positive rate (TPR), true negative rate (TNR), and area under curve (AUC), all computed from the test dataset. An independent Student's t test comparison showed that the ridge classifiers performed better than decision trees in ≥ 3 metrics. Based on this, the ridge classifiers were selected for classifying bird and insect radar echoes. All the ridge classifiers are predictive with ACC > 0.81, TPR > 0.82, TNR > 0.77, and AUC > 0.86. The addition of other features generally show an improvement, with about a 4% increase on these metrics; however, later evidence suggests that this is probably due to overfitting to cases of large-scale migration.

Further investigation was conducted to assess the classifiers on more diverse orientation of biota. The first case contained bird and insect migration scans from the test set. They are chosen because all inputs are available. The ridge classifiers detected birds in 94.4%–95.4% of range gates for the bird scan and insects in 92.3%–92.7% of gates for the insect scan, consistent with our hypothesis of the source of these echoes. Overall, the addition of the remaining features to the base model has a limited impact on performance, with an increase of at most 1% and 0.4% in the proportion of birds and insects detected for their respective cases. The next case contained bird migration with heavily aliased velocities, so sector could not be obtained. The ridge classifiers detected a bird majority, with 82.9%–90.5% of gates classified as birds. The base classifier and the one with range interval detected the highest proportion of birds. The addition of texture decrease the proportion of birds detected by about 7%.

All training data were collected from KTLX. Thus, a classifier's ability to detect biological patterns from other WSR-88Ds would provide strong evidence that it can be applied across NEXRAD. Furthermore, the training data contained no bird roosts. The ability to detect these roosts would demonstrate that the classifiers are effective for diverse orientation of biota. The next case explores bird roosts collected by the KHTX (located in Huntsville). Previous studies (Stepanian et al. 2016) identified groups of bird, insect, and weather echoes for this scan. Sector could not be recovered because of the heterogenous mix of scatterers. The base classifier and the one with range interval detects the roosts as bird dominated. Their results also match the groups identified in Stepanian et al. (2016). The addition of texture seems to degrade performance on the roosts, probably because it is less suited for capturing finer features. The final case also contains bird roosts, collected by KTLX. We identified four flocks of roosting birds against a background of insect echoes. Additionally, a bird/insect fuzzy logic classifier developed in Jatau and Melnikov (2019)

identifies three of these roosts (the fourth roost was outside the fuzzy classifier's range limit) as birds. Again, the base classifier and the one with range interval detect all the roosts as birds. The results also match those in Jatau and Melnikov (2019).

Overall, the base classifier shows consistently strong performance across all the scenarios considered. The addition of range interval and sector seems to cause no noticeable improvement, while texture seem to favor detecting broader-scale migration at the expense of finer features. The base classifier was also demonstrated to be effective when applied to new situations involving roosting birds and a different radar (KHTX), showing its potential to be used on the U.S. national network of S-band polarimetric weather radars. As such, we recommend that this classifier can be implemented as an accessory to the HCA. For every range gate, the HCA would first be applied. If the HCA detects biological echoes, the base classifier is then applied to further distinguish the gate as being either bird or insect echoes.

There are many potential improvements to classifying biological echoes. We have broadly looked at birds and insects in this study; however, even within these classes different species can have different scattering characteristics. More work is needed to understand and incorporate this information into a classification model. Ground truth validation remains the biggest challenge. One approach being considered is validating our classifiers using measurements from other sensors. We hope this research also encourages more data collection and verification efforts. We would like to conclude with a note on machine learning. With the advent of its algorithms to the field of weather radars, questions might arise about the role of more traditional techniques (e.g., VAD) for analyzing radar data. We hope this research demonstrates the importance of these techniques for preparing input features to be used in machine learning algorithms.

Data availability statement. The dataset for this project was downloaded from the National Centers for Environmental Information (NCEI).

REFERENCES

Bachmann, S., and D. Zrnić, 2007: Spectral density of polarimetric variables separating biological scatterers in the VAD display. *J. Atmos. Oceanic Technol.*, **24**, 1186–1198, https://doi.org/10.1175/JTECH2043.1.

Bottou, L., 2010: Large-scale machine learning with stochastic gradient descent. *Proc. COMPSTAT*, Paris, France, IASC, https://doi.org/10.1007/978-3-7908-2604-3_16.

Bovik, A. C., 2009: Basic binary image processing. *The Essential Guide to Image Processing*, A. Bovik, Ed., Academic Press, 69–96.

Breiman, L., J. Friedman, R. Olshen, and C. Stone, 1984: *Classification and Regression Trees*. Wadsworth and Brooks, 358 pp.

Bridge, E. S., and Coauthors, 2011: Technology on the move: Recent and forthcoming innovations for tracking migratory birds. *BioScience*, **61**, 689–698, https://doi.org/10.1525/bio.2011.61.9.7.

Browning, K. A., and R. Wexler, 1968: The determination of kinematic properties of a wind field using Doppler radar. *J. Appl. Meteor.*, **7**, 105–113, https://doi.org/10.1175/1520-0450(1968)007<0105:TDOKPO>2.0.CO;2.

Chandrasekar, V., R. Keränen, S. Lim, and D. Moisseev, 2013: Recent advances in classification of observations from dual polarization weather radars. *Atmos. Res.*, **119**, 97–111, https://doi.org/10.1016/j.atmosres.2011.08.014.

Cheng, Z., S. Gabriel, P. Bhambhani, D. Sheldon, S. Maji, A. J. Laughlin, and D. W. Winkler, 2020: Detecting and tracking communal bird roosts in weather radar data. *Proc. AAAI Conf. on Artificial Intelligence*, New York, NY, AAAI, https://doi.org/10.1609/aaai.v34i01.5373.

Chilson, C., K. Avery, A. McGovern, E. Bridge, D. Sheldon, and J. Kelly, 2019: Automated detection of bird roosts using NEXRAD radar data and convolutional neural networks. *Remote Sens. Ecol. Conserv.*, **5**, 20–32, https://doi.org/10.1002/rse2.92.

Contreras, R. F., and S. J. Frasier, 2008: High-resolution observations of insects in the atmospheric boundary layer. *J. Atmos. Oceanic Technol.*, **25**, 2176–2187, https://doi.org/10.1175/2008JTECHA1059.1.

Dokter, A. M., F. Liechti, H. Stark, L. Delobbe, P. Tabary, and I. Holleman, 2011: Bird migration flight altitudes studied by a network of operational weather radars. *J. Roy. Soc. Interface*, **8**, 30–43, https://doi.org/10.1098/rsif.2010.0116.

Doviak, R., and D. Zrnić, 1993: *Doppler Radar and Weather Observations*. Academic Press, 562 pp.

Drake, V. A., and R. A. Farrow, 1988: The influence of atmospheric structure and motions on insect migration. *Annu. Rev. Entomol.*, **33**, 183–210, https://doi.org/10.1146/annurev.en.33.010188.001151.

——, and D. R. Reynolds, 2012: *Radar Entomology: Observing Insect Flight and Migration*. CABI Wallingford, 489 pp.

Efron, B., and R. Tibshirani, 1986: Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Stat. Sci.*, **1**, 54–75, https://doi.org/10.1214/ss/1177013815.

Fawcett, T., 2006: An introduction to ROC analysis. *Pattern Recognit. Lett.*, **27**, 861–874, https://doi.org/10.1016/j.patrec.2005.10.010.

Gauthreaux, J., A. Sidney, J. W. Livingston, and C. G. Belser, 2008: Detection and discrimination of fauna in the aerosphere using Doppler weather surveillance radar. *Integr. Comp. Biol.*, **48**, 12–23, https://doi.org/10.1093/icb/icn021.

Gauthreaux, S. A., Jr., and C. G. Belser, 1998: Displays of bird movements on the WSR-88D: Patterns and quantification. *Wea. Forecasting*, **13**, 453–464, https://doi.org/10.1175/1520-0434(1998)013<0453:DOBMOT>2.0.CO;2.

——, and R. Diehl, 2020: Discrimination of biological scatterers in polarimetric weather radar data: Opportunities and challenges. *Remote Sens.*, **12**, 545, https://doi.org/10.3390/rs12030545.

Hardy, K. R., and I. Katz, 1969: Probing the clear atmosphere with high power, high resolution radars. *Proc. IEEE*, **57**, 468–480, https://doi.org/10.1109/PROC.1969.7001.

Jatau, P., and V. Melnikov, 2019: Classifying bird and insect radar echoes at S band. *35th Conf. on Environmental Information Processing Technologies*, Phoenix, AZ, Amer. Meteor. Soc., 830, https://ams.confex.com/ams/2019Annual/webprogram/Paper351588.html.

Kilambi, A., F. Fabry, and V. Meunier, 2018: A simple and effective method for separating meteorological from nonmeteorological targets using dual-polarization data. *J. Atmos. Oceanic Technol.*, **35**, 1415–1424, https://doi.org/10.1175/JTECH-D-17-0175.1.

Kohavi, R., 1995: A study of cross-validation and bootstrap for accuracy estimation and model selection. *Proc. 14th Int. Joint Conf. on Artificial Intelligence*, San Francisco, CA, AAAI, 1137–1143.

Kumjian, M. R., 2013: Principles and applications of dual-polarization weather radar. Part I: Description of the polarimetric radar variables. *J. Oper. Meteor.*, **1**, 226–242, https://doi.org/10.15191/nwajom.2013.0119.

Lang, T. J., S. A. Rutledge, and J. L. Stith, 2004: Observations of quasi-symmetric echo patterns in clear air with the CSU–CHILL polarimetric radar. *J. Atmos. Oceanic Technol.*, **21**, 1182–1189, https://doi.org/10.1175/1520-0426(2004)021<1182:OOQEPI>2.0.CO;2.

Liu, S., Q. Xu, and P. Zhang, 2005: Identifying doppler velocity contamination caused by migrating birds. Part II: Bayes identification and probability tests. *J. Atmos. Oceanic Technol.*, **22**, 1114–1121, https://doi.org/10.1175/JTECH1758.1.

Martin, W. J., and A. Shapiro, 2007: Discrimination of bird and insect radar echoes in clear air using high-resolution radars. *J. Atmos. Oceanic Technol.*, **24**, 1215–1230, https://doi.org/10.1175/JTECH2038.1.

Melnikov, V. M., R. R. Lee, and N. J. Langlieb, 2012: Resonance effects within S-band in echoes from birds. *IEEE Geosci. Remote Sens. Lett.*, **9**, 413–416, https://doi.org/10.1109/LGRS.2011.2169933.

Park, H. S., A. V. Ryzhkov, D. S. Zrnić, and K.-E. Kim, 2009: The hydrometeor classification algorithm for the polarimetric WSR-88D: Description and application to an MCS. *Wea. Forecasting*, **24**, 730–748, https://doi.org/10.1175/2008WAF2222205.1.

Pedgley, D. E., 1982: *Windborne Pests and Diseases: Meteorology of Airborne Organisms*. Ellis Horwood, 250 pp.

Pedregosa, F., and Coauthors, 2011: Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.*, **12**, 2825–2830.

Peterson, A. T., and R. A. J. Williams, 2008: Risk mapping of highly pathogenic avian influenza distribution and spread. *Ecol. Soc.*, **13**, art15, https://doi.org/10.5751/ES-02532-130215.

Radhakrishna, B., F. Fabry, and A. Kilambi, 2019: Fuzzy logic algorithms to identify birds, precipitation, and ground clutter in S-band radar data using polarimetric and nonpolarimetric variables. *J. Atmos. Oceanic Technol.*, **36**, 2401–2414, https://doi.org/10.1175/JTECH-D-19-0088.1.

Rinehart, R., 2004: *Radar for Meteorologists*. Rinehart Publications, 482 pp.

Stepanian, P. M., and K. G. Horton, 2015: Extracting migrant flight orientation profiles using polarimetric radar. *IEEE Trans. Geosci. Remote Sens.*, **53**, 6518–6528, https://doi.org/10.1109/TGRS.2015.2443131.

——, P. B. Chilson, and J. F. Kelly, 2014: An introduction to radar image processing in ecology. *Methods Ecol. Evol.*, **5**, 730–738, https://doi.org/10.1111/2041-210X.12214.

——, K. G. Horton, V. M. Melnikov, D. S. Zrnić, and S. A. Gauthreaux Jr., 2016: Dual-polarization radar products for biological applications. *Ecosphere*, **7**, e01539, https://doi.org/10.1002/ecs2.1539.

Van Den Broeke, M. S., 2013: Polarimetric radar observations of biological scatterers in Hurricanes Irene (2011) and Sandy (2012). *J. Atmos. Oceanic Technol.*, **30**, 2754–2767, https://doi.org/10.1175/JTECH-D-13-00056.1.

Zhang, P., S. Liu, and Q. Xu, 2005: Identifying Doppler velocity contamination caused by migrating birds. Part I: Feature extraction and quantification. *J. Atmos. Oceanic Technol.*, **22**, 1105–1113, https://doi.org/10.1175/JTECH1757.1.

Zrnić, D. S., and A. V. Ryzhkov, 1998: Observations of insects and birds with a polarimetric radar. *IEEE Trans. Geosci. Remote Sens.*, **36**, 661–668, https://doi.org/10.1109/36.662746.

——, and ——, 1999: Polarimetry for weather surveillance radars. *Bull. Amer. Meteor. Soc.*, **80**, 389–406, https://doi.org/10.1175/1520-0477(1999)080<0389:PFWSR>2.0.CO;2.