Predicting the Quality of Ionosphere-Reflected Pulsed Radar Returns Using a Machine Learning Ensemble Model

Sean Pompea

2023

Contents

Introduction

Background

The Earth's ionosphere consists of three primary regions, referred to as regions D, E, and F. The ionosphere is a dynamic system that undergoes continual change—for example, the density of electrons increases during the daytime and decreases at night (Guest 2003). Other phenomena include sudden ionospheric disturbances (SIDs) which are instigated by solar flares (Loudet 2013).

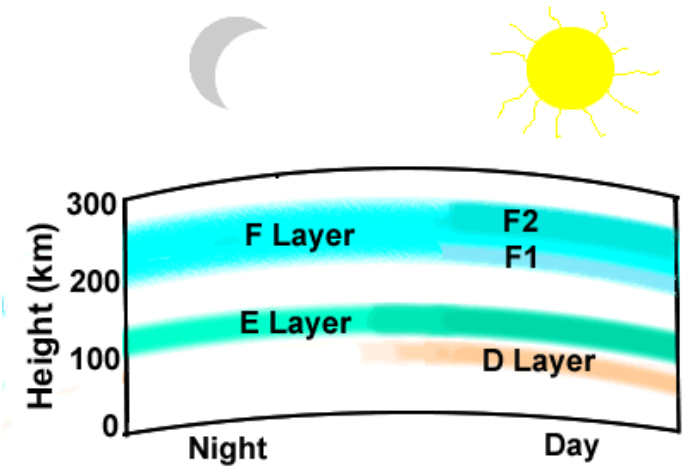

Figure 1: Diagram of the ionosphere (Guest 2003).

Certain radio waves of specific frequencies will travel through the ionosphere and out into space, while others, such as AM and shortwave (both known as high-frequency or HF waves), reflect off the ionosphere and travel back to Earth.

Radar is a technology that generates radio signals and captures the reflected signal returns, making it a technology that scientists and researchers can use to study the dynamics and physics of Earth's upper ionosphere (specifically, E and F regions), including, in particular, its irregularities (Greenwald et al 1985). One such radar is the installation at Goose Bay, Labrador, an HF (8-20 Mhz) pulsed radar installation consisting of an array of 16 log-period antennae which can send pulsed signals to the ionosphere; the return signals are then processed into data which can be studied (Greenwald et al 1985, Walker et al 1987).

Data collection

The Goose Bay radar sends out a multipulse signal to the ionosphere, and receives a reflected return signal that is then used to calculate the autocorrelation function (ACF). (In practice, the term ACF is more often used to refer to the calculated data than to the function itself.) The method of calculatng ACFs is outside the scope of the current study, but for details, see Sigillito et al 1989 (and also Wollf). Each ACF value is a complex number consisting of a real part and an imaginary part; in practice, these are handled as two separate values. In the case of the Goose Bay installation, the manner of its specific multiphase pattern emitted from the 16 antennae results in a return signal that resolves into a vector of 17 distinct ACF value pairss—which, for practical purposes, become 34 separate values. Each second, the radar can generate 25 sets of these 17 ACF pairs (Sigillito et al 1989). Scientists use these data to study the ionosphere.

Clutter

In the context of radar, the term clutter refers to unusable data due to noise, interference, and backscatter from unknown or incidental objects, e.g., rain or birds (O'Donnell 2008). In the case of ionosphere-reflected radar returns, the primary sources of clutter are due to signals passing on through the ionosphere (instead of reflecting back), cancellation due to reflection from an overabundance of ionospheric structures, interference from other transmitters (Sigillito et al 1989), and self-clutter arising from signals arriving at the same time but originating from distinct pulses (Reimer and Hussey 2015).

ACF data therefore needs to cleaned up prior to use, by identifying and removing unusable or 'bad' ACF values, leaving the 'good' values for further study.

Description of the data set

This study made use of the publicly-available Ionosophere data set (Sigillito et al 1988), which was generated by the Goose Bay radar installation in Labrador, and is currently archived at the University of California at Irvine's Machine Learning Repository online archive.

The data set consists of 351 rows. This is about the amount of ACF data that the radar would generate in a time period of 45 seconds (though, that is not to imply that the current data set consists of temporally contiguous readings.) There are 34 independent variables (17 pairs of ACF values, which all together represent the output of a single occurrence of the radar signal and return cycle), and one response variable.

The response variable is a binary variable which is either 'g' or 'b', for 'good' or 'bad' signal returns.

Current approach

In the past, the cleaning of ACF data was done by hand. Nowadays, automated methods are used. Previous work by Sigillito et al (1989) demonstrated the use of neural networks for this task.

For the current study, I hypothesized that a machine learning ensemble model consisting of linear regression, k-nearest neighbors (KNN), and random forest would provide accurate predictions of ACF data quality. Random forest (standalone) was also tested for comparison.

Performance

The ensemble model achieved an F-score of 0.941 with the holdout data set. Standalone random forest performed better, with an F-score of 0.956. See Outcomes for further details.

Materials and methods

Data preparation

The response variable was converted to values of 1 and 0.

One column had a standard deviation of 0, and was therefore deemed to have no predictive power and was subsequently removed, leaving a total of 33 independent variables.

Model description

A machine learning ensemble model was built by first fitting three individual models—linear regression, k-nearest neighbors (KNN), and random forest—and then combining their predictions into a single set of predictions. While the current study is ultimately a classification task, it can be useful to perform intermediate steps using regression, where predicted outcomes can fall anywhere between 0 and 1. In each case, predictions were first generated as continuous outcomes, and then subsequently these values were then standardized to 0 or 1: values above 0.5 were treated as good outcomes (1), while 0.5 and below were treated as unusable outcomes (0), i.e., clutter. For the ensemble model, each prediction was calculated by taking the average of the continuous versions of predictions from the three underlying models, and then similarly standardizing to 0 or 1. Previous work by Irizarry (2019) provides precedence for using this sort of regression-like approach for a classification task; specifically, see section 27.8 of Irizarry 2019, in which they demonstrate the use of linear regression to classify an image of a written digit as either a 2 or a 7. The benefit of this approach is that, when constructing the ensemble, averages can be calculated, rather than resorting to, e.g., a "majority wins" approach.

The linear regression model had no tunable parameters. After tuning via 10-fold cross validation, KNN's `k` parameter was set to 3. Similarly, after tuning via 10-fold cross validation, random forest's `mtry` parameter was set to 10.

Models were developed using the training and validation sets. Afterward, random forest and the ensemble model were evaluated using the holdout set.

Results

Exploratory analysis

36% of the data represent good signals; 64% represent bad/unusable signals (i.e., clutter).

Since all independent variables in the data set were calculated using the same ACF method (see the Introduction), it's reasonable to think that distinctions amongst predictors, while present, probably won't be meaningful from an exploratory standpoint (insofar as there is likely no semantic distinction amongst the variables).

Two approaches were used to visualize the data in the hopes of gaining insights: heatmap, and principal component analysis (PCA).

Distance between good and bad ACF values: heatmap

A heatmap (see Figure 2) of the data set provided some indication, though not a strong one, that there were distinctions between good radar signal ACFs and signal clutter.

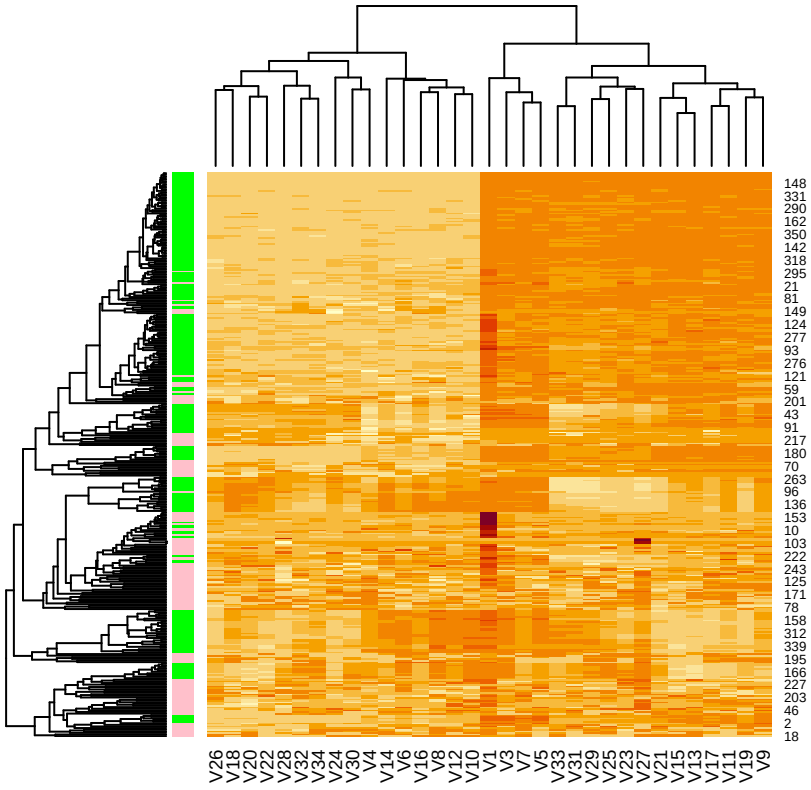

Figure 2: Heatmap of ACF data; green indicates good; pink indicates bad.

Principal component analysis

PCA (principal component analysis) showed that there was clustering present between good and bad radar returns. The data set was scaled and centered prior to performing PCA. Figures 3-5 show visualizations of the first three combinations of PCs (principal components). (Note that the response variable was recast as numeric in preparation for development of the ensemble model.)

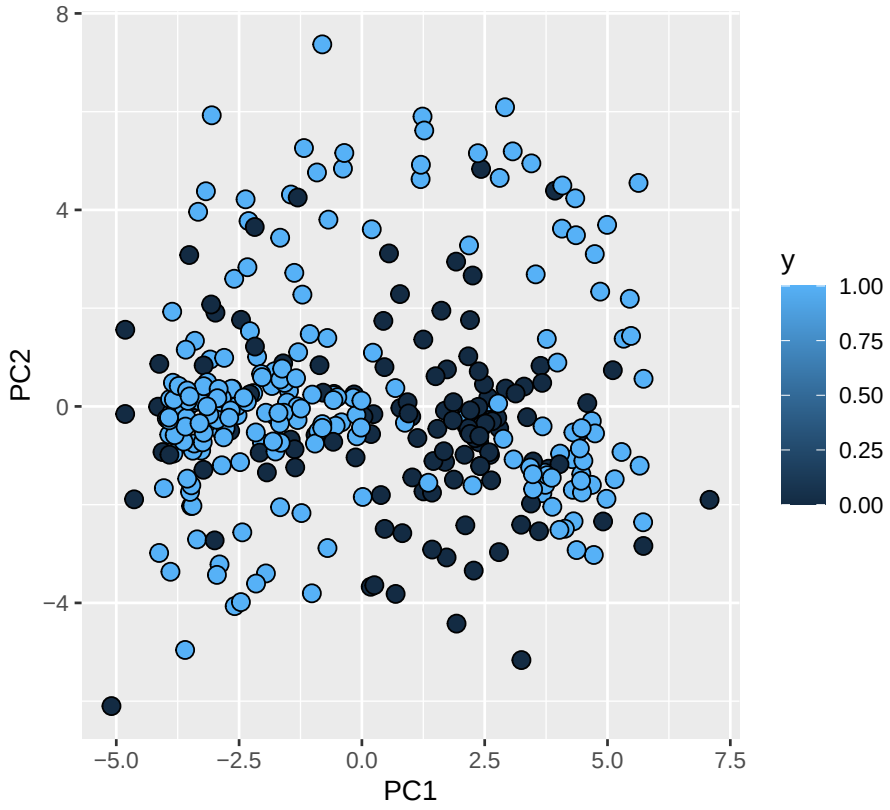

Figure 3: Plot of principal components 1 and 2.

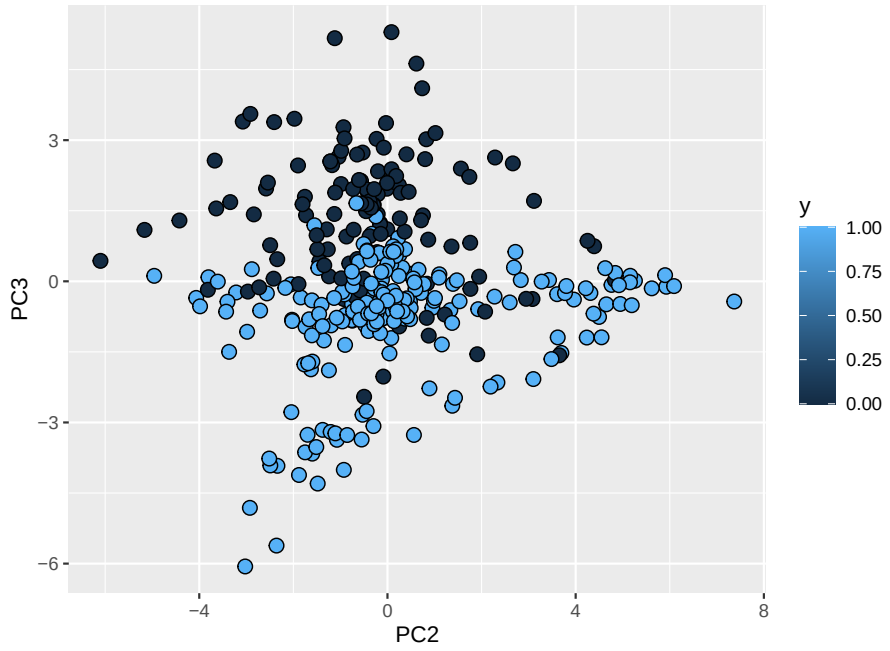

Figure 4: Plot of principal components 2 and 3.

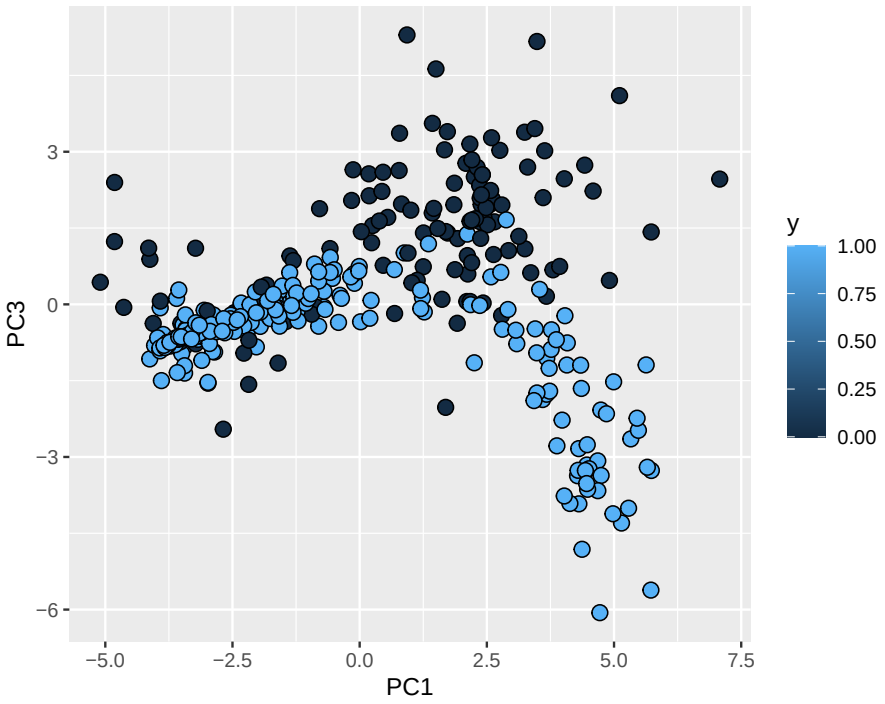

Figure 5: Plot of principal components 1 and 3.

PCA showed that about 48% of the variance in the data could be explained by the first three PCs, and 95% of the variance could be explained by the first 23 PCs. (There were 33 PCs total, corresponding to the 33 independent variables.)

Table 1: Statistics (standard deviation, variance, and cumulative variance) for the first three principal components.

Measure	PC1	PC2	PC3
Standard deviation	2.96852525783212	2.05879687537709	1.6481046197788
Proportion of Variance	0.26703	0.12844	0.08231
Cumulative Proportion	0.26703	0.39548	0.47779

Figure 6 visualizes the proportion of variance by PC.

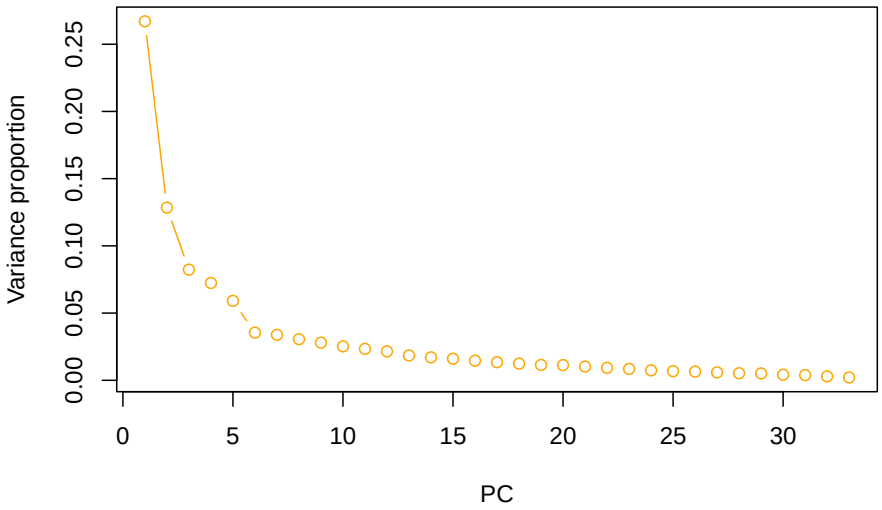

Figure 6: Proportion of variance by PC.

Data splitting

The data was split into train, validation, and holdout sets—50%, 25%, and 25%, respectively, as per Hastie et al (2013)—using a randomized sampling technique.

After splitting the data, the proportion of outcomes in the training set roughly matched the original data set prior to splitting:

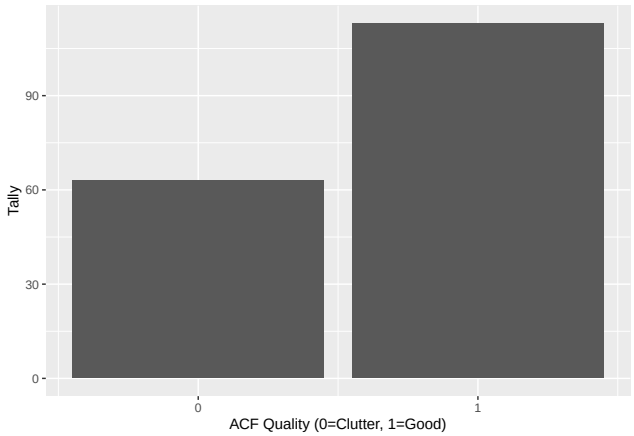

Figure 7: Distribution of the response variable in the training set.

Modeling

First, three models were each employed individually one by one (linear regression, KNN, and random forest). Then, they were combined into an ensemble model.

Linear regression

The first model employed was linear regression. The training data set was fitted with a linear regression model, predictions were generated, and subsequently those predictions were standardized to 0 or 1 (bad or good ACF values). Its performance was evaluated using the validation set. The results were as follows: the LR model achieved an overall accuracy of 0.862, a sensitivity of 0.982 (an indicator of how often the model

correctly identified good ACF values as good), a specificity of 0.645 (an indicator of how often the model correctly identified clutter ACF as clutter), and an F-score of 0.902.

Table 2: Statistical results for LR on the validation set.

	Results for linear regression (validation set).
Accuracy	0.862
Sensitivity	0.982
Specificity	0.645
F1	0.902

The receiver operating characteristic (ROC) curve for LR predictions produced an area-under-curve (AUC) value of 0.8137. (AUC values closer to 1 are indicative of an accurate model.) Note that in the case of binary outcomes, the associated ROC curve is, visually, not very exciting, but nonetheless provides some information, particularly regarding the AUC value; this is also the case for ROC curves used to describe clinical tests (see Lalkhen and McCluskey 2008 for example plots). (For an interesting discussion on the topic of visualizations of ROC curves for binary outcomes, see Muschelli 2021.)

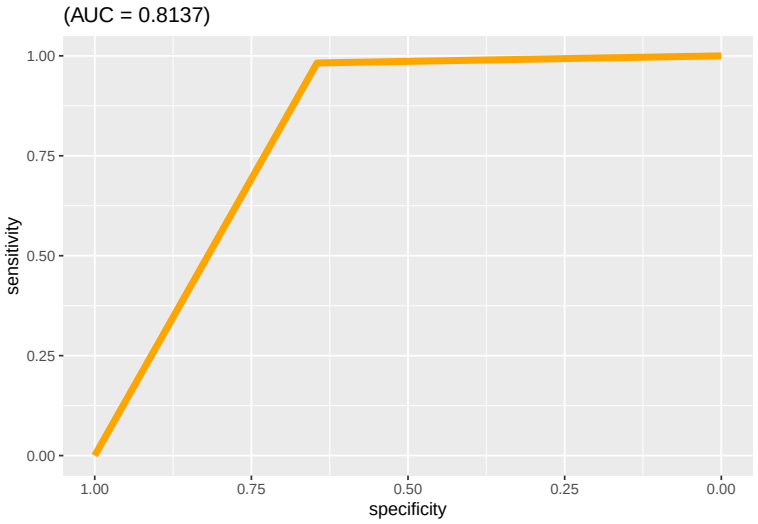

Figure 8: ROC curve for LR.

KNN

The second model employed was k-nearest neighbors (KNN). 10-fold cross validation was used to determine an optimal value of 3 for tuning parameter `k`. The model performed slightly worse overall than LR, with an accuracy of 0.839, and an F-score of 0.885.

Table 3: Statistical results for KNN (validation set).

	Results for KNN (validation set).
Accuracy	0.839
Sensitivity	0.964
Specificity	0.613
F1	0.885

The ROC curve for KNN indicated an AUC of 0.7886—a lower AUC than LR.

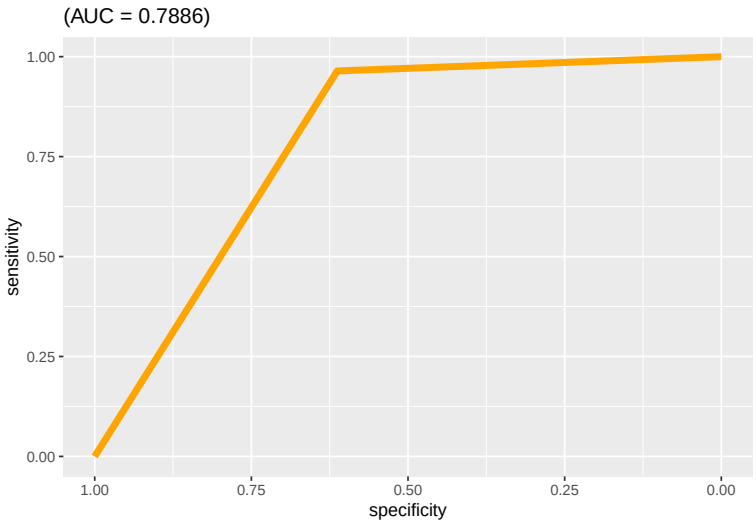

Figure 9: ROC curve for KNN.

Random forest

Similarly to the approach with LR and KNN, a random forest model was trained on the training data set. 10-fold cross validation was used to choose an optimal value of 10 for the `mtry` tuning parameter (a parameter which determines the number of features considered at successive points where decision trees which make up a random forest are being formed). RF performed well with the validation set, with accuracy of 0.908 and F-score of 0.927—better performance than both LR and KNN.

Table 4: Statistical results for random forest (validation set).

Results for random forest (validation set).	
Accuracy	0.908
Sensitivity	0.911
Specificity	0.903
F1	0.927

For random forest, an AUC of 0.907 was obtained for the ROC curve, which is the highest AUC out of the three models.

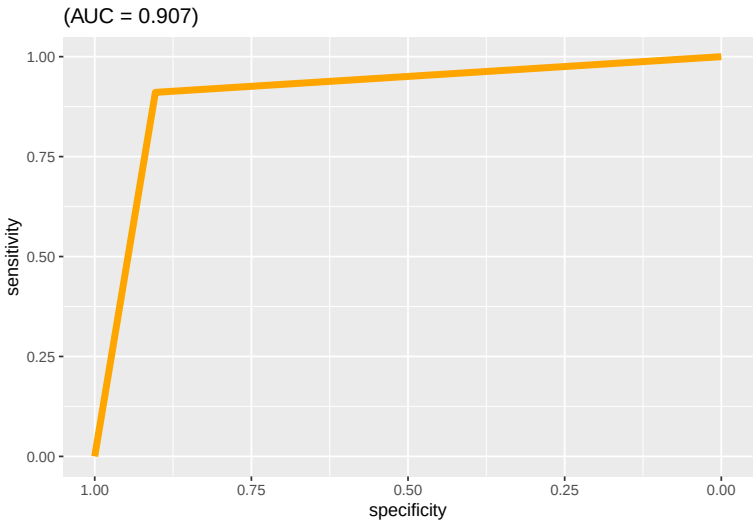

Figure 10: ROC curve for random forest.

Ensemble model

For the ensemble model, continuous versions (not yet standardized to 0 or 1) of the predicted outcomes from the three individual models (linear regression, KNN, and random forest) were averaged, and each average was then standardized to 0 or 1. Using the validation set, the ensemble achieved accuracy of 0.885, sensitivity of 0.964, specificity of 0.742, and an F-score of 0.915. This would seem to indicate slightly better performance with the ensemble model than with standalone LR or KNN, but not better than standalone RF.

Table 5: Statistical results for ensemble (validation set).

	Results for ensemble model (validation set).
Accuracy	0.885
Sensitivity	0.964
Specificity	0.742
F1	0.915

For the ensemble model, an AUC of 0.8531 was obtained for the ROC curve.

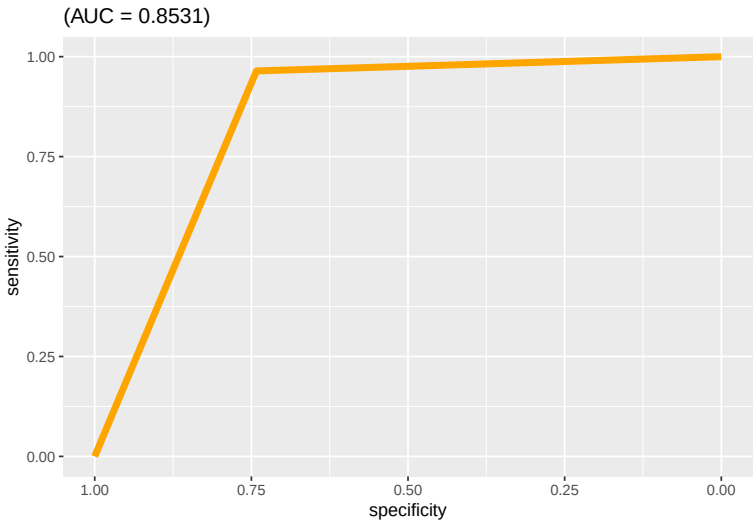

Figure 11: ROC curve for ensemble.

Testing with the holdout data set

At this point, both random forest and the ensemble model were then tested with the holdout set. See the next section (Outcomes) for details and overall performance statistics.

The ROC curve for ensemble, when testing with the holdout set, yielded an AUC of 0.8906. For RF, the AUC was 0.9353 (compared with the validation AUC of 0.907), which tends to point to the RF model's robustness with new data.

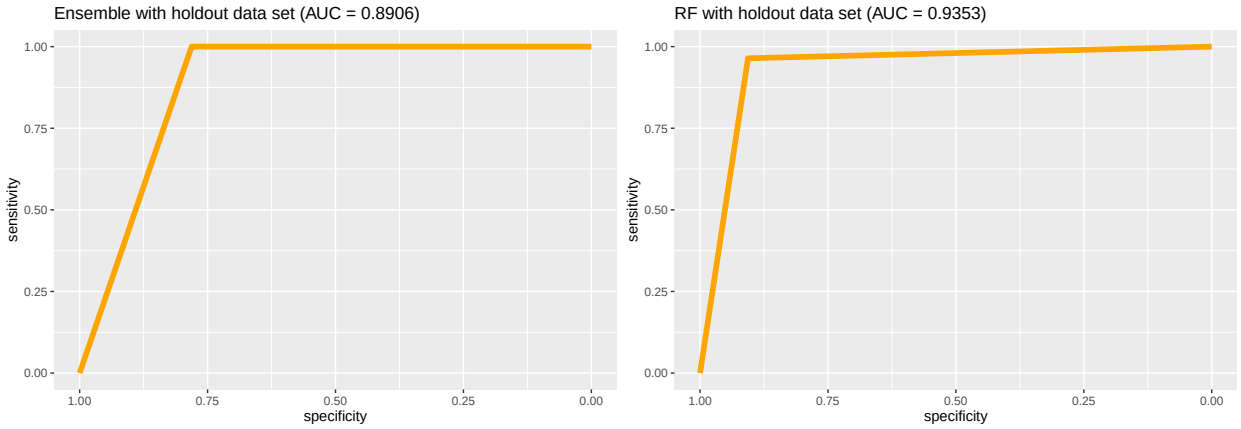

Figure 12: ROC curve for ensemble and RF (using holdout set).

Outcomes

When applied to the holdout data set, the ensemble model achieved a sensitivity of 1, and specificity of 0.781. Its F-score of 0.941 improved over the F-score that the model achieved with the validation set (0.915).

The random forest, when applied to the holdout set, performed better overall than the ensemble model, with a sensitivity of 0.964, specificity of 0.906, and F-score of 0.956 (versus 0.927 during validation).

Overall, when tested with the holdout set, random forest performed better than the ensemble model.

Table 6 summarizes this study's findings.

Table 6: Summary of outcomes (unless noted otherwise, results are from the validation set).

	LR	KNN	RF	Ensemble	RF_holdout	Ensemble_holdout
Accuracy	0.862	0.839	0.908	0.885	0.943	0.920
Sensitivity	0.982	0.964	0.911	0.964	0.964	1.000
Specificity	0.645	0.613	0.903	0.742	0.906	0.781
F1	0.902	0.885	0.927	0.915	0.956	0.941

Discussion

Sigillito et al's work involved using neural networks. They reported that they achieved a sensitivity and specificity, respectively, of 95.9% and 66.7% for a linear perceptron; 98.4% and 63% for a nonlinear perceptron, and 100% and 88.9% for a multilayer feedforward network (MLFN). The random forest in this study had the closest performance to their MLFN. It seems that various approaches (in their study and this one) leave something to be desired when it comes to specificity.

The current study hints at the idea that a single, apt machine learning model may in cases be preferable to ensemble approaches—or that, really, random forest is simply a generally performant model.

Ensemble's performance with respect to specificity is likely due in part to the underlying models' performance in that category; it is also possible that there is a hidden disadvantage to the simple averaging technique that was employed. Further exploration into more robust methods of constructing ensemble models (perhaps, e.g., weighted averages based on the performance of underlying models, etc.) could prove fruitful.

The small size of the training set (87 rows) may have contributed to overfitting—this possibility cannot be ruled out.

Potential future work in this area would likely need to involve using a larger data set, to ward off overfitting. Ribeiro et al (2013) have developed a simulator capable of generating artificial ACFs, so it could be possible to use that simulator (or a similar one) to generate a very large data set that could facilitate improved model fits.

Conclusions

Machine learning is a feasible approach for predicting the quality of radar ACFs. For this study, an ensemble model was constructed, and then both random forest and the ensemble model were tested with a holdout data set. Random forest performed better than the ensemble model. Pathways of refinement in the future might include exploring more sophisticated methods of constructing ensemble models, and training with a larger data set. Random forest serves as a promising method for the current task of classifying radar returns.

References

R.A. Greenwald, B. Baker, R.A. Hutchins, and C. Hanuise. "An HF phased-array radar for studying small-scale structure in the high-latitude ionosphere." 1985. Radio Science, Volume 20, Number 1, Pages 63-79, January-February, 1985.

P. Guest. "HF and Lower Frequency Radiation - The Ionosphere and the Sun." 2003. https://www.met.nps.edu/~psguest/EMEO_online/module3/module_3_2.html. (Academic Web server with no copyright present.)

T. Hastie, R. Tibshirani, J. Friedman. Elements of Statistical Learning, Second Edition. 2013. Springer.

A.G. Lalkhen and A. McCluskey. "Clinical tests: sensitivity and specificity." Continuing Education in Anaesthesia, Critical Care & Pain, Volume 8, Number 6, 2008, pp. 221-223.

L. Loudet. Sudden Ionospheric Disturbances Monitoring Station A118 website. 2013. https://sidstation.loudet.org/.

R. Irizarry. Introduction to Data Science. 2019. LeanPub. https://leanpub.com/datasciencebook.

J. Muschelli. "ROC and AUC with a Binary Predictor: a Potentially Misleading Metric." J Classif. 2020;37(3):696-708. doi:10.1007/s00357-019-09345-1.

R.M. O'Donnell, "Introduction to Radar Systems, Lecture 7: Radar Clutter and Chaff (slides)." 2008. https://www.ll.mit.edu/outreach/radar-introduction-radar-systems-online-course.

A.J. Ribeiro, P.V. Ponomarenko, J.M. Ruohoniemi, J.B.H. Baker, L.B.N. Clausen, R.A. Greenwald, and S. de Larquier. "A realistic radar data simulator for the Super Dual Auroral Radar Network." Radio Sci., 2013. doi:10.1002/rds.20032.

V. Sigillito, S. Wing, L. Hutton, K. Baker. Ionosophere data set. 1988. UC Irvine Machine Learning Repository. https://archive.ics.uci.edu/dataset/52/ionosphere. DOI: 10.24432/C5W01B.

V. Sigillito, S. Wing, L. Hutton, K. Baker. "Classification of Radar Returns from the Ionosphere using Neural Networks." Johns Hopkins APL Technical Digest, Volume 10, Number 3, 1989.

A.D. M. Walker, R. A. Greenwald, and K. B. Baker. "Determination of the fluctuation level of ionospheric irregularities from radar backscatter measurements." Radio Science, Volume 22, Number 5, Pages 689-705, September-October, 1987.

C. Wolff. "Radar Basics: Correlation." No publication date. https://www.radartutorial.eu/10.processing/sp54.en.html.