Final Report: Classification of Aircraft Wildlife Strikes from FAA Reports

By: Jerry Tafoya, Siavash Norouzi

1) What is the "core" research problem that the group is using ML modeling for its investigation?

When we started this project we wanted to find a useful way to apply Machine Learning to aircraft wildlife strikes provided by the FAA. When we began our research into how ML had been applied to similar works, the both us decided that it would be useful if we could create the start of some kind of "Heads Up" or "Recommendation" system that could let aircraft personnel and air traffic control of what they should be expecting after an incident. In our data we had a lot of parameters that were very similar and or not really needed so we were able to reduce from 66 down to 24 that seemed mostly independent of each other by hand picking them. With ML in mind we decided to choose test parameters, these ended up being "Injury", "Flight Impact" and "Aircraft Damage". These parameters were chosen as they seemed to be the most important due to their monetary, safety and reputational ties to the aircraft business.

2) **All related literature works by peer researchers, and their obtained modeling results**

When looking for examples in the literature that would be applicable to our dataset and to our "core" problem we found some interesting results. We categorized these papers into two groups, those being applicable to the dataset and to those that were applicable to our "core" problem. To those in the first group we picked three papers, these being from Sonawane (Heart Disease), Mustafa (River Sediment) and Morfidis (Seismic Damage). The paper from Sonawane in particular was interesting to us, in that we found that for their dataset they created a prediction based system that used a Multilayer Perceptron NN, it accepted 13 clinical features as input and trained using back propagation with up to 98% accuracy. We found this to be a very applicable approach to the dataset since it used a variety of test parameters and many thousands of instances. In the papers regarding our "core" problem, we found that they were looking at improvements upon standard ML models, in particular they fine tuned their models to maximise their True Positive Performance which in our case for predicting what happens in an aircraft incident is very important information to have as soon as possible.

3) **Describe the utilized ML models in the presented research, and comparison to those used in (2) above**

In our ML research we were able to choose 7 different models with justification for each. We implemented a MLP model due to its strong presence in the literature. We chose Least Regression for its strong binary classification which is how we encoded our data. LDA was used due to its generative probabilistic approach which we thought would be good due to the number of test parameters. We added a KNN and SVM model so that they could potentially find strong clusters in the data. Finally we had CART and NB models implemented for their Binary Tree properties and for its normally distributed data performance respectively (done in later testing). This follows the ideology in our dataset papers where they tested many ML models in the hopes of finding one that fit the dataset well. After our work training the models, we found that CART had performed the best in most cases and decided to work further on that to start working towards our “core” problem. From here we began to fine tune our CART model and apply a reduction of complexity through our PCA. This followed the style of approach found in the literature by fine tuning their models and providing improvements in training time.

4) **Research obtained results and any advancement over the researches in (2) above**

Before applying PCA, we observed that the CART model performed the most reliably at the highest level (about 73 - 78% per test parameter). We decided to fine tune our model based on PCA for the three classes of Injury, Aircraft Damage, and Flight Impact. The following is the comparison between the model’s accuracy on these parameters as targets when the input parameters were handpicked and when they were selected based upon PCA. For Injury, the model’s weighted accuracy improved from 92% with handpicked parameters to 95% with PCA interpreted parameters. Fig. 1 shows the accuracy results for Injury.

```
Accuracy: 0.9491525423728814
[[60  5]
 [ 1 52]]
              precision    recall  f1-score   support

           0       0.98      0.92      0.95        65
           1       0.91      0.98      0.95        53

    accuracy                           0.95       118
   macro avg       0.95      0.95      0.95       118
weighted avg       0.95      0.95      0.95       118
```

Fig 1. Weighted accuracy of CART model on Injury as target

The low association of Injuries with the Operator could be the reason for this improvement. In the case of Aircraft Damage, PCA did not change the weighted accuracy. The weighted accuracy remained 74% before and after applying PCA as shown in Fig 2.

```
Accuracy: 0.7443809017156536
[[2512 1001]
 [ 921 3085]]
              precision    recall  f1-score   support

           0       0.73      0.72      0.72      3513
           1       0.76      0.77      0.76      4006

    accuracy                           0.74      7519
   macro avg       0.74      0.74      0.74      7519
weighted avg       0.74      0.74      0.74      7519
```

Fig 2. Weighted accuracy of CART model on Aircraft Damage as target

At least it can be stated that PCA helped reduce the dimensionality and accelerated the training without losing any performance. This scenario, however, was different for Flight Impact. The model’s weighted accuracy dropped from 84% to 74% when using the PCA interpreted parameters. Fig 3 contains the details of the model accuracy.

```
Accuracy: 0.7441166139796277
[[2013  693]
 [ 764 2224]]
              precision    recall  f1-score   support

           0       0.72      0.74      0.73      2706
           1       0.76      0.74      0.75      2988

    accuracy                           0.74      5694
   macro avg       0.74      0.74      0.74      5694
weighted avg       0.74      0.74      0.74      5694
```

Fig 3. Weighted accuracy of CART model on Flight Impact as target

It should be noted that we are running the model along the strongest correlation revealed by the PCA analysis. This correlation has the Operator on its other side which itself is the source of skewness in the data. While counterintuitive, it is not unexpected; even with cutting out unnecessary features, the principal components PC1 and PC2 have improved to control only around 50% of the data variation. Moreover, given the data skewness, it cannot be expected that PC1 and PC2 can control any higher variation. At this point, we can conclude that other analyses should be used such as LASSO to maybe help gain more insights into our skewed data. This is discussed in the next section as a suggestion for future work.

5) Conclusion/Observations/Future-Work

In the project as a whole we found that the PCA provided some strong insights into the data and helped us with dimensionality reduction. It revealed some correlations in the data that we had not thought about before which could potentially explain some of our results in terms of performance. Still we could have tried other methods such as LASSO that would have been different in the fact that it performs well with highly skewed data which was very prevalent in ours, it is even able to automate certain parts of model selection (ie. variable selection/parameter elimination) which could have been very helpful. In the future we believe that transfer learning would find great results, our data contains values (outside of our test parameters) that are not available at the moment of an accident, these being “species name” and “species number”. If we were able to apply these classifications to another model and then feed those to our current implementations we could create a more autonomous system that would need less maintenance. In conclusion, we feel that we had reached a strong foundation in trying to solve our “core” problem of creating some kind of heads up system for airlines. We also feel that this work was useful as it would have the potential to improve the quality of aircraft management, passenger safety and incident response times.

Citations/Refernces

[1] Konstantinos Morfidis, Konstantinos Kostinakis,
Approaches to the rapid seismic damage prediction of r/c buildings using artificial neural networks, Engineering Structures, Volume 165, 2018, Pages 120-141, ISSN 0141-0296, https://doi.org/10.1016/j.engstruct.2018.03.028.

[2] Kurt, Imran, Mevlut Ture, and A. Turhan Kurum. "Comparing performances of logistic regression, classification and regression tree, and neural networks for predicting coronary artery disease." *Expert systems with applications* 34.1 (2008): 366-374.

[3] Sonawane, Patil “Prediction of heart disease using multilayer perceptron neural network” *Research gate* Feb 2015

[4] Li, Deying, et al. "Landslide Susceptibility Prediction Using Particle-Swarm-Optimized Multilayer Perceptron: Comparisons with Multilayer-Perceptron-Only, BP Neural Network, and Information Value Models." *Applied Sciences* 9.18 (2019): 3664.

[5] Mustafa, M. R., et al. "River suspended sediment prediction using various multilayer perceptron neural network training algorithms—a case study in Malaysia." *Water resources management* 26.7 (2012): 1879-1897.