Explainable AI: Rotorcraft Attitude Prediction

Hikmat Khan
Rowan University
Glassboro, New Jersey,
United States

Ghulam Rasool
Rowan University
Glassboro, New Jersey,
United States

Nidhal Carla Bouaynaya
Rowan University
Glassboro, New Jersey,
United States

Tyler Travis
Federal Aviation Administration
Atlantic City, New Jersey,
United States

Lacey Thompson
Federal Aviation Administration
Atlantic City, New Jersey,
United States

Charles C. Johnson
Research Engineer, Flight Test Engineer, and Program Lead,
Federal Aviation Administration
Egg Harbor Township,New Jersey,
United States

ABSTRACT

Rotorcrafts are generally subject to a higher fatal accident rate than other segments of aviation, including commercial and general aviation. The safety improvement for rotorcrafts would directly improve the efficiency of air traffic control, since rotorcrafts operate primarily within low-level airspace; an area that is becoming increasingly complex with new entrants, such as unmanned aircraft systems and urban air mobility. The recent impact of artificial intelligence and deep learning algorithms on various aspects of our lives has led to the investigation of the application of these algorithms in the aviation domain; as it may offer a prime opportunity to enhance safety within the aviation community. In this research, we explore the efficacy, reliability, and, more importantly, the explainability of modern deep learning algorithms. We use machine learning models to predict the attitude (pitch and yaw) of rotorcrafts using video data recorded with ordinary cameras. The cameras were mounted inside the helicopter cockpit and recorded outside view through windshield continually during the flight. We train four different architectures of convolutional neural networks (CNNs), i.e., VGG16, VGG19, ResNet50, and Xception. The models achieved 90%, 91%, 88%, and 88%, respectively, average attitude prediction accuracy on the test video dataset. Furthermore, we use gradient class activation maps (grad-CAM) to ascertain the features and regions of the image that influenced the model to make a specific prediction. We show that CNNs learn to focus on similar features as human operators (pilots), i.e., the natural horizon curve. Our findings demonstrate the feasibility of using deep learning models for attitude prediction from flight videos recorded using ordinary inexpensive cameras. The proposed video analytics framework provides a cost-effective means to supplement traditional Flight Data Recorders (FDR); a technology that is often beyond the financial reach of most general aviation rotorcraft operators.

INTRODUCTION

As the premier agency for promoting and ensuring aviation safety, the Federal Aviation Administration (FAA) continues to highlight the importance of participating in Helicopter Flight Data Monitoring (HFDM) programs to improve flight safety and operational efficiency. Indeed, rotorcraft safety was one of the agency's top ten most wanted list of safety improvements in 2017-2018 and continues to be a high priority in 2020. Organizations including FAA, National Transportation Safety Board (NTSB), and the United States Helicopter Safety Team (USHST) are strong proponents of flight data recorders (FDRs). These organizations and other industry partners are working together to promotes the use of FDRs as a possible mechanism for reducing the fatal accident rate. However, despite their best efforts, barriers to implementation exist. These include, but are not limited to, technical skills required to operate an FDR and costs associated with the acquisition and installation of the FDR. The traditional FDRs require a Supplemental Type Certificate (STC) or Field Approval (FA) to install and operate the device following the Rotorcraft Flight Manual (RFM). The initial acquisition cost of an FDR can range from $9,000 - $50,000, on average. Given a range of factors, rotorcrafts, in general, have a lower participation rate

Presented at the VFS International 76th Annual Forum & Technology Display, Virginia Beach, Virginia, October 6–8, 2020.

in the FDM programs than other forms of aviation, including commercial fixed-wing or part 121 air carriers.

Inexpensive and off-the-shelf video cameras mounted inside the cockpit may offer a potential alternative to traditional FDRs. Even small helicopter operators often have access to, or have the financial means to purchase, one or more off-the-shelf video cameras. These cameras can potentially record all the data that traditional FDRs provide. Moreover, on-board cameras may provide supplementary data that, depending on the type of the FDR, may not be available. Video data from on-board cameras can be used for a variety of tasks, including the flight parameter estimation from instrument panel gauges, flight replay during post-accident investigations, rotorcraft attitude estimation, and any other visual information analysis that can be extracted from video data. In this research project, we focused on the problem of estimation of the rotorcraft attitude, i.e., pitch and roll using video data from on-board cameras. Furthermore, we attempt to explain the model's predictions to establish the reliability and increase the trustworthiness of the AI model.

In this paper, we trained, validated, and tested various machine learning models using a large dataset of videos (broken down into frames) and the ground truth attitude measurements recorded using an on-board Attitude Heading and Reference System (AHRS). We were able to achieve an attitude estimation accuracy of 92% on the test data (i.e., a part of the dataset that was never used for training or validation purposes). We were also interested in elucidating the discriminative features or regions in the image that trigger the deep learning algorithm to make a specific decision. For human experts, the 'horizon curve' generally serves as the discriminative feature for estimating the rotorcraft attitude. We were able to confirm that the discriminative features for the deep learning algorithms matched that of human experts.

Our results demonstrate the feasibility of an inexpensive alternative in the cockpit, i.e., a camera that would facilitate the participation of the aviation community in the FDM programs even for legacy helicopters. Our research efforts provide useful data collection and analysis tools that could significantly improve the safety and operational efficiency of rotorcrafts as well as general aviation.

In this paper, section "related work" starts with a brief overview of existing methods used to enhance rotorcraft safety. In section "methodology", we present the experimental setup and elaborate on the proposed machine learning methods. Section "results and discussion" presents the results and provides a discussion related to the obtained results. Finally, we provide concluding remarks in Section "conclusion and future work" and also present possible directions for future research work.

RELATED WORK

Recently, machine and deep learning algorithms have been successfully used to solve a variety of computer vision problems, including object identification, localization, and pixel-level segmentation in natural images, medical diagnostics, and security applications. (Refs. 1–3). A variety of deep neural networks, especially convolutional neural networks (CNNs) have been proposed and implemented that can achieve above-human accuracy on multiple computer vision tasks, e.g., VGG-16, VGG-19, ResNet, Inception, and Xception, to name a few (Refs. 4–6). Deep learning techniques have shown the ability to learn highly discriminative features in increasingly complex hierarchy directly from the data.

The aviation research community has also investigated machine learning and deep learning techniques for various applications (Refs. 7–9). Khan *et al.* demonstrated that deep learning methods could process videos recorded using off-the-shelf cameras in the helicopter cockpit to infer flight state information (Ref. 9). The authors reported promising predictive accuracy results for multiple flight parameters estimated from video streams of instrument panel gauges (Ref. 9). Alligier *et al.* used machine learning to improve airspeed prediction during the aircraft climb (Ref. 7). In another work, the same authors used machine learning for mass estimation of ground-based aircraft climb prediction (Ref. 8). Kenneth applied structural topic modeling to the Aviation Safety Reporting System (ASRS) corpus to identify topics, trends, and areas that required further investigation (Ref. 10).

Gianazza *et al.* trained a variety of machine learning algorithms to predict the workload of air traffic controllers (Ref. 11). Shin and Hwang used classical computer used hand-engineered features to predict rotorcraft attitude from on-board cameras (Ref. 12). The authors extracted the natural horizon line and used the line as a feature to estimate the roll and bank angles (Ref. 12).

Explainability, trustworthiness, and reliability of modern AI and deep learning algorithms are active areas of research. Several methodologies and approaches have been proposed in the literature, including class activation maps (CAM) and gradient class activation maps (g-CAM) (Refs. 13–17). These methodologies help investigators visualize the features of the input data that influenced the AI/machine learning algorithm to make a specific classification decision. Apart from helping in interpretation, CAM/grad-CAM techniques also explain model behavior and assist in developing trust in AI/machine learning algorithms.

METHODOLOGY

We developed (trained, validated, and tested) deep learning algorithms for the prediction of rotorcraft attitude (i.e., pitch and yaw) using on-board video data. Our deep learning models were able to achieve test predictive accuracy of more than 92%. Furthermore, we used a grad-CAM algorithm to visualize features of the input that influenced trained deep learning algorithms to make specific decisions, i.e., explaining predictions performed by the deep learning algorithm at test time (Refs. 13–17). We found that our trained deep learning models based their predictions on the 'natural horizon curve' just like a human expert.

Figure 1. Dataset: Sample images for nine different classes. Class labels, defined in Table 1, are indicated under each image.

Data Acquisition and Annotation

We used video data recorded using cameras mounted inside the cockpit of S-76 helicopters. The cameras continuously recorded the outside view as seen through the windshield. The cameras used a fisheye lens, and the field of view of the lens was adjusted to capture the broadest possible view of the natural horizon and external scene. We used video data recorded through ten different flights (10 videos), totaling approximately 12 hours of video data. The time and data information (timestamp) provided by the FDR was embedded in the video stream to help in synchronization of the FDR data with the video data during data analysis.

The FDR data (pitch and roll real values) served as the ground truth (annotations) for our training, and validation datasets. The sensors' data from the FDR and cameras were recorded at different sampling rates. The pitch and roll values from the sensors were recorded and stored by the FDR at 10Hz; whereas, the videos were recorded with different frame rates ranging from 15 frames to 45 frames per second depending upon the camera type and settings.

In order to eliminate the possibility of introducing noisy samples in data (i.e., annotating frame with wrong FDR record), we used a comprehensive strategy to synchronize individual frames (extracted from videos) with the FDR data (pitch and roll). First, we extracted all individual frames (images) from a given flight video and then extracted the timestamp (embedded in the video from the FDR) from each frame using the algorithm proposed in (Ref. 20). Later, we matched the extracted timestamps with the corresponding sensor readings from the FDR. Then, we annotated individual frames with their corresponding attitude values (pitch and roll). Finally, we compiled the annotated frames in the form of videos that can be quickly viewed by a human expert to double-check the correctness of the annotation process. In total, the labeled flight videos resulted in approximately 120,000 annotated frames, which defined our dataset for attitude prediction. We divided the dataset into three bins, i.e., training, validation, and test using a 70 : 20 : 10 split ratio.

The attitude data recorded by the FDR consists of two real numbers representing continuous measurements from pitch and roll sensors. For our deep learning predictive framework, we introduced a threshold (α) on the pitch and roll values and defined nine different bins that represent mutually exclusive nine discrete classes. The nine classes are: class 0 - nose down (ND), class 1 - nose up (NU), class 2 - roll positive (RP), class 3 - roll negative (RN), class 4 - ND and RP, class 5 - NU and RP, class 6 - ND and RN, class 7 - NU and RN, and class 8 - level and steady-state (L). We used a threshold value of $\alpha = \pm 3$ for all of our experiments reported in this paper. For example, in case when both the roll and pitch values are within an interval of -3 to $+3$, the state of the rotorcraft was annotated to class 8, i.e., level or steady-state. Detailed description of our class annotation scheme and data distribution is provided in Table 1 and Figure 1 presents representative sample images. Figure 3 presents the dataset distribution (which includes training, validation and test sets) and it can easily be seen that all nine classes are not equally represented; such phenomenon is called *class-imbalance*. We addressed the class-imbalance challenge by assigning higher weighs to minority classes in the final optimization function of our algorithms (Ref. 18). Figure 4 presents the class-wise weighs that were calculated using the *sklearn* built-in function (Ref. 18) (inspired by (Ref. 19)). To better demonstrate the explainability of deep learning models, we generated two versions of the dataset based on cockpit information in the frame. Fig-

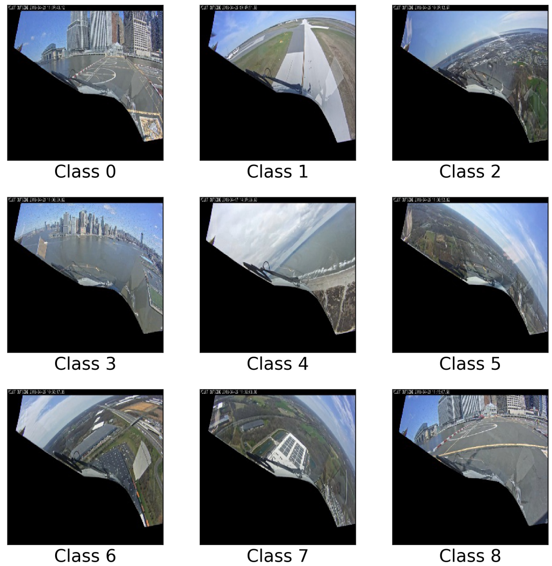

Figure 2. Cockpit subtraction: Sample images for nine different classes. Class labels, defined in Table 1, are indicated under each image.

ure 1 presents sample images of the dataset, which contains cockpit information, while Figure 2 presents samples images of the dataset that do not contain cockpit information. In the latter case, we masked out the cockpit to better understand the inner-workings of the deep learning models and to explain their decisions (more on this will be discussed in the sequel).

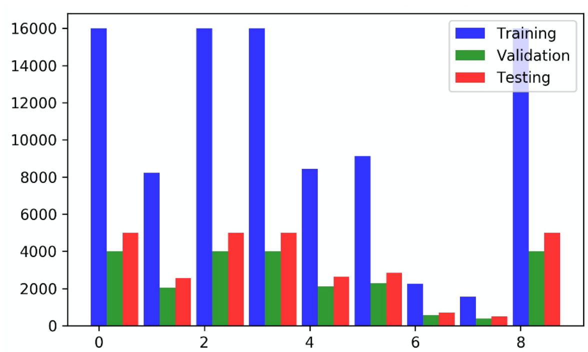

Figure 3. Data distribution of the 9 different classes. Class labels are defined in Table 1

Experimental Setup

We used four different CNNs; VGG16, VGG19, ResNet50, and Xception (Refs. 4–6, 21, 22) for both datasets (i.e., with and without cockpit information in input image). During training, all models were initialized with publicly available ImageNet weights (Ref. 23). We trained VGG16 (Ref. 4) and VGG19 (Ref. 4) using the transfer learning regime (Ref. 23). The first 70% layers of both VGG16 and VGG19 were frozen to the ImageNet weights, while the remaining 30% were up-

Figure 4. Weights of nine different classes. Class labels are defined in Table 1

dated using our dataset. However, for the other two CNNs models, i.e., ResNet50 and Xception, we did not use transfer learning, i.e., all weights were initialized to ImageNet weights (Refs. 5, 6) and updated during training phase. We used early stopping criteria during our training, which uses the Adam optimizer (Ref. 24), to prevent overfitting. All experiments were performed using the Adam optimizer (Ref. 24) with learning rate of 0.001. All other parameters related to initialization, training, and optimization were kept to default values, as discussed in (Ref. 24).

RESULTS AND DISCUSSION

To better investigate the applicability/feasibility of deep learning models for attitude prediction, we trained, validated, and tested four different CNN architectures (VGG16, VGG19, ResNet50, and Xception) on our rotorcraft dataset, under two

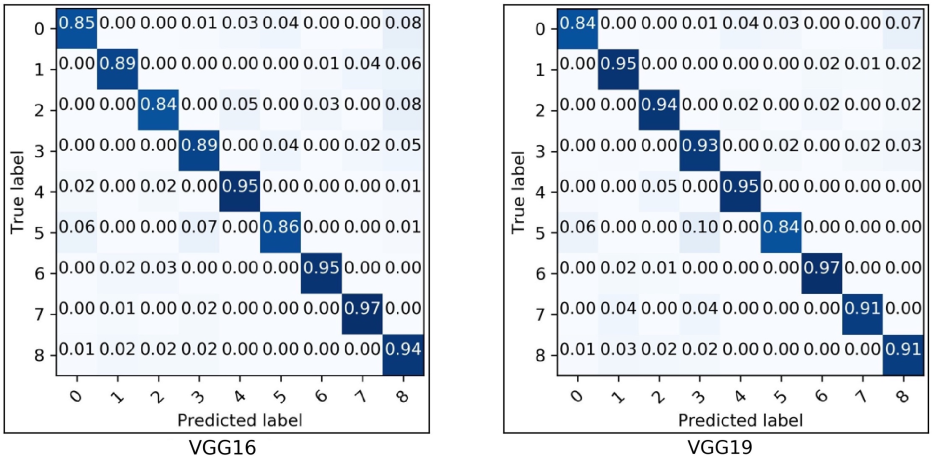

Figure 5. Normalized confusion matrices for 2 CNN architectures with cockpit information dataset, i.e., VGG16, VGG19.

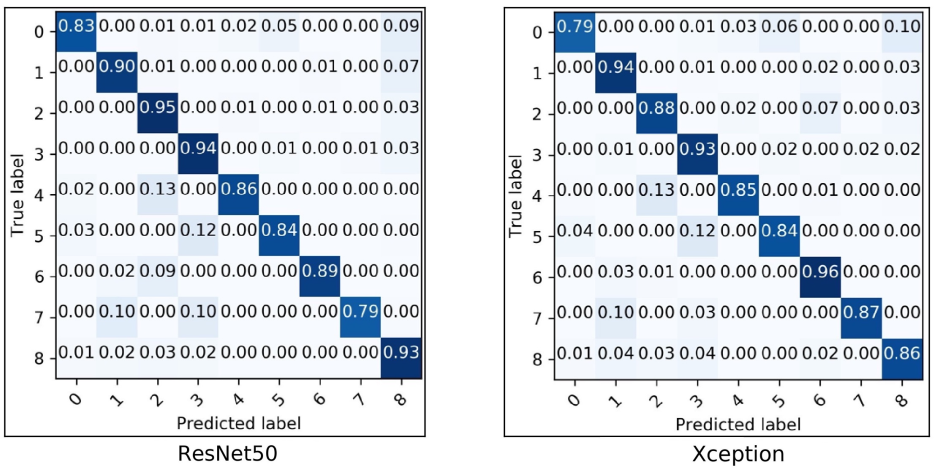

Figure 6. Normalized confusion matrices for 2 CNN architectures with cockpit information dataset, i.e., ResNet50, and Xception.

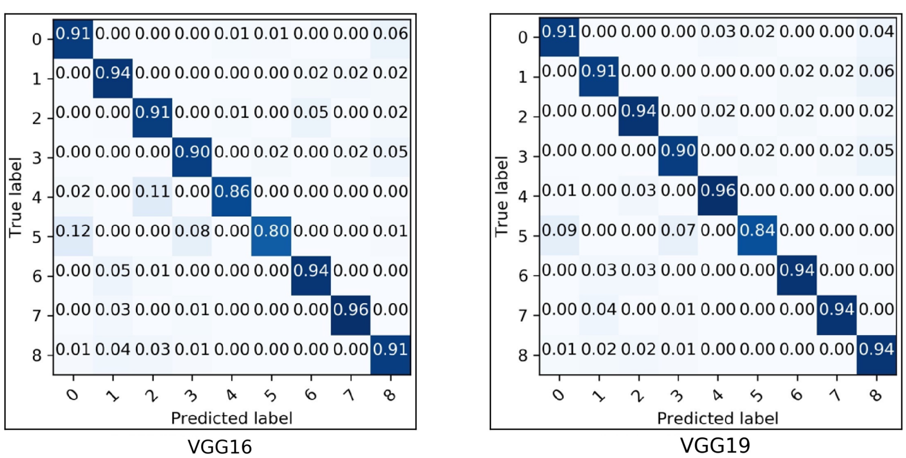

Figure 7. Normalized confusion matrices for 2 CNN architectures without (i.e. masked out) cockpit information dataset, i.e., VGG16, VGG19.

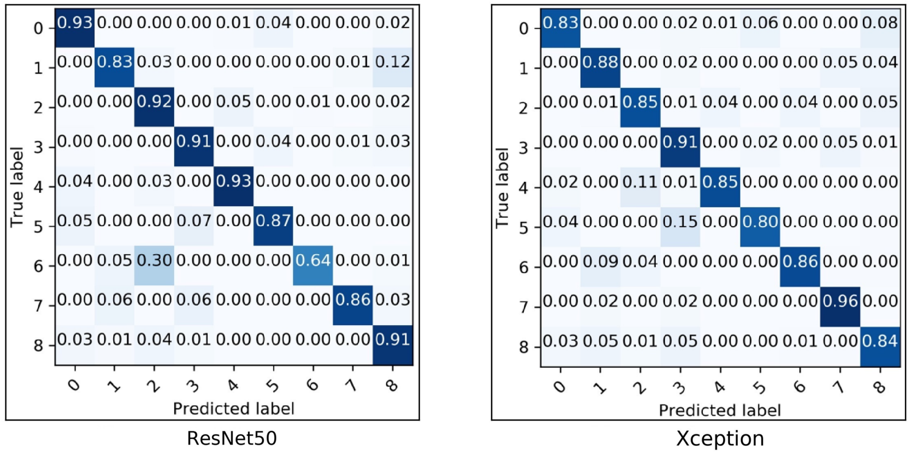

Figure 8. Normalized confusion matrices for 2 CNN architectures without (i.e. masked out) cockpit information dataset, i.e., ResNet50, and Xception.

Table 1. Definition of attitude classes. We used a threshold $\alpha = 3$ and defined 9 discrete classes. Abbreviations used: NU - nose up, ND - nose down, RP - roll positive; RN - roll negative, and L - level or steady-state.

Class	Description	Pitch(P)	Roll(R)
0	NU	$P > \alpha$	$-\alpha \leq R \leq +\alpha$
1	ND	$P < -\alpha$	$-\alpha \leq R \leq +\alpha$
2	RR	$-\alpha \leq P \leq +\alpha$	$R > \alpha$
3	RL	$-\alpha \leq P \leq +\alpha$	$R < -\alpha$
4	NU & RP	$P > \alpha$	$R > \alpha$
5	NU & RN	$P > \alpha$	$R < -\alpha$
6	ND & RP	$P < -\alpha$	$R > \alpha$
7	ND & RN	$P < -\alpha$	$R < -\alpha$
8	L	$-\alpha \leq P \leq +\alpha$	$-\alpha \leq R \leq +\alpha$

different versions of dataset i.e., with and without cockpit information. In the former case, we fed the complete frame, including the whole cockpit image while in the latter case, we masked out the cockpit region from the frame and left only the outside view as input to the deep learning models. The core purpose behind creating two versions of dataset and designing separate experiments was to; First, understand discriminative features in view (i.e., input frame) that the deep learning methods have based their decisions on. Second, it is important to validate the reliability of such features from a human expert point of view. In Tables 2 and 3, we provide a summary of validation and test accuracies for all four models for datasets with and without cockpit information, respectively.

We noted that VGG19 achieved the highest test accuracy, on both with and without cockpit datasets, leading other models by more than 1%. The test accuracy values of ResNet50 and Xception architecture on bath datasets (i.e., with and without cockpit) were below 90%. We also noted that both Xception and ResNet50 architectures performed exceptionally well on the validation datasets; however, these models could not generalize well to test datasets. We further noted models trained on the dataset with cockpit information perform well as compared to their counterparts (without cockpit). From human expert point of view, i.e., helicopter pilots, the cockpit information provides useful information that further aids the machine learning model to better distinguish/detect the horizon curve.

In Figures 5 and 6, we present normalized confusion matrices that present the class-wise accuracies for all four architectures trained on the dataset with cockpit information. In Figures 7 and 8, we present normalized confusion matrices that present the class-wise accuracies for all four architectures trained on the dataset without cockpit information. A confusion matrix expounds the accuracy metric and provides classification errors for all classes separately and can identify class imbalance issues. We noted that, in general, all architectures (VGG19, VGG16, ResNet50, and Xception) had difficulty learning to correctly discriminate 'class 0 - nose down' from 'class 8 - level and steady-state'. We used confusion matrices to study the behavior of our CNN architectures as well as improve their performance by collecting more data for the relevant minority classes. Keeping in view the test accuracy and confusion matrices, we recommend using the VGG19 architecture for the attitude estimation task.

In Figures 9 and 10, we present explainability results using the grad-CAM (Ref. 13) technique for several test images in daylight as well as night settings, for cockpit and without cockpit inputs, respectively. The first and third rows in both Figures 9 and 10 present grad-CAM overlays, respectively. The second and fourth rows in Figures 9 and 10 present corresponding input images, respectively. The red color in overlay images represents a large contribution from these pixels to the prediction, while the blue color represents a small contribution. In our discussions with human experts, i.e., helicopter pilots, we found that the only discriminative and reliable source of information for predicting the attitude is the "natural horizon". In our experiments, we also observed that deep learning models, in both with and without cockpit scenarios, also focus on the horizon while predicting the attitude (note red areas in overlay images). We believe that such visualizations could help in establishing the reliability and trustworthiness in predictions and also provide explainability of the trained neural network models.

Table 2. Validation and test accuracy on the dataset with cockpit information.

Model	Validation Accuracy	Test Accuracy
VGG19	94.2%	91.5%
VGG16	92.9%	90.4%
ResNet50	94.7%	88.1%
Xception	95.0%	88.0%

Table 3. Validation and test accuracy on the dataset without cockpit information.

Model	Validation Accuracy	Test Accuracy
VGG19	94.3%	92.0%
VGG16	94.6%	90.3%
ResNet50	94.6%	86.6%
Xception	94.1%	86.4%

CONCLUSION AND FUTURE WORK

In this paper, we focused on developing deep learning algorithms to estimate the attitude of a rotorcraft using flight video data recorded with inexpensive cameras mounted inside the cockpit, continuously recording outside view through the windshield. We used sensor data from the FDR to annotate/label video datasets and later split it into training, validation, and testing sets for four different state-of-the-art CNN architectures, i.e., VGG16, VGG19, ResNet50, and Xception. We used class activation maps to visualize the discriminative regions of the image that the model relied on to make its prediction. We found that our trained deep learning models also used the horizon curve as the most discriminative region in the input image, just like a human expert, i.e., human pilot. Our results demonstrate the applicability of reliable and trustworthy deep learning models for aviation safety applications. In future research work, we will focus on proposing deep learning algorithms for attitude prediction using the attitude indicate gauges inside the cockpit of the S-76 helicopter. The availability of such auxiliary prediction, which is based on the cockpit gauge, would be important to consider, particularly in bad weather conditioning or bad outside visibility. Further, we will cross-evaluate the predictive performance of the models trained on an outside view (i.e., windshield) and attitude indicator gauge (i.e., cockpit) for the attitude prediction task. Such comparison will further develop reliability and trustworthiness in deep learning predictions and will serve in enhancing the deployment of AI systems in aviation safety.

Authors' Contacts:

- Hikmat Khan: khanhi83@students.rowan.edu
- Dr. Ghulam Rasool: rasool@rowan.edu
- Dr. Nidhal C. Bouaynaya: bouaynaya@rowan.edu
- Charles C. Johnson: Charles.C.Johnson@faa.gov

ACKNOWLEDGMENTS

This work was supported by the Federal Aviation Administration (FAA) Cooperative Agreement Number 16- G-015 and NSF Awards OAC-2008690 and DUE-1610911. Ghulam Rasool was also partly supported by NSF Award OAC-2008690.

REFERENCES

1. Zhao, Q. Z., Zheng, P., Xu, S., and Xindong, W., "Object Detection with Deep Learning: A Review," *IEEE transactions on neural networks and learning systems,* Jan, 2019, DOI:10.1109/tnnls.2018.2876865.

2. Stefan, B., Roland, W., Lutz-P, N., and Mauricio, R.,"A survey of MRI-based medical image analysis for brain tumor studies." *Journal of Physics in medicine and biology,* Vol.58(13), 2013, pp. R97-129. DOI:10.1088/0031-9155/58/13/r97.

3. Suraj, S.,Ravi, K. S., Reddy, M., Nikita, P., Srinivas, S. S. K., and Venkatesh, B., "A Taxonomy of Deep Convolutional Neural Nets for Computer Vision," *Journal of Front. Robotics and AI,* Vol. 2, 2016, DOI:10.3389/frobt.2015.00036.

4. Karen, S., and Zisserman, A., "Very deep convolutional networks for large-scale image recognition," International Conference on Learning Representations Conference Proceedings, San Diego, May 2015.

5. Kaiming, H., Xiangyu, Z., Shaoqing, R., and Jian, S., "Deep Residual Learning for Image Recognition," IEEE Conference on Computer Vision and Pattern Recognition Conference Proceedings, pp.770-778, Las Vegas, Nevada, June, 2016. DOI:10.1109/CVPR.2016.90.

6. Chollet, F., "Xception: Deep Learning with Depthwise Separable Convolutions," IEEE Conference on Computer Vision and Pattern Recognition, pp. 1800-1807, July, 2017, DOI:10.1109/cvpr.2017.195.

7. Alligier, R., Gianazza, D., and Durand, N., "Machine Learning Applied to Airspeed Prediction During Climb," In Proceedings of the 11th USA/Europe Seminar on Air Traffic Management Research and Development, Vol. 16, No. 6, 2015, pp. 3138–3149, June, 2015.

8. Alligier, R., Gianazza, D., and Durand, N., "Machine learning and mass estimation methods for ground-based aircraft climb prediction," IEEE Transactions on Intelligent Transportation Systems, Vol. 16, No. 6, 2015, pp. 3138–3149

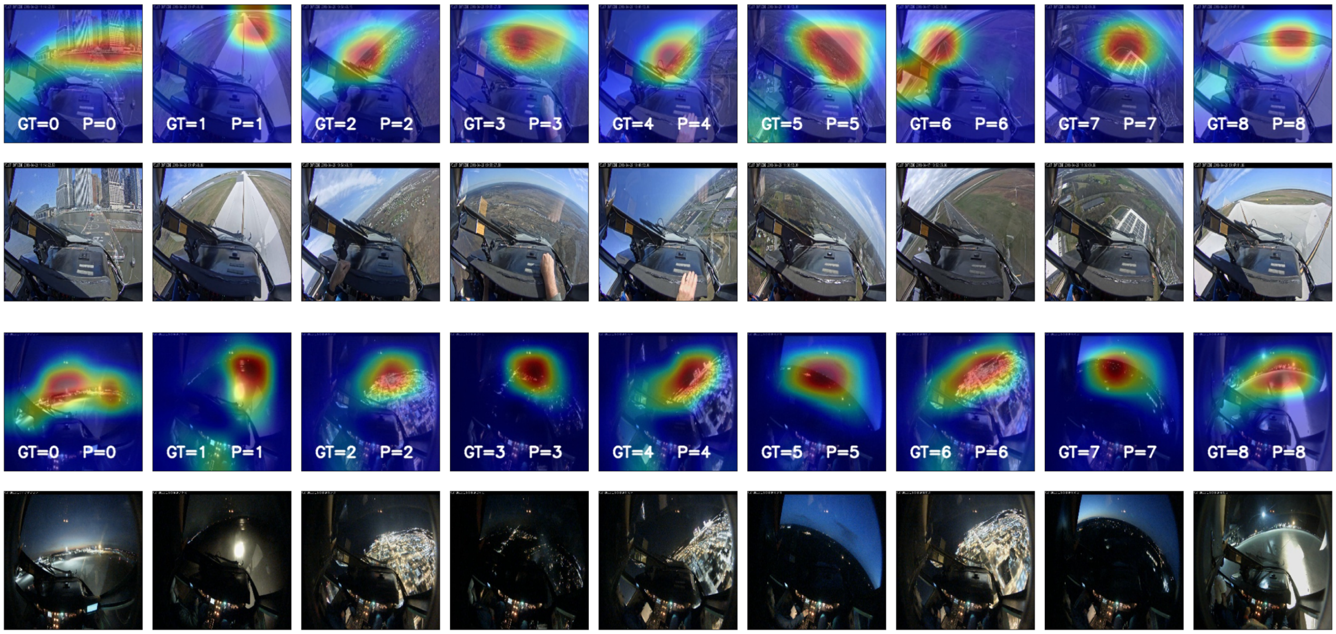

Figure 9. Explainability results using grad-CAM are presented for the day (first two rows) and night settings (last two rows) with input that contain the cockpit information. The first and third rows show grad-CAM overlays drawn on the input images (shown in rows two and four). The red color in an overlay image represents a large contribution from these pixels to the prediction, while the blue color represents a small contribution. Abbreviation used are: GT - ground truth and P - predicted class.

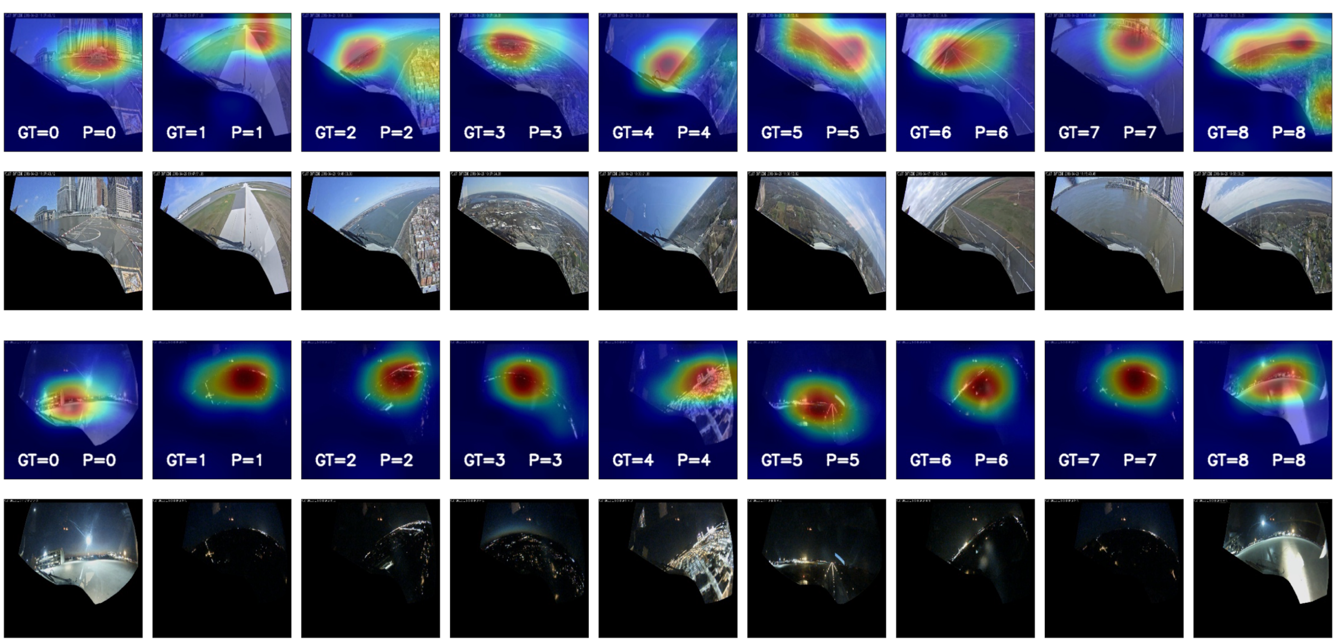

Figure 10. Explainability results using grad-CAM are presented for the day (first two rows) and night settings (last two rows) with input that does not contains the cockpit information. The first and third rows show grad-CAM overlays drawn on the input images (shown in rows two and four). The red color in an overlay image represents a large contribution from these pixels to the prediction, while the blue color represents a small contribution. Abbreviation used are: GT - ground truth and P - predicted class.

9. Hikmat, K., Nidhal, C. B., Ghulam, R., and Charles C. J., “Rotorcraft Flight Information Inference from Cockpit Videos using Deep Learning,” Presented at American Helicopter Society 75th Annual Forum, Philadelphia, Pennsylvania, USA, May 2019.

10. Kenneth, D. K., “Topics and Trends in Incident Reports,” In Proceedings of the 12th USA/Europe Seminar on Air Traffic Management Research and Development, June, 2017.

11. Gianazza. D., “Learning Air Traffic Controller Workload from Past Sector Operations,” In Proceedings of the 12th USA/Europe Seminar on Air Traffic Management Research and Development, June, 2017.

12. Shin, S., and Hwang, I., “Helicopter Cockpit Video Data Analysis for Attitude Estimation using DBSCAN Clustering,” AIAA Infotech at Aerospace, 2016, DOI:10.2514/6.2016-0920.

13. Ramprasaath, R. R., Cogswell, M., Abhishek, D., Vedantam, R., Parikh, D., and Betra, D., “Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization,” IEEE International Conference on Computer Vision, pp. 618-626, Venice, Italy, December, 2017, DOI:10.1109/ICCV.2017.74.

14. Smilkov, D., Thorat, N., Kim, B., Viégas, F. and Wattenberg, M., “SmoothGrad: removing noise by adding noise,” In Proceedings of the 34th International Conference on Machine Learning Representations, Sydney, Australia, August, 2017.

15. Sundararajan, M., Taly, A. and Yan, Q., “Axiomatic attribution for deep networks,” In Proceedings of the 34th International Conference on Machine Learning, Vol 70, pp. 3319-3328, Sydney, Australia, August, 2017.

16. Springenberg, J, T., Dosovitskiy, A., Brox, T. and Riedmiller, M.,“Striving for Simplicity: The All Convolutional Net,” In Proceedings of the 32th International Conference on Learning Representations, Lille, France, December, 2014.

17. Shrikumar, A., Greenside, P., and Kundaje, A., “Learning important features through propagating activation differences,” In Proceedings of the 34th International Conference on Machine Learning, Vol 70, pp. 3145-3153, August, 2017.

18. Scikit-learn Machine Learning in Python, Estimate class weights for unbalanced datasets https://scikit-learn.org/stable/modules/generated/sklearn.utils.class_weight.compute_class_weight.html

19. King, G., and Langche, Z., “Logistic regression in rare events data,” Journal of Political analysis, Vol. 9(2), pp. 137-163, 2001, DOI:10.1093/oxfordjournals.pan.a004868.

20. Smith, R., “An overview of the Tesseract OCR engine,” Ninth International Conference on Document Analysis and Recognition Conference Proceedings, Parana, Sept 2007. DOI:10.1109/ICDAR.2007.4376991.

21. Hu, J., Shen, L., and Sun, G., “Squeeze-and-excitation networks,” IEEE Conference on Computer Vision and Pattern Recognition, pp. 7132-7141, June, 2018, DOI:10.1109/CVPR.2018.00745.

22. Chen, K., Jiangmiao, P., Jiaqi, W., Xiong, Y., Li, X., Shuyang, S., and others, “Hybrid task cascade for instance segmentationm,” IEEE Conference on Computer Vision and Pattern Recognition, Long Beach, CA, June, 2019.

23. Applications - Keras Documentation, Available models for image classification with weights trained on ImageNet, https://keras.io/applications/

24. Diederik, P., Kingma and Jimmy, B., “Adam: A Method for Stochastic Optimization,” International Conference on Learning Representations Conference Proceedings, San Diego, CA, May 7-9, 2015.

25. Gopalakrishnan, K., and Hamsa, B., “A comparative analysis of models for predicting delays in air traffic networks,” In Proceedings of the 12th USA/Europe Seminar on Air Traffic Management Research and Development, June, 2017.

26. Alexander, S. E., Michael, O. B., and Lovell, J. D., “Predicting performance of ground delay programs,” In Proceedings of the 12th USA/Europe Seminar on Air Traffic Management Research and Development, June, 2017.

AUTHORS BIOGRAPHY

Hikmat Khan is currently a PhD student at Rowan University. He is a research fellow supporting the Federal Aviation Administration (FAA) via a research grant/cooperative agreement by evaluating the feasibility of applying deep learning approaches to increase safety within the rotorcraft industry. His research interests include deep learning, continual learning, few-shot learning and optimization.

Nidhal C. Bouaynaya received her Ph.D. degree in Electrical and Computer Engineering from the University of Illinois at Chicago, in 2007. From 2007-2013, she was an Assistant then Associate Professor with the Department of Systems Engineering at the University of Arkansas at Little Rock. In Fall 2013, she joined the Department of Electrical and Computer Engineering at Rowan University, where she is currently a Professor and Associate Dean for Research and Graduate Studies. Dr. Bouaynaya won several Best Paper Awards, the latest being the Best Student Paper Award at the *2019 IEEE Workshop on Machine Learning for Signal Processing*. Her main research interests are in multidimensional signal processing, machine learning and optimal estimation.

She is the Co-Founder and CEO of MRIMath, LLC, a start up company in artificial intelligence for healthcare.

Ghulam Rasool is an Assistant Professor of Electrical and Computer Engineering at Rowan University. He received a BS. in Mechanical Engineering from the National University of Sciences and Technology (NUST), Pakistan, in 2000, an M.S. in Computer Engineering from the Center for Advanced Studies in Engineering (CASE), Pakistan, in 2010, and the Ph.D. in Systems Engineering from the University of Arkansas at Little Rock in 2014. He was a postdoctoral fellow with the Rehabilitation Institute of Chicago and Northwestern University from 2014 to 2016. He joined Rowan University as an adjunct professor and later as a lecturer in the year 2018. Currently, he is the co-director of the Rowan AI Lab. His current research focuses on machine learning, artificial intelligence, data analytics, signal, image, and video processing. His research is funded by National Science Foundation (NSF), U.S. Department of Education, U.S. Department of Transportation (through the University Transportation Center (UTC), Rutgers University), Federal Aviation Administration (FAA), New Jersey Health Foundation (NJHF), and Lockheed Martin, Inc. His recent work on Bayesian machine learning won the Best Student Award at the 2019 IEEE Machine Learning for Signal Processing Workshop.

Charles C. Johnson works as a research engineer, program manager, and technical expert for the Aviation Research Division at the FAA William J. Hughes Technical Center in Atlantic City, NJ. During his 10+ year-career with the FAA, he has led several rotorcraft and unmanned aircraft systems (UAS) research and development simulation/flight test activities that seek to improve aviation safety. Cliff is qualified on the ScanEagle UAS platform. He holds a Bachelor's degree in Mechanical Engineering from Rowan University. He also holds a private pilot's license (single engine land/fixed-wing) and is pursuing his instrument, commercial, and helicopter add-on ratings.

Tyler Travis serves as a Research Analyst at the Federal Aviation Administration's (FAA) William J. Hughes Technical Center in Atlantic City, NJ. She supports several research and development activities that seek to improve aviation safety. During her 5+ years at the FAA Technical Center, Tyler has worked on several UAS and Rotorcraft Human-In-The Loop simulations involving new technologies and procedural changes impacting the National Airspace System. Prior to joining the FAA, Tyler completed her B.S. in Business Administration with a concentration in Management Information Systems at Drexel University in September 2012.

Lacey Thompson works as an Operations Research Analyst for the Unmanned Aircraft Systems (UAS) Engineering Branch at the Federal Aviation Administration's (FAA) William J. Hughes Technical Center in Atlantic City, NJ. During her 7.5 years with the FAA, Lacey has managed several UAS Human-In-the-Loop (HITL) simulations. In addition, she serves as an Aviation Science, Technology, and Mathematics (AvSTEM) Ambassador, creating the curriculum for the first ever module on UAS for the Aviation Monthly Mentoring Program. Lacey holds a Bachelor's degree in Physics from Northwestern State University of Louisiana and a Master's degree in Aeronautics from Embry-Riddle Aeronautical University Worldwide.