

**Naval Information
Warfare Center**



PACIFIC

TECHNICAL REPORT 3315
JULY 2023

Readiness Predictions for the MH-60 Seahawk with the Digital Aviation Readiness Technology Engine

Jamal Rorie, PhD
Benjamin Michlin, PhD
Josh Duclos
Dean Lee
Gary Williams

NIWC Pacific

Approved for public release. Distribution is unlimited.

Naval Information Warfare Center (NIWC) Pacific
San Diego, CA 92152-5001

This page is intentionally blank.

TECHNICAL REPORT 3315
JULY 2023

Readiness Predictions for the MH-60 Seahawk with the Digital Aviation Readiness Technology Engine

Jamal Rorie, PhD
Benjamin Michlin, PhD
Josh Duclos
Dean Lee
Gary Williams

NIWC Pacific

Approved for public release. Distribution is unlimited.

Administrative Note:

This report was approved through the Release of Scientific and Technical Information (RSTI) process in July 2023 and formally published in the Defense Technical Information Center (DTIC) in July 2023.



NIWC Pacific
San Diego, CA 92152-5001

NIWC Pacific
San Diego, California 92152-5001

P. M. McKenna, CAPT, USN
Commanding Officer

M. J. McMillan
Executive Director (Acting)

ADMINISTRATIVE INFORMATION

The work described in this report was performed by the Command and Control (C2) Technology & Experimentation Division of the Command & Control and Enterprise Engineering (C2E2) Department (53629), Naval Information Warfare Center (NIWC) Pacific, San Diego, CA. The NIWC Pacific Naval Innovative Science and Engineering (NISE) Program provided funding for this Rapid Prototyping project.

Released by
Emily Nguyen, Division Head
Command and Control (C2) Technology &
Experimentation Division

Under authority of
Scott Crellin, Department Head
Command & Control and Enterprise
Engineering (C2E2) Department

ACKNOWLEDGMENTS

This is a work of the United States Government and therefore is not copyrighted. This work may be copied and disseminated without restriction.

The citation of trade names and names of manufacturers is not to be construed as official government endorsement or approval of commercial products or services referenced in this report.

Editor: LMG

EXECUTIVE SUMMARY

The Digital Aviation Readiness Technology Engine (DARTE) paradigm, proven highly successful with fixed-wing aircraft such as the F/A-18 Super Hornet and E-2 Hawkeye, has had the mission capable (MC) model expanded to rotary wing aircraft for the first time with predictions for the MH-60 Seahawk. Readiness datasets for the MH-60 are explored and artificial intelligence / machine learning (AI/ML) models are created to predict future MC aircraft in squadrons. New challenges, including different squadron structures and deployment profiles, are addressed. Additionally, new insights are accessible with the addition of new readiness datasets and interrogation of model predictions with eXplainable AI (XAI).

This page is intentionally blank.

CONTENTS

EXECUTIVE SUMMARY	V
1. INTRODUCTION	1
2. DATA	3
3. MISSION CAPABLE STATUS PREDICTION MODEL	5
3.1 DATA PREPARATION	5
3.2 MODEL DEVELOPMENT - TIME-BASED HOLDOUT SPLIT	5
3.3 MODEL DEVELOPMENT - OTHER HOLDOUT SPLITS	6
4. PREDICTION UNCERTAINTY	9
5. XAI	11
6. CONCLUSIONS	13
REFERENCES	15

FIGURES

1. MH-60 monthly average MC histograms, present versus future	4
2. Squadron MC prediction residuals plots.....	7
3. Model performance evaluation, cross-validation with random split.....	8
4. Model performance evaluation, holdout set from datetime-based split.....	9
5. Global feature importance using SHAP.....	11
6. Individual prediction feature importance using SHAP.	12

TABLES

1. Datetime-Based Split r^2 Results.	5
---	---

This page is intentionally blank.

1. INTRODUCTION

Predicting readiness metrics and improving readiness has been a goal of the Naval Aviation Enterprise since the release of Secretary of Defense James N. Mattis’s memo [1] on the readiness of critical aviation platforms in 2018. In response to the imperative set out in this memo, the Commander, Naval Air Forces Force Readiness Analytics Group (CNAF FRAG) partnered with the Naval Information Warfare Center (NIWC) Pacific to develop models using readiness datasets of manning, training, and equipment data to predict F/A-18 readiness metrics, such as squadron-level mission capable (MC) status. This resulted in the Digital Aviation Readiness Technology Engine, or “DARTE” [2]. These F/A-18 models and their successors proved to be very successful, eventually predicting squadron-level monthly average MC metrics and quarterly flight hour execution up to three months in advance [3]. The models also lead to new actionable insights into F/A-18 readiness, including identifying factors that affected performance at both the squadron and enterprise level, some of which spurred the creation of the Resource Optimization, Allocation, and Management (ROAM) model [4]. The DARTE development paradigm proved to be flexible: these models were also successfully adapted to other fixed-wing aircraft such as the E-2 Hawkeye and applied at different time scales. Because of this flexibility, it is expected that the DARTE paradigm can be successfully applied to a wider variety of platforms than fixed-wing aircraft.

The expansion to the MH-60 Seahawk is the first application of the DARTE framework to rotary-wing aircraft. While it is expected that its readiness metrics will similarly be predictable, the new domain is also expected to have some important differences. One such difference is that MH-60 squadrons often form “detachments,” or smaller units consisting of a few MH-60 aircraft each. The MH-60 also have different mission profiles, primarily split into Helicopter Sea Combat (HSC) squadrons—for combat search-and-rescue, combat support, Anti-Surface Warfare (ASuW), replenishment, and troop support missions—and Helicopter Maritime Strike (HSM) squadrons—for surveillance, attack, and Anti-Submarine Warfare (ASW) missions. There are also squadrons that do not neatly fit into the HSC/HSM paradigm. Finally, the MH-60 is known to have less variation over time in the number of MC aircraft in a squadron, so creating a model that appreciably outperforms the benchmark persistence model is a challenge.

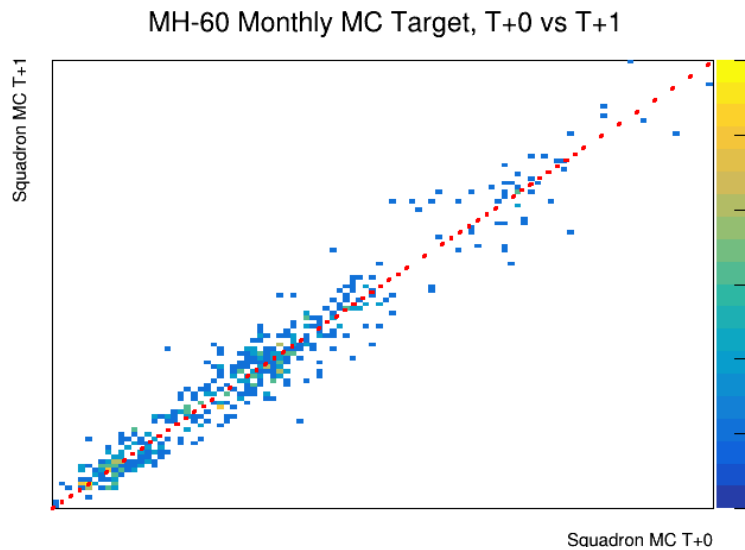
This paper is divided as follows: Sec. 2 describes initial data exploration and the insights gained from it; Sec. 3 explores data preparation, model building, and validation; Sec. 4 discusses the addition of uncertainties to the predictions; Sec. 5 discusses the application of eXplainable AI (XAI) for model and prediction insights. Finally, a summary of findings and next steps are discussed in the conclusion.

This page is intentionally blank.

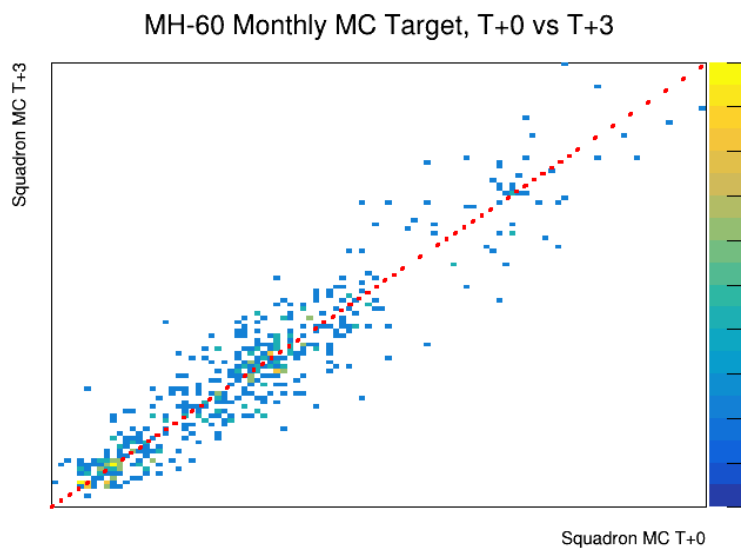
2. DATA

Individual manning, training, and equipment data sources that were used in previous DARTE F/A-18 models also contain MH-60 data. Datasets with records reported daily were aggregated to a monthly level using statistical quantities, such as their mean, max, min, et cetera. Once aggregated, they were then joined to datasets with monthly-level reporting using the reporting month and Unit Identification Code (UIC) as keys. Additional future planning information datasets were later added to the aforementioned joined MH-60 dataset, also using reporting month and UIC as keys.

The resulting dataset was then examined for correlations between features and future squadron-level monthly average MC status. Strong correlations were found between certain features and future monthly average MC status, most notably several elements of future planning information as well as a squadron's current month average MC status. Histograms of current MH-60 MC status versus future months' MC status, denoted as "T+1" for one month in the future and "T+3" for three months in the future, are presented in Figure 1. Note that while there is still a correlation between current and future squadron-level monthly average MC at T+3, the dispersion is higher than at T+1. This, combined with the identification of correlated features, suggests an opportunity to improve predictions with artificial intelligence / machine learning (AI/ML) techniques.



(a) T+1



(b) T+3

- The 2D histograms of MH-60 monthly average MC status comparing the current month, T+0 with (a) one month in the future, T+1 and (b) three months in the future, T+3. The dashed red line represents the condition where the MC status of the squadron does not change. As can be seen by comparing (a) to (b), both one- and two- month average monthly MC status correlates to the MC status of the current month, but the correlation degrades over time.

Figure 1. MH-60 monthly average MC histograms, present versus future

A concern with the MH-60 data that was not present in the F/A-18 data arises from the fact that MH-60 squadrons often form detachments. These detachments, due to their small size and the transient nature, might behave with respect to MC in a way that would skew model predictions and/or performance.

3. MISSION CAPABLE STATUS PREDICTION MODEL

3.1 DATA PREPARATION

The model was built using a Random Forest Regressor from Sci-Kit Learn [5] with `min_samples_leaf` set to one, effectively allowing the model to serve as a quantile regression forest [6]. This model was selected due to the robustness, interpretability of model and predictions, speed of training tree-based learners, and simplicity of data preparation. While random forests are not particularly well-suited to identify patterns in sequences, they can benefit from the inclusion of temporal (“lagged”) features created from historical data and from future features derived from planning data available at prediction time. Several months of these features were added to the dataset. Numeric data (such as floats, integers, et cetera) were retained as-is, categorical features (such as squadron names) were one-hot encoded, and datetime features were reinterpreted as days since or until the record’s index date. Additional features derived from the knowledge of subject matter experts were also included. Initial cleaning of the feature set removed any feature that was more than 50% missing data. Missing data from the remaining features were imputed with the mean value for the feature; positive and negative infinities were imputed with maximum and minimum value for that feature, respectively.

3.2 MODEL DEVELOPMENT - TIME-BASED HOLDOUT SPLIT

Once prepared, the dataset was split into training, test, and validation sets based on datetime boundaries. Care was taken so that no information spillover occurred between sets. For each $T+n$ model, the final division was approximately 70%, 15%, and 15%, respectively. The models were then trained and evaluated in comparison to a persistence model. The performance of the models on the holdout sets, characterized by the r^2 coefficient of determination, are presented in Table 1. While the $T+1$ model prediction r^2 is consistent with the persistence model, higher n predictions using a random forest regressor increasingly outperformed the persistence model. This behavior is shown in Figure 3. The model prediction residuals were also plotted and the distributions were fit in ROOT [7] to determine if they were Gaussian and if they were consistent with a mean of zero; this is presented in Figure 2 for $T+1$ and $T+3$ and shown to be true. The model is shown to be able to predict squadron-level monthly average MC to within a single aircraft $\sim 70\%$ of the time, up to three months into the future.

Table 1. Datetime-Based Split r^2 Results.

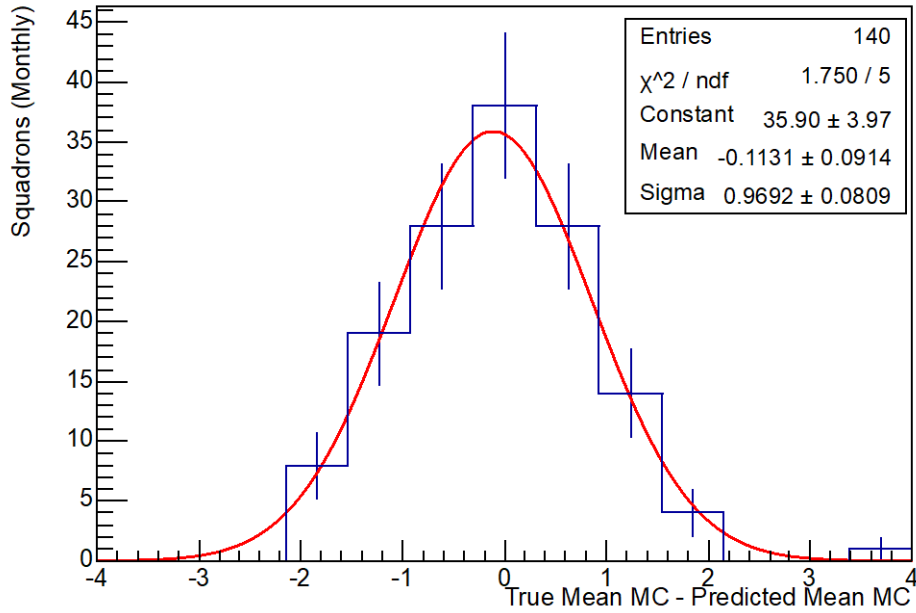
	T+1	T+2	T+3
Model	0.93	0.90	0.90
Persistence	0.92	0.86	0.82

- Comparison of model and persistence predictions for $T+n$ on a validation set created with a datetime-based split.

3.3 MODEL DEVELOPMENT - OTHER HOLDOUT SPLITS

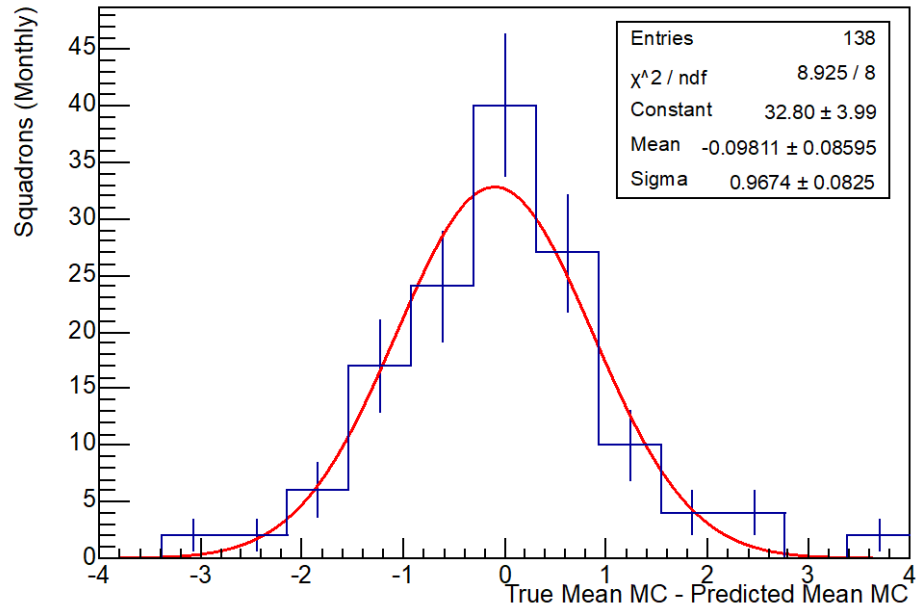
The models were then tested for robustness using other splits of the data. As mentioned in Section 1, there was a concern that the tendency of MH-60 squadrons to create detachments would affect model performance. The detachments have lower Primary Aircraft Authorization (PAA) values than their parent squadrons, so the data was split between squadrons with high and low PAA values. The models remain similarly performant for both high and low PAA groups, thus removing this concern. The introduction also noted that the HSC and HSM squadrons perform different roles. As such, it is possible that the models may be able to take advantage of different features and patterns to make more accurate predictions if the datasets they are presented contain only one type of squadron. However, this also reduces the number of records that the models will see during training by roughly half. Models trained using either only HSC or only HSM data were tested; segregating the data did not show an appreciable difference. The models were further studied with five-fold cross-validation that was stratified on the target and a second feature. This method provided an r^2 value with uncertainties, thus allowing for quantitative comparisons of model performance. However, because this technique does not implement datetime-based boundary conditions on the folds, it likely resulted in some temporal cross-contamination between the folds that may have unfairly assisted the models in making predictions on some folds. Even with this consideration, model performance on each fold was consistent with performance seen in datetime-based splits.

MC Residuals Month T+1, Time Split



(a) T+1

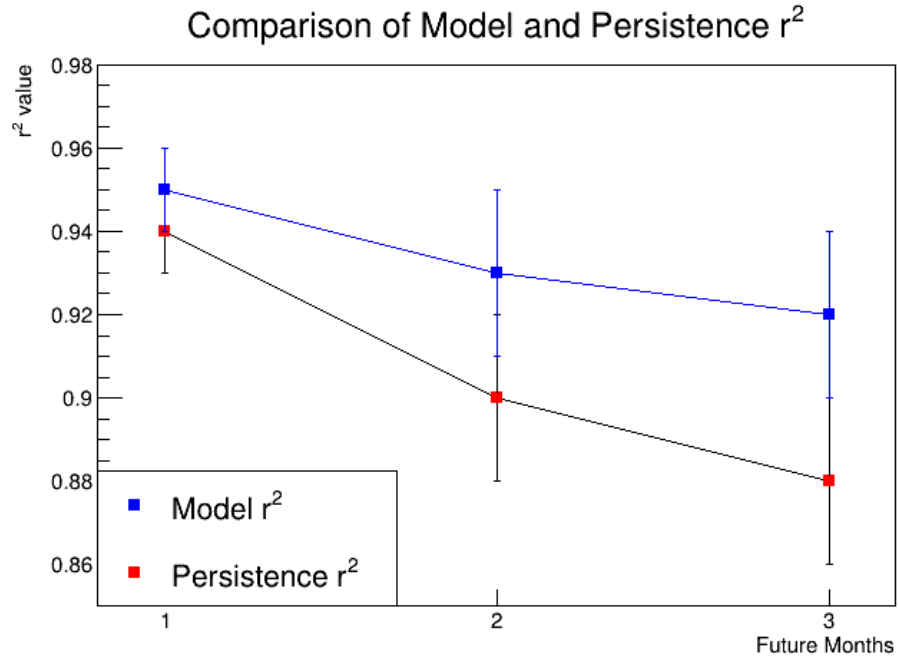
MC Residuals Month T+3, Time Split



(b) T+3

- Residuals of mean squadron MC prediction vs. true mean squadron MC values for months (a) T+1 and (b) T+3. In both cases, the mean of the gaussian distribution is consistent with zero and the χ^2 is very close to unity.

Figure 2. Squadron MC prediction residuals plots.



- A comparison of r^2 values for the model (blue) and persistence (red) predictions using a five-fold cross-validation. At T+3, the 1σ error bars for the model and persistence do not overlap.

Figure 3. Model performance evaluation, cross-validation with random split.

4. PREDICTION UNCERTAINTY

As described in Section 3.1, the random forest regressor was constructed such that it was effectively a quantile regression forest. This allows for each prediction to have a confidence interval ascribed to it derived using the result of each tree; this technique is described in [8]. For this model, an unweighted approach was employed using the outputs directly from the trained model to find the percentiles. The lower and upper boundaries were set at the fifth and ninety-fifth percentiles, yielding a prediction confidence interval into which 90% of the true values fell. A plot illustrating this is presented in Figure 4.

The prediction confidence intervals are essential for decision makers to understand how much credence to give an individual prediction. Additionally, since the intervals are derived from quantiles, they can be asymmetric. This gives added utility to the model for decision makers who may weigh their concerns for under-prediction and over-prediction for a given metric differently.

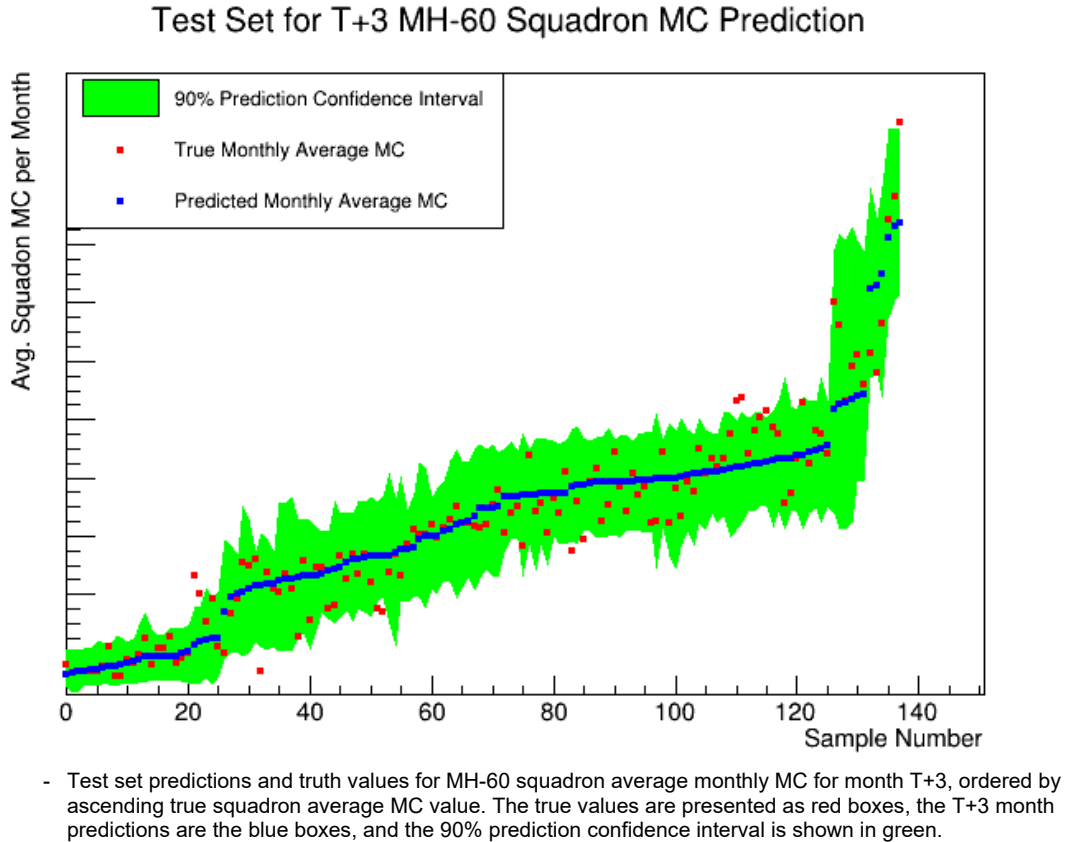
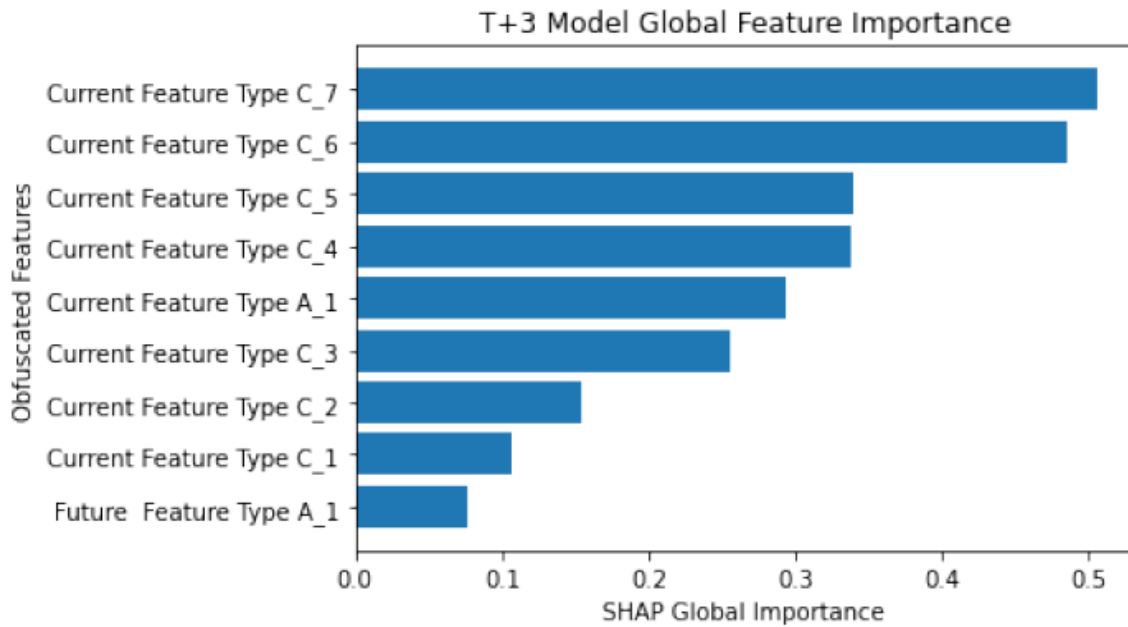


Figure 4. Model performance evaluation, holdout set from datetime-based split.

This page is intentionally blank.

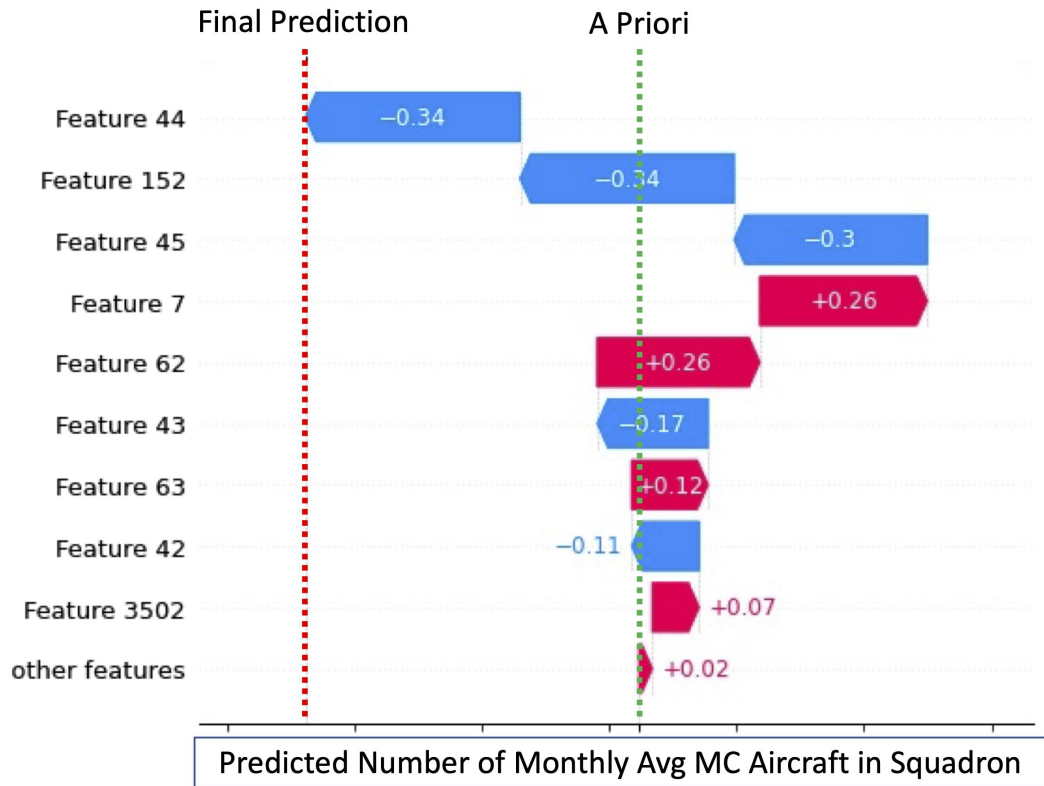
5. XAI

To interrogate the model and determine what features were important for its predictions, the SHapley Additive exPlanations (SHAP) method [9] was applied to the trained models and holdout set predictions. This package uses a technique based in game theory that determines relative feature importances by creating coalitions of features and determining how their manipulation affects the model outputs. During model development, this was used to identify any possible “cheating” features, learn what types of features (e.g., current MH-60 squadron MC status, squadron personnel profile, maintenance actions, et cetera) the model was focusing on, and determine the utility of historical lagged features and future planning information. A global SHAP feature importance plot for T+3 in a datetime-based split is shown in Figure 5. The global feature importance was derived using the mean absolute Shapely values. The model appears to give considerable importance to some historical lagged features as well as future planning information, suggesting that observed historical and expected future performance might be further explored in other MH-60 AI/ML readiness models. Further, this type of analysis has previously lead to the identification of “levers” and “drivers”: features that affect outcomes that either can or cannot be controlled, respectively. The SHAP technique gives explanations at the individual prediction level, resulting in detailed explanations for each squadron. An illustration of this given in Figure 6.



- A global SHAP feature importance plot for T+3 predictions in a datetime-based split.

Figure 5. Global feature importance using SHAP.



- A SHAP waterfall plot of an individual MH-60 prediction for month T+3 in a datetime-based split. How each feature contributes to shifting the model's a priori prediction of the monthly average number of MC MH-60 in a squadron to the final prediction based on this squadron's data is shown.

Figure 6. Individual prediction feature importance using SHAP.

6. CONCLUSIONS

The Digital Aviation Readiness Technology Engine is a flexible paradigm that can be applied to make predictive models for a wide variety of systems. This paper describes the first application of the paradigm to a rotary-wing aircraft, the MH-60 Seahawk. The random forest regressor model can successfully predict the monthly average squadron MC up to three months into the future to within ± 1 aircraft approximately 70% of the time. This has been shown to outperform the baseline persistence model at timestep T+3. Uncertainty and explainable AI have been added to each model prediction, giving added utility to decision makers and subject matter experts (SMEs). Future work to improve the model may focus on using different model architectures, such as Long Short Term Memory Networks, to identify and utilize temporal patterns in the data. The dataset could also be enhanced with features engineered in cooperation with SMEs; further tuning could be performed on the random forest regressor. Finally, more evaluation and testing of this and successive models should be performed before it is deployed for the warfighter.

This page is intentionally blank.

REFERENCES

1. Mattis, J. 2018. "SUBJECT: NDS Implementation - Mission Capability of Critical Aviation Platforms," Tech. rep.
2. Michlin, B., Chang, R., Cruz, R., Duclos, J., Lee, D., Siu, V., and Yetman, C. 2019. "Predicting FA-18 Squadron Readiness and Quarterly Flight Hour Execution Using Machine Learning," Tech. rep., Naval Information Warfare Center Pacific San Diego United States.
3. Michlin, B., Duclos, J., and Williams, G. 2020. "Interpretable Deep Learning Ensembles for Improving Naval Aviation Readiness," *25th International Command and Control Research and Technology Symposium*.
4. Michlin, B., Blackman, L., and Williams, G. 2020. "Improving fa 18 squadron readiness: The resource optimization, allocation, and management model," Tech. rep., Naval Information Warfare Center, Pacific.
5. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E. 2011. "Scikit-learn: Machine Learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830.
6. Meinshausen, N. 2006. "Quantile Regression Forests," *Journal of Machine Learning Research*, vol. 7, no. 35, pp. 983–999, URL <http://jmlr.org/papers/v7/meinshausen06a.html>.
7. Brun, R., Rademakers, F., Canal, P., Naumann, A., Couet, O., Moneta, L., Vassilev, V., Linev, S., Piparo, D., GANIS, G., Bellenot, B., Guiraud, E., Amadio, G., wverkerke, Mato, P., TimurP, Tadel, M., wlav, Tejedor, E., Blomer, J., Gheata, A., Hageboeck, S., Roiser, S., marsupial, Wunsch, S., Shadura, O., Bose, A., CristinaCristescu, Valls, X., and Isemann, R. 2019. "root-project/root: v6.18/02," URL <https://doi.org/10.5281/zenodo.3895860>.
8. Saabas, A. 2015. "Prediction intervals for Random Forests," URL <https://blog.datadive.net/prediction-intervals-for-random-forests/>.
9. Lundberg, S. M. and Lee, S.-I. 2017. "A Unified Approach to Interpreting Model Predictions," I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, eds., *Advances in Neural Information Processing Systems 30*, Curran Associates, Inc.

This page is intentionally blank.

INITIAL DISTRIBUTION

84310	Technical Library/Archives	(1)
53629	J. Rorie	(1)
53629	B. Michlin	(1)
71740	J. Duclos	(1)
53629	D. Lee	(1)
53629	G. Williams	(1)

Defense Technical Information Center
Fort Belvoir, VA 22060-6218 (1)

This page is intentionally blank.

REPORT DOCUMENTATION PAGE

*Form Approved
OMB No. 0704-01-0188*

The public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing the burden to Department of Defense, Washington Headquarters Services Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.

PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.

1. REPORT DATE (DD-MM-YYYY) July 2023		2. REPORT TYPE Final		3. DATES COVERED (From - To)	
4. TITLE AND SUBTITLE Readiness Predictions for the MH-60 Seahawk with the Digital Aviation Readiness Technology Engine				5a. CONTRACT NUMBER	
				5b. GRANT NUMBER	
				5c. PROGRAM ELEMENT NUMBER	
				5d. PROJECT NUMBER	
6. AUTHORS Jamal Rorie, PhD Benjamin Michlin, PhD Josh Duclos Dean Lee Gary Williams NIWC Pacific				5e. TASK NUMBER	
				5f. WORK UNIT NUMBER	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) NIWC Pacific 53560 Hull Street San Diego, CA 92152-5001				8. PERFORMING ORGANIZATION REPORT NUMBER TR 3315	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) The NIWC Pacific Naval Innovative Science and Engineering (NISE) Program 53560 Hull Street San Diego, CA 92152-5001				10. SPONSOR/MONITOR'S ACRONYM(S) NISE	
				11. SPONSOR/MONITOR'S REPORT NUMBER(S)	
12. DISTRIBUTION/AVAILABILITY STATEMENT Approved for public release. Distribution is unlimited.					
13. SUPPLEMENTARY NOTES This is a work of the United States Government and therefore is not copyrighted. This work may be copied and disseminated without restriction.					
14. ABSTRACT The Digital Aviation Readiness Technology Engine (DARTE) paradigm, proven highly successful with fixed-wing aircraft such as the F/A-18 Super Hornet and E-2 Hawkeye, has had the mission capable (MC) model expanded to rotary wing aircraft for the first time with predictions for the MH-60 Seahawk. Readiness datasets for the MH-60 are explored and artificial intelligence / machine learning (AI/ML) models are created to predict future MC aircraft in squadrons. New challenges, including different squadron structures and deployment profiles, are addressed. Additionally, new insights are accessible with the addition of new readiness datasets and interrogation of model predictions with explainable AI (XAI).					
15. SUBJECT TERMS Readiness of critical aviation platforms; squadron structures; deployment profiles; mission Capable (MC) aircraft in squadrons					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Dr. Jamal Rorie
U	U	U			19b. TELEPHONE NUMBER (Include area code) 619-990-6543
			SAR	30	

This page is intentionally blank.

This page is intentionally blank.

Approved for public release. Distribution is unlimited.



Naval Information Warfare Center (NIWC) Pacific
San Diego, CA 92152-5001