Defence Science Journal, Vol. 75, No. 1, January 2025, pp. 35-43, DOI : 10.14429/dsj.20014

Predictive Factor Analysis of Air-to-Air Engagement Outcomes Using Air Combat Manoeuvring Instrumentation Data

Gyuejeong Lee[#], Yong-hwan Kim [$], and Daeyoung Choi[!,*]

[#]*SAKAK Co., Ltd., Seoul - 04147, Republic of Korea*
[$]*Defense Acquisition Program Administration, Gwacheon - 13809, Republic of Korea*
[!]*Division of Artificial Intelligence and Data Science, The Cyber University of Korea, Seoul - 02708, Republic of Korea*
[*]*E-mail: choidy@cuk.edu*

ABSTRACT

This study presents a novel predictive factor analysis of air-to-air engagement outcomes using a decade of air combat manoeuvring data (2009-2019) from the Air Combat Manoeuvring Instrumentation (ACMI) system of the Republic of Korea Air Force (ROKAF). The objective was to construct and evaluate an air-to-air combat hit prediction model using the ACMI system data to identify the critical factors influencing engagement outcomes. This methodology encompasses data preprocessing, feature engineering, binary classification model development, and model interpretation. This study utilises 17 features, including the attitude and speed of both aircraft, along with five additional features derived from the domain knowledge of the relative positions of the two aircraft. Four machine-learning algorithms were employed: logistic regression, random forest, XGBoost, and CatBoost. The best-performing model achieved an accuracy of 83.0 %, noticeably outperforming the baseline at 76.2 %. The analysis revealed that positional information is more crucial than attitude information in predicting engagement outcomes, with the spatial separation between aircraft emerging as the most influential factor. This study showcasings a standard procedure for utilising ACMI system data and demonstrating the effectiveness of machine learning in analysing air combat data.

Keywords: Air combat manoeuvring instrument (ACMI); Air-to-air engagement; Machine learning; Air-to-air combat hit-prediction model

1. INTRODUCTION

Air superiority is essential in modern warfare[1-3]. Air superiority refers to controlling the battlefield sky against an enemy. Once air superiority is achieved, friendly forces, including ground forces, can manoeuvre without prohibitive interference from enemy forces[4,5]. Air combat is a tactical method used to achieve air superiority, and various studies have been conducted to improve its efficiency[6-9]. In this study, we focus on the critical factors of air combat against an enemy's aerial vehicle regarding Air Combat Manoeuvres (ACM).

Regarding ACM, it is essential to develop effective combat tactics and train fighter pilots to improve the win rate in air-to-air combat. However, due to costs, the use of fighters and weapons for developing or evaluating tactics and training or testing pilot skills is limited[10]. Thus, air-to-air combat training is mostly conducted in virtual environments, and the development of precise ACM performance measurements is becoming increasingly important to ensure the reliability of air combat tactics and pilot skills in real-world scenarios.

Existing research approaches to ACM performance measurements mainly focus on combining analytical and empirical methodologies to develop appropriate measurement structures and algorithms[11]. Candidate measurements such as positional advantage and weapon events have been developed based on the state information of both aircraft and weapons, and various studies have utilised these candidates[12-17]. Waag[18], *et al.* proposed a composite measure to predict engagement outcomes during ACM. Krusmark[12], *et al.* assessed the effectiveness of the traditional Grade sheet used to measure air-combat performance. ARAR[19], *et al.* proposed a flexible rule-based framework for a pilot performance analysis.

However, while the utility and effectiveness of both simulation systems and ACM performance measurements have been demonstrated regarding training fighter pilots and developing air combat tactics, more debate still needs to be had on their reliability and validity in real-world environments[20-21]. Balcerzak[22], *et al.* insisted that there was a shortage of research demonstrating the validity of simulation systems, citing the case of civilian aircraft, and that it was more apparent whether the skills learned in simulations were appropriately applied to actual flights. This debate has significant implications for the military domain. Therefore, providing feedback based on actual manoeuvring track data analysis is essential for calibrating measurements developed in a virtual environment. However, a statistical approach to ACM based on actual data has rarely been studied in this domain because acquiring the actual manoeuvring data of an aircraft is limited because of cost and safety concerns.

Received : 26 March 2024, Revised : 15 September 2024
Accepted : 04 November 2024, Online published : 10 January 2025

Air Combat Manoeuvring Instrumentation (ACMI) systems may be an alternative to resolve these limitations. An ACMI system records in-flight data, such as positional information, aircraft state, and weapon events, using pod devices attached to the aircraft, and the recorded data are used for debriefing. The system consists of aircraft pods and a ground system. ACM data are transmitted from the pod to the ground system for recording, displaying, and debriefing[23]. In addition, these data have been consistently accumulated and managed for over a decade. Thus, given the various attributes and quantities of ACMI data, they can be used in data-driven research[24-25].

Motivated by the need for more realistic and data-driven analyses of air combat engagements, this study presents a comprehensive study based on extensive real-world ACMI data from training engagements. Our objectives are threefold: First, to demonstrate a standard procedure for utilising ACMI system data, encompassing feature extraction, selection, and effective modelling of a hit-prediction problem. Second, an air-to-air engagement hit prediction model was constructed using machine learning algorithms, which allowed us to determine the most dominant components of the ACM in deciding engagement outcomes. Third, interpretable machine-learning techniques were applied to rank the key factors for successful engagement. We analyze feature importance using correlation coefficients, feature importance scores, and SHAP (SHapley Additive exPlanations) values[26]. This approach also allowed us to validate conventional methods, differentiating our work from previous studies that relied primarily on simulated or limited flight test data.

The ACMI data are provided by the Republic of Korea Air Force (ROKAF) for research purposes only and are not publicly accessible.

The remainder of this paper is organized as follows. Section 2 describes the problem definition and data. Sections 3 and 4 demonstrate the results of feature engineering and the analysis details, respectively, followed by a discussion and conclusion in Section 5.

2. PROBLEM DEFINITION AND DATA

According to the ROKAF training protocol, air-to-air combat training can be divided into the five categories listed in Table 1. This study only focused on the BFM training procedure. Let BLUE be a fighter of friendly forces and RED

Table 1. Categories of air-to-air combat training

Category	Description
BFM	Basic Fighter Manoeuvring (Most basic form) - Two fighters train together (attacker, defender)
ACM	Air Combat Manoeuvring (Advanced BFM) - Two fighters attack or defend against a RED
ACT/ DACT	Air combat tactics/dissimilar air combat training air-to-air combat without prior agreements between BLUE and RED. (2:2, 2:4, 4:4, 4:2, etc.)
TI/TIN	Tactical intercept/tactical intercept night capture enemy aircraft using fighter radar with the assistance of air traffic control
WA	Weapon to air fighter practices gun firing.

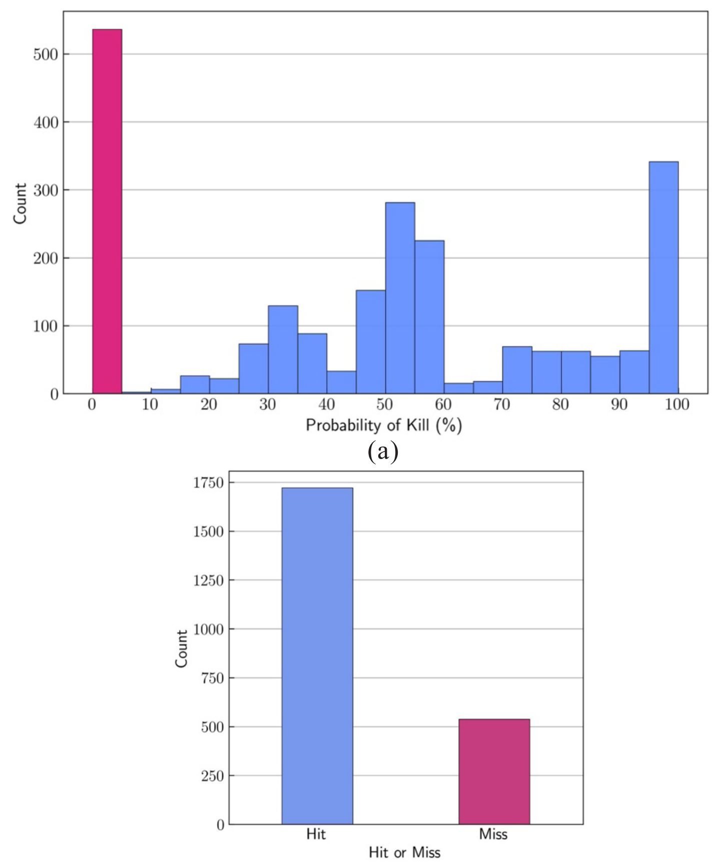

Figure 1. Distributions of Attributes. (a) Probability of kill (PK) ranges between 0 and 1; and (b) 'Hit' and 'Miss' are defined by whether the probability of kill is greater than 0.

Table 2. Attributes of the data

Attribute		Type	Description	Unit
BLUE	RED			
B_xpos	R_xpos	Position	X position coordinate	degree
B_ypos	R_ypos	Position	Y position coordinate	degree
B_zpos	R_zpos	Position	Z position coordinate	m
B_roll	R_roll	Attitude	Rotation around the front-to-back axis	radian
B_pitch	R_pitch	Attitude	Rotation around the side-to-side axis	radian
B_yaw	R_yaw	Attitude	Rotation around the vertical axis	radian
B_aoa	R_aoa	Attitude	Angle between the oncoming air and a reference line on the aircraft	radian
B_speed	R_speed	Kinetic energy	Speed of an aircraft	Mach
B_g,	R_g,	Kinetic energy	Gravity of an aircraft	G

be an adversarial fighter in an air combat training scenario. BLUE and RED are the same type of fighter, F-16, who engage in Within-Visual Range (WVR) combat. BLUE fires AIM-9 IR (infrared) tracking-guided air-to-air missiles to shoot down RED[27]. During training, the ACMI pods collected the maneuvering data of both aircraft, except for the RED probability of kill (PK) value. The PK value, which represents the extent to which BLUE's missile damages RED and ranges from 0 to 1, was calculated internally using the ACMI system. This calculation method has not yet been publicly disclosed. Thus, this study assumed that the PK value calculated by the system adequately reflects the damage to the actual air-to-air engagement.

Based on maneuvering data and PK values, we formulate the hit-prediction model to predict a 'Hit' or 'Miss' from the maneuvering and weapon event data of BLUE and RED. The '0' PK value indicates 'Miss,' which means no damage to RED, and the others are converted to 'Hit,' which means sufficient damage to RED. The distribution of PK values and the distribution of 'Hit' and 'Miss' are shown in Fig. 1.

The data for training the hit prediction model were obtained from the ACMI system operated by the ROKAF, where the collection period was from 2009 to 2019. To prepare the data, we applied several pre-processing steps. First, we addressed data quality issues by removing outliers and missing data points, which often result from the high-speed data acquisition inherent to the ACMI system. Next, data consistency was ensured by standardizing the units of speed and angle across all attributes. However, we did not perform data normalization because the machine-learning algorithms employed were designed to appropriately handle varying scales of input features. After pre-processing, the dataset contains 2,258 instances corresponding to 2,258 missile launches (hits or misses). Of the total, 1,721 instances were labeled as 'Hit' and 537 as 'Miss,' yielding a hit ratio of 76.2 % and establishing the baseline performance. Table 2 lists the 18 attributes used in this study.

3. FEATURE ENGINEERING

In this section, we leverage the domain knowledge extracted from the data to facilitate air-to-air missile hit predictions. We performed feature extraction by focusing on identifying pertinent features. Although the original attributes in the dataset alone may be sufficient for missile hit prediction, extracting additional features can enhance the predictive performance of machine-learning algorithms. To conclude this section, we examined the correlations to ascertain the relationship between the features and missile hits.

3.1 Feature Extraction

In this study, domain knowledge was employed to extract five features. Here, domain knowledge refers to the specific methodology of BFM used in air-to-air combat, which provides insight into attacking adversaries. In BFM air combat scenarios, BLUE manoeuvres to achieve an optimal position and energy state relative to RED before launching a missile. Based on the methodology concerning relative position, we first considered the differences in three-dimensional spatial distances (BR_dist) and altitudes (BR_alt) between BLUE and RED. These differences were computed from attributes representing the position types, namely, B_xpos, B_ypos, B_zpos, R_xpos, R_ypos, and R_zpos. These features are significant because air-to-air missiles can only be hit within a specific range. Second, energy is divided into potential and kinetic energies, with the

Table 3. Extracted features

Feature	Description	Unit
BR_dist	Distance between blue and red	m
BR_alt	Difference of the altitude of blue and that of red	m
BR_speed	Difference of the speed of blue and that of red	Mach
BR_hca	Angular difference between the heading of blue and that of red	degree
BR_aa	Angle measured from the tail of red to blue	degree

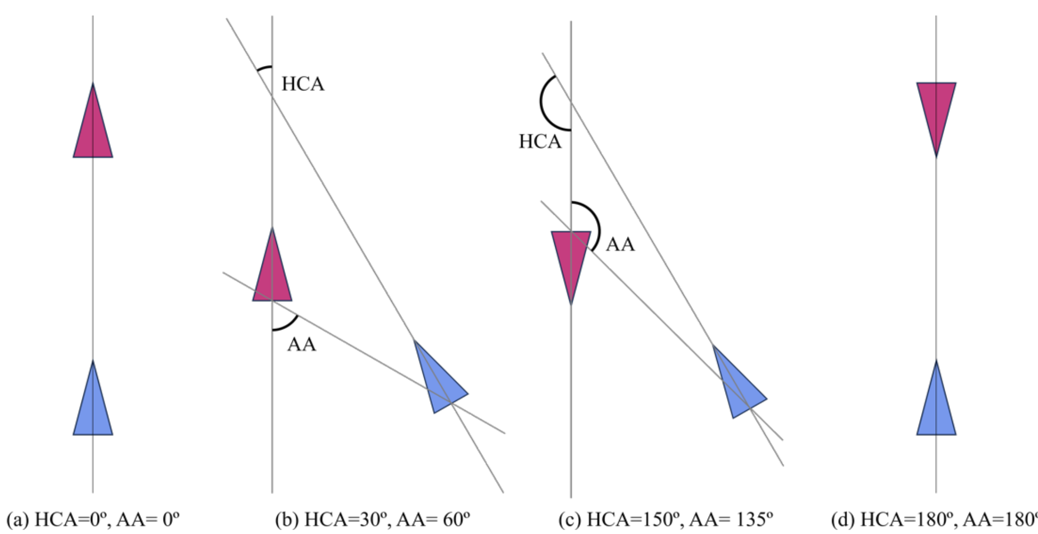

Figure 2. Illustration of HCA and AA. The triangles represent blue and red aircraft. It is shown that (a) and (d) have the same relative position but have different HCA and AA according to the aircraft's heading, which also can be applied to (b) and (c).

altitude difference (BR_alt) and speed difference (BR_speed) playing a pivotal role. Higher altitudes and speeds increase the potential and kinetic energies, allowing aircraft to strategically exchange altitude and speed in three positions based on the BFM principles during air-to-air combat.

In addition to the features based on the relative values above, we further incorporated crucial features considering the BFM. These features encompass the Heading Cross Angle (HCA) and Aspect Angle (AA), as shown in Fig. 2. HCA represents the angular difference in the headings between the two aircrafts, and AA indicates the angle from the tail of RED to the direction of BLUE. Because missiles exhibit higher hit probabilities within specific angular ranges, HCA and AA are recognised as significant features. A summary of the five features mentioned above and their respective units is presented in Table 3. Similarly, in addition to considering the relative positions of the two aircraft, one may also regard the relative values of attributes, such as attitude and gravity, as features.

However, based on domain knowledge, the relative values of the aircraft attitude and gravity have limited significance. By contrast, the absolute values of an aircraft's attitude and gravity are more important than their relative values. Consequently, we refrained from using the relative values of attitude and gravity as additional features.

3.2 Feature Selection

Finally, we obtained 23 features comprising 18 original attributes and five additional features derived through feature extraction. To refine the feature-selection process, we utilized domain knowledge to exclude unnecessary features. Specifically, we omitted six attributes related to the aircraft position. Although three-dimensional terrain information is crucial in air-to-air combat, the data under analysis lacks such terrain data. In addition, utilizing terrain information in model construction may hinder generalization. Ultimately, we obtained a final set of 17 features: B_roll, B_pitch, B_yaw, B_aoa, B_speed, B_g, R_roll, R_pitch, R_yaw, R_aoa, R_speed, R_g, BB_dist, BR_alt, BR_speed, BR_hca, and BR_aa, as shown on the vertical axis in Fig. 3.

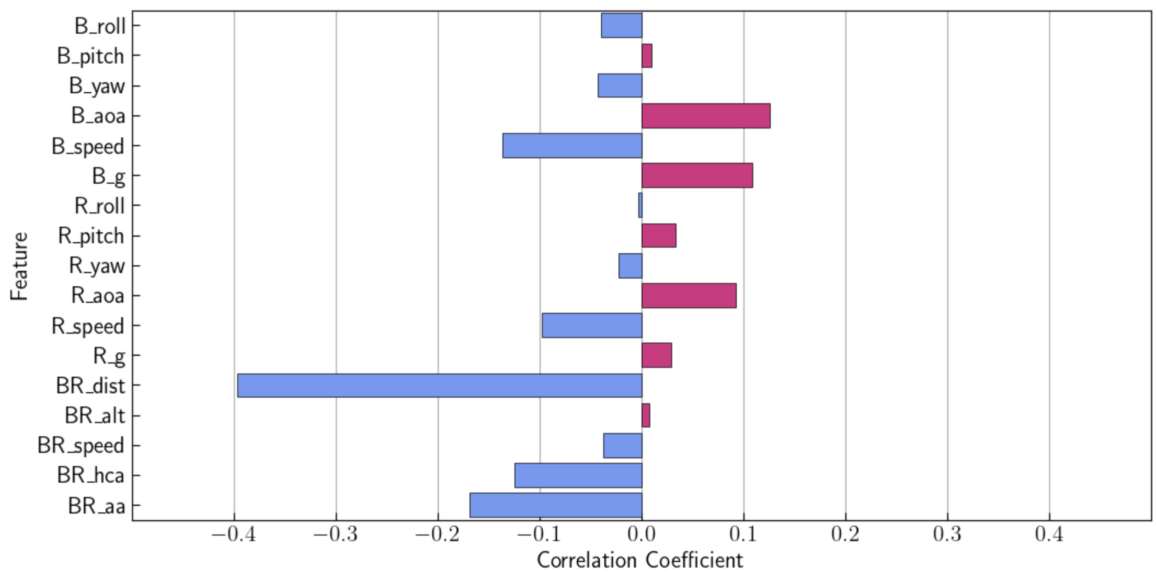

Figure 3. Correlation coefficient between 'Hit' and 17 features. Only one feature (BR_dist) has a significant correlation coefficient, and the features using relative position between BLUE and RED tend to be relatively more significant than the others.

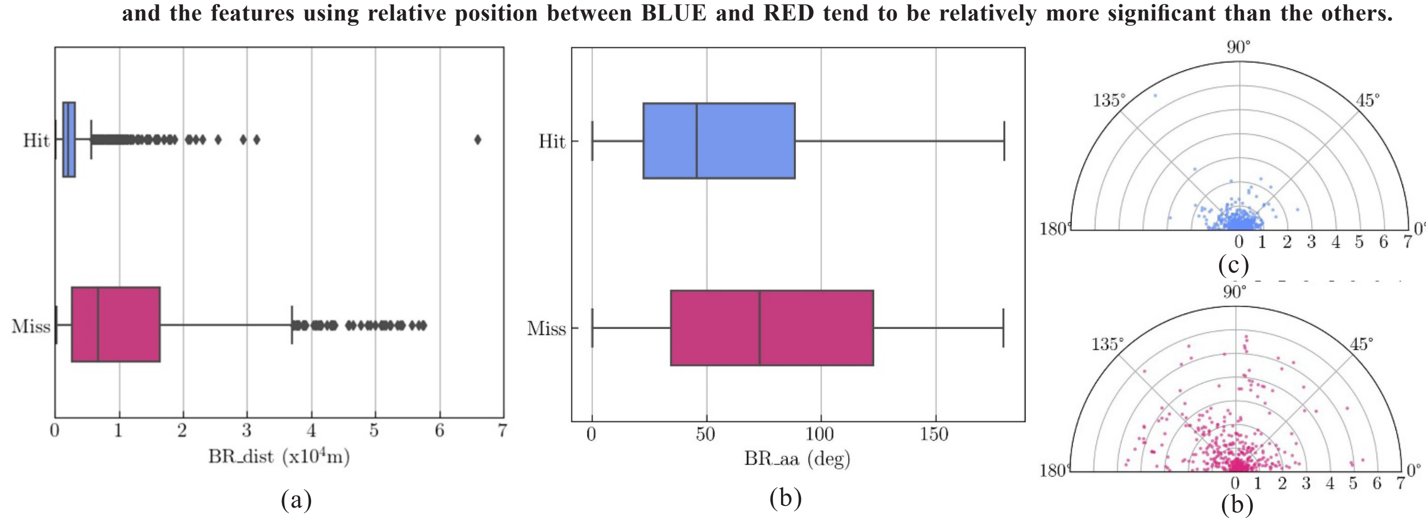

Figure 4. The distribution of missile hit and miss. Blue and red represent 'Hit' and 'Miss.'; (a) boxplots for BR_dist (b) BR_aa.; (c) half-polar plots for BR_dist; and (d) BR_aa.

3.3 Relation Between Missile Hit and Features

We investigated the impact of the features on missile hit prediction using correlation coefficients. The correlations between each feature and the missile hits are shown in Fig. 3. Most of the correlation coefficients were relatively low. Only BR_dist showed a strong linear correlation with the number of missile hits. In addition, features using the relative positions between BLUE and RED, such as BR_hca and BR_aa, tended to be relatively significant.

We also constructed hit prediction models using a single feature. However, the results demonstrated overfitting, indicating poor generalisation. Although the training accuracy ranged from 60 % to 80 % depending on the feature, the test accuracy for most features remained at approximately 50 %. Only BR_dist achieved a test accuracy of approximately 60 %, which is below the baseline performance. This implies that single features alone cannot adequately distinguish between hits and misses and that multiple features must be combined for successful classification.

When combining the features for modeling, it outperforms the baseline, as discussed in Section 4. For instance, Fig. 4 shows two boxplots for BR_dist (left) and BR_aa (middle), and two half-polar plots of BR_dist (upper) and BR_aa (lower) for missile hits and misses. The combination of these two features improves the prediction performance.

4. EXPERIMENT

4.1 Experimental Setting

In this study, we investigate the performance of four machine learning algorithms: logistic regression (LR), random forests (RF), XG Boost (XGB), and Cat Boost (CATB) for classification. Logistic regression served as the baseline model, providing a simple yet effective means of examining the relationships among variables. Random forests, XG Boost, and Cat Boost, while all tree-based ensemble algorithms, differ in their approach: random forests use bagging techniques to create independent trees, XG Boost employs gradient boosting to sequentially improve weak learners, and Cat Boost introduces ordered boosting and processing of categorical features. These algorithms excel in handling tabular data classification problems, each leveraging its unique strengths[28,35-37]. Furthermore, we also explored the performance of Gradient Boosting and Light GBM within the boosting family[38-39]. However, a comparative evaluation revealed that their performances were closely aligned with those of XG Boost and Cat Boost. We also evaluated Multilayer Per Ceptrons (MLPs) that are generally known to underperform on tabular data such as those used in our experiments; indeed, the results were not promising[28]. The training and test data were divided at an 8:2 ratio. The hyperparameter selection for each algorithm was accomplished through Bayesian optimization, and the determination of optimal hyperparameters was achieved through 5-fold cross-validation.

4.2 Performance Result

Table 4 presents the experimental outcomes of the four algorithms using the two feature sets. The first set, labelled ‘All,’ encompasses all 17 features, while the second, ‘Observable,’ is composed of only 12 features, excluding four features that cannot be acquired in (near) real-time from RED. The ACMI data included information from both the BLUE and RED gathered from POD sensors in the training scenarios. However, in actual air-to-air combat cases, BLUE can only access a partial, near real-time stream of RED’s information, with ‘observation’ referring to data obtained through sensors or surveillance systems and transmitted to BLUE almost instantly. Capturing real-time observations of RED’s attitude features (R_roll*, et al.*) and gravity (R_g) from BLUE is difficult. In contrast, positional features (R xpos*, et al.*) and speed (R_speed) are more easily observable and collectible. Thus, the features BR dist, BR_alt, BR_speed, BR_hca, and BR_aa were derived from the observable positional features and speed to construct the model.

Performance assessment was based on accuracy, precision, recall, F1 score, and area under the receiver operating characteristic curve (AUC). Given the class imbalance of the data, it is crucial to interpret the accuracy carefully. Table 4 shows that the performances of the four algorithms are similar, with random forests and XG Boost slightly outperforming the others for the five performance metrics. This indicates a potential link between algorithmic behavior and data characteristics, which can affect performance measures differently.

The performance of the model with the’ Observable’ feature set is nearly on par with using all features, as demonstrated in Table 4. This is consistent with the fact that the RED features have less influence, as reflected by their reduced importance in the evaluation process, as illustrated in Figs. 3, Fig. 5, and Table 5. In summary, given a baseline accuracy of 76.2 %, the performance enhancement with the ‘All’ feature set ranges from approximately 5.9 % to 6.5 %. In contrast, with the ‘Observable’ feature set, it falls within the range of approximately 5.0 % to 6.8 %.

Table 4. Performance comparison of ‘All’ and ‘Observable’ feature sets using four algorithms

Feature set	Algorithm	Accuracy	Precision	Recall	F1 score	AUC
All	LR	0.821	0.815	**0.985**	0.892	0.653
	RF	**0.827**	0.826	0.976	**0.895**	0.676
	XGB	0.821	**0.828**	0.962	0.890	**0.677**
	CATB	0.823	0.823	0.974	0.892	0.670
Observable	LR	0.812	0.807	**0.985**	0.887	0.636
	RF	0.825	0.827	0.971	0.893	0.677
	XGB	**0.830**	**0.833**	0.968	**0.895**	**0.689**
	CATB	0.827	0.828	0.974	**0.895**	0.679

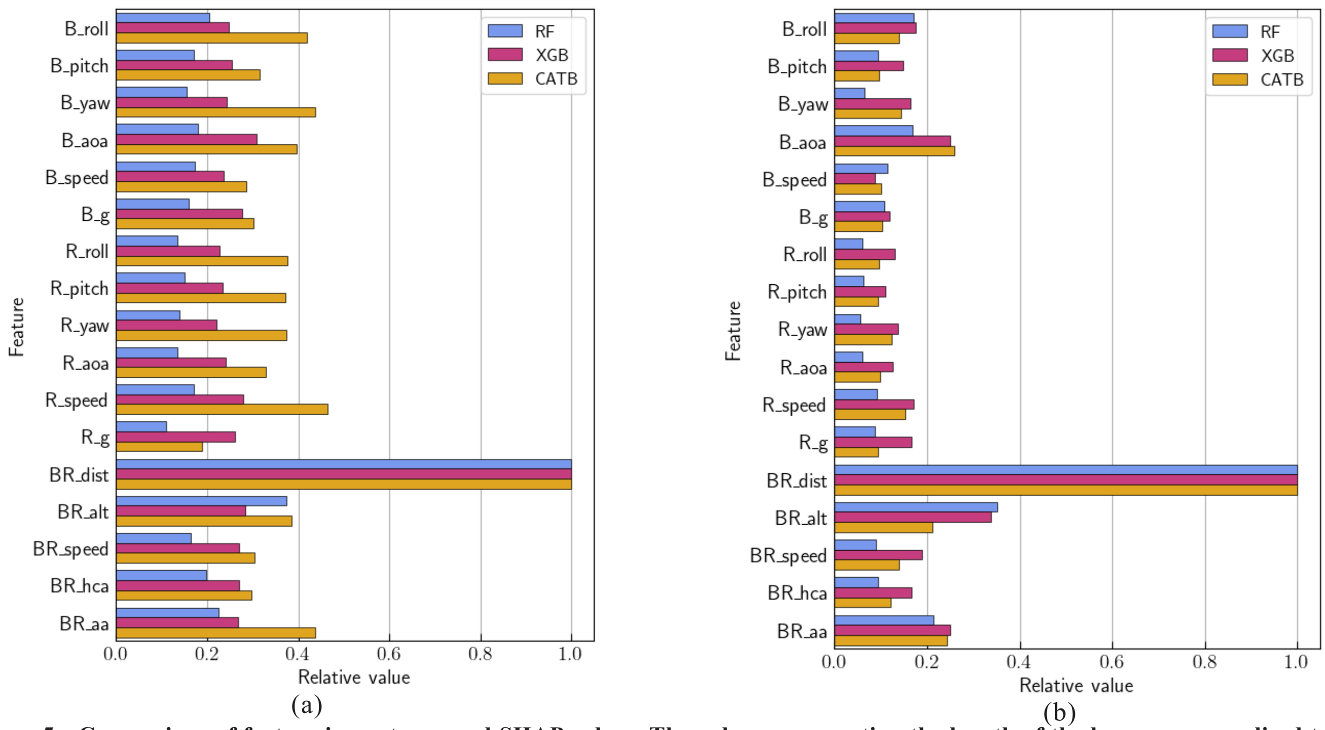

Figure 5. Comparison of feature importance and SHAP values. The values representing the length of the bars are normalised to the largest value per algorithm and measure, thus presented as relative values; (a) Feature importance; and (b) SHAP value.

Table 5. Feature rankings based on both feature importance and SHAP values. Orange, blue, and red represent features belonging to BR, BLUE, and RED, respectively. Each column lists the rankings of the 17 features in descending order, with higher rankings indicating more significant influence.

Rank	Overall	Correlation coefficient	Feature importance			SHAP value		
			RF	XGB	CATB	RF	XGB	CATB
1	BR_dist	BR_dist	BR_dist	BR_dist	BR_dist	BR_dist	BR_dist	BR_dist
2	BR_alt	BR_aa	BR_alt	B_aoa	R_speed	BR_alt	BR_alt	B_aoa
3	BR_aa	B_speed	BR_aa	BR_alt	B_yaw	BR_aa	BR_aa	BR_aa
4	B_aoa	B_aoa	B_roll	R_speed	BR_aa	B_roll	B_aoa	BR_alt
5	R_speed	BR_hca	BR_hca	B_g	B_roll	B_aoa	BR_speed	R_speed
6	B_roll	B_g	B_aoa	BR_hca	B_aoa	B_speed	B_roll	B_yaw
7	BR_hca	R_speed	B_speed	BR_speed	BR_alt	B_g	R_speed	BR_speed
8	B_g	R_aoa	B_pitch	BR_aa	R_roll	B_pitch	BR_hca	B_roll
9	B_yaw	B_yaw	R_speed	R_g	R_yaw	BR_hca	R_g	R_yaw
10	B_speed	B_roll	BR_speed	B_pitch	R_pitch	R_speed	B_yaw	BR_hca
11	R_pitch	BR_speed	B_g	B_roll	R_aoa	BR_speed	B_pitch	B_g
12	BR_speed	R_pitch	B_yaw	B_yaw	B_pitch	R_g	R_yaw	B_speed
13	R_yaw	R_g	R_pitch	R_aoa	BR_speed	B_yaw	R_roll	R_aoa
14	R_aoa	R_yaw	R_yaw	B_speed	B_g	R_pitch	R_aoa	R_roll
15	R_roll	B_pitch	R_roll	R_pitch	BR_hca	R_roll	B_g	B_pitch
16	B_pitch	BR_alt	R_aoa	R_roll	B_speed	R_aoa	R_pitch	R_pitch
17	R_g	R_roll	R_g	R_yaw	R_g	R_yaw	B_speed	R_g

4.3 Feature Importance

For missile hit prediction, we analysed the feature importance and SHAP values of random forests, XG Boost, and Cat Boost to assess the individual significance of the features. The decrease in the average impurity within each tree determines the feature importance values in random forests. In XG Boost and Cat Boost, the feature importance is evaluated by the number of times a feature is used to split the data across all trees. The SHAP values represent the average of all the marginal contributions across all possible coalitions. In Fig. 5, these values are normalised to the largest value per algorithm and measure and are therefore presented as relative values. Despite potential variations owing to algorithmic differences, both the feature importance and the SHAP value consistently emphasize that BR_dist is significantly more influential than the other features, echoing the results in Fig 3.

To identify primary features, we conducted a single-feature ranking analysis using the method outlined by Guyon[34]. We ranked features based on seven normalized measures: feature importance, SHAP values from three tree-based ensemble methods, and correlation coefficients. The overall rank for each feature was determined by averaging the rankings across the seven measures.

The features in Table 5 are ranked and color-coded according to each measure. BLUE, RED, and BR are represented by orange, blue, and red, respectively. The rankings of the 17 features were listed in descending order, with higher rankings indicating a more significant influence. While slight color variations exist across measures, orange features generally dominate the top positions, followed by blue in the middle, and red at the bottom. In the overall ranking, three of the top five positions belonged to BR, whereas five of six features of RED ranked outside the top ten.

5. DISCUSSION AND CONCLUSION

5.1 Comparison with the Conventional Performance Measurement

Our hit prediction model, built on ACMI data, shows promising effectiveness in predicting engagement outcomes. With accuracies ranging from 82.1 % to 83.0 % across the different algorithms and feature sets (Table 4), the model outperformed the baseline accuracy by 76.2 % and by 5.9 % to 6.8 %. This improvement suggests that the model effectively captures the complex dynamics of hit predictions from ACMI data. The results in Fig. 5 and Table 5 show that the positional features of the two fighters were significant for the outcome of the air-to-air engagement. This result is analogous to those of conventional ACM performance-measurement studies. As demonstrated in reference[18,35], positional advantage measurements, such as the all Aspect Manoeuvring Index (AAMI)[15], were the most related to air-to-air engagement outcomes. The AAMI includes a range of fighters.

Experimental results can also be rationally translated using the air combat manoeuvring manual. According to reference[36], BLUE requires the ability of the BFM to enter the RED weapons envelope, and this BFM aims to reduce the range, aspect angle, and angle off to ensure that it can fire weapons at the RED. Based on this analogy and rationality, a data-driven analysis can be used as a verification or refinement methodology for conventional performance measurements in simulation systems.

5.2 Limitation

One of the two limitations of this study is that only the data from the time and from 0.1 s before the missile launch of BLUE was utilised for analysis out of the entire manoeuvring data. Although the information at the missile launch moment is crucial for predicting the hit probability of air-to-air missiles, it is probably necessary to consider the manoeuvres of both BLUE and RED before the launch, as they influence the positioning at the launch moment. Therefore, we must incorporate data from the period preceding a missile launch to extend the applicability of the findings beyond hit prediction and utilize them as feedback information in actual training scenarios. Utilizing such time-series data and employing deep learning algorithms of the RNN family, such as Long Short-Term Memory (LSTM) or Gated Recurrent Unit (GRU), could potentially enhance both predictive performance and interpretability[37-42].

The second is the inherent limitation caused by the use of the ACMI system. While manoeuvring information is acquired from the pods, weapon events, including engagement outcomes, are simulated by the ACMI system. Therefore, the data utilized can be regarded as partially simulated data and analyzed with respect to hit probability.

5.3 Contribution and Future Work

This study demonstrates a standard procedure for utilising ACMI system data, encompassing feature extraction, selection, and effective modelling of a hit-prediction problem. By employing interpretable machine learning techniques, we developed an accurate predictive model and uncovered the most influential factors affecting air-to-air engagement outcomes. This approach bridges the gap between data-driven analysis and traditional air combat performance metrics, thereby offering valuable insights for tactical development and training.

In future work, a refinement model design for fine-tuning the parameters of conventional performance measurements using a data-driven analysis can be suggested. In addition, building an enhanced hit prediction model can be recommended using the RNN family algorithm to exploit the time-series features of the ACMI data. Finally, a multimodal hit prediction model can be proposed for development. Various aspects of air-to-air engagement can be analyzed to train the hit-prediction model using different types of information, such as the aircraft state or pilot information.

REFERENCES

1. Saunders, R. & Souva, M. Air superiority and battlefield victory. *Res. & Politics*, 2020, **7**(4). doi: 10.1177/2053168020972816
2. Jones, R.D. Israeli Air Superiority in the 1967 Arab-Israeli War: An Analysis of Operational Art. Naval War Coll Newport Ri Joint Military Operation Dept, Technical Report No. ADA311683. June 1996.
3. Keaney, T.A. & Cohen, E.A. Gulf war air power survey: Summary Report Vol. 2. Office of the Secretary of the Air Force, USA, 1993. 276 p.
4. AAP-06: NATO glossary of terms and definitions. North Atlantic Treaty Organization (NATO), 2021.
5. Meilinger, P.S. 10 Propositions regarding air power. Air Force History and Museums Program, USA, 1995. 86 p.
6. Dwivedi, P.P. & Sharma, D.K. Selection of combat aircraft by using shannon entropy and VIKOR method. *Def. Sci. J.*, 2023, **73**(4). 411-419

doi: 10.14429/dsj.73.17996

7. Jain, S.K.; Jain, T.K.; Shanker, S. & Srivastava, S. Rapid and accurate INS transfer alignment for air launched tactical missile using kalman filter. *Def. Sci. J.*, 2023, **73**(4) 496-502.
doi: 10.14429/dsj.73.18435
8. Lee, G.; Kwon, J.; Kwak, N. & Yang, E. Robust measurement validation for radar target tracking using prior information. *IET Radar, Sonar & Navigation*, 2019, **13**(10) 1842-1849.
doi: 10.1049/iet-rsn.2019.0019
9. Lee, G.; Lee, S.; Kim, K. & Kwak, N. Probabilistic track initiation algorithm using radar velocity information in heavy clutter environments. *In* 15th European Radar Conference(EuRAD) IEEE 2018, Madrid, Spain, 2018.
doi: 10.23919/EuRAD.2018 .8546666
10. Thurman, R.A. & Mattoon, J.S. Virtual reality: Toward fundamental improvements in simulation-based training. *Edu. Technol.*, 1994 , **34**(8), 56-64.
11. Kelly, M.J.; Lee Wooldridge, A.; Hennessy, R.T. & Reed, J.C. Air combat manoeuvring performance measurement. Proceedings of the Human Factors Society Annual Meeting. 1979, **23**(1), pp. 324-328.
doi: 10.1177/107118137902300183
12. Krusmark, M.; Schreiber, B.T. & Bennett, W. The effectiveness of a traditional gradesheet for measuring air combat team performance in simulated Distributed Mission Operations. United States Air Force Research Laboratory, Scientific and Technical Information No. AFRL-HE-AZ-TR-2004-0090. May 2004.
13. Van Horn, H.J. Can simple models predict air combat results? *J.Aircraft*, 2011,**48**(2), 652-659.
doi: 10.2514/1.C031202
14. Mansikka, H.; Virtanen, K.; Harris, D. & Salomäki, J. Live–virtual–constructive simulation for testing and evaluation of air combat tactics, techniques, and procedures, Part 1: Assessment framework. *J. Def. Model. Simul.*, 2021, **18**(4), 285-293.
doi: 10.1177/ 1548512919886375
15. McGuinness, J.; Forbes J.M. & Rhoads J.E. Air combat manoeuvring performance measurement system design. Williams Air Force, Technical Report No. AFHRL-TP-83-56. Oct 1984.
16. Breidenbach, S.T.; Clavarelli, A.P. & Sievers, R. Measurement methods and metrics for aircrew assessment during close-in air-to-air combat. Naval Training Systems Center, Naval Air Systems Command, Technical Report No.ADA1168066. September 1985.
17. Rajabally, E.; Valiusaityte, I. & Kalawsky, R. Aircrew performance measurement during simulated military aircrew training: A review. *In* AIAA Modeling and Simulation Technologies Conference, Chicago, USA, 2009.
doi:10.2514/6.2009-5829
18. Waag, W.L.; William B.R. & Jeffrey L.L. Development of a composite measure for predicting engagement outcome during air combat manoeuvring. Armstrong Laboratory, Air Force Systems Command, Technical Report No.AL-TR-1992-0002, May 1992.
19. ARAR, Ö.F. & Ayan, K.A flexible rule-based framework for pilot performance analysis in air combat simulation systems. *Turkish J. Electric. Engin. Comput. Sci.*, 2013, **21**(8), 2397-2415.
doi: 10.3906/elk-1201-50
20. Saastamoinen, K. & Maunula, K. Usefulness of flight simulator as a part of military pilots training–case study: Grob G 115E. *Procedia Comput. Sci.*, 2021, **192**, 1670-1676.
doi: 10.1016/j.procs. 2021.08.171
21. Cross, J.; Boag-Hodgson, C.; Ryley, T.; Mavin, T.J. & Potter, L.E. Using extended reality in flight simulators: a literature review. *IEEE Transact. Visual. Comput. Graphics*, 2022, **29**(9), 3961-3975.
doi: 10.1109/TVCG.2022.3173 921
22. Balcerzak, T. & Katarzyna, K. Flight simulation in civil aviation: Advantages and disadvantages. Revista europea de derecho de la navegación marítima y aeronáutica, 2018, **35**, 35-68.
23. Panarisi, M.T. A comparative analysis of internal and external solutions to provide air combat manoeuvring instrumentation functionality. Air University Press, USA, 2001. 88 p.
24. Dantas, J.P.; Costa, A.N.; Medeiros, F.L.; Geraldo, D.; Maximo, M.R. & Yoneyama, T. Supervised machine learning for effective missile launch based on beyond visual range air combat simulations. *In* 2022 Winter Simulation Conference (WSC) IEEE, Singapore, 2022.
doi: 10.1109/WSC 57314.2022.10015384
25. Kim, D.; Arjit, S. & Rhea P.L. Data-enhanced dynamic flight simulations for flight performance analysis. *Aerospace Sci. and Technol.*, 2022, **121**, 1-28.
doi: 10.1016/j.ast.2022.107357
26. Lundberg, S.M. & Lee, S. A unified approach to interpreting model predictions. Advances in neural information processing systems(NIPS 2017), CA, USA, 2017.
27. https://www.af.mil/About-Us/Fact-Sheets/Display/Article/104557/aim-9-sidewinder/,
(Accessed on 25 March 2024).
28. Borisov, V.; Leemann, T.; Seßler, K.; Haug, J.; Pawelczyk, M. & Kasneci, G. Deep Neural Networks and Tabular Data: A Survey. IEEE Transact. *Neural Networks and Learning Syst.*, 2022, Early Access, 1-21.
doi: 10.1109/ TNNLS.2022.3229161
29. Breiman, Leo. Random forests. Machine learning, 2001, 45. 5-32.
doi: 10.1023/A:1010933404324
30. Chen, T. & Carlos, G. Xgboost: A scalable tree boosting system. Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining, NY, USA, 2016.
doi: 10.1145/2939672.2939785
31. Prokhorenkova, L.; Gusev, G.; Vorobev, A.; Dorogush, A.V.; & Gulin, A. CatBoost: Unbiased boosting with categorical features. Advances in neural information processing systems(NIPS 2018). Montreal, Canada, 2018.

32. Friedman, J.H. Stochastic gradient boosting. *Comput. Stat. Data Analysis*, 2002, **38**(4), 367-378. doi:10.1016/S0167-9473(01)00065-2
33. Ke, G.; Meng, Q.; Finley, T.; Wang, T.; Chen, W.; Ma, W. & Liu, T.Y. Lightgbm: A highly efficient gradient boosting decision tree. Advances in neural information processing systems(NIPS 2017) , CA, USA, 2017.
34. Guyon, I. & Elisseeff, A. An introduction to variable and feature selection. *J. Machine Learn. Res.*, 2003, **3**, 1157-1182. doi: 10.55 55/944919.944968
35. Kelly, M.J. Performance measurement during simulated air-to-air combat. *Human Factors*, 1988 , **30**(4), 495-506 doi: 10.1177/001872088803000 410
36. Shaw, R.L. Fighter Combat: Tactics and Manoeuvring. Naval Institute Press Annapolis, USA, 1985. 428 p.
37. Zhang, J.D.; Yu, Y.F.; Zheng, L.H.; Yang, Q.M.; Shi, G.Q. & Wu, Y. Situational continuity-based air combat autonomous manoeuvring decision-making. *Def. Technol.*, 2022, **29**, 66-79. doi: 10.1016/j.dt.2022.08.010
38. Shi, Z.; Xu, M.; Pan, Q.; Yan, B. & Zhang, H. LSTM-based flight trajectory prediction. *In* 2018 International Joint Conference on Neural Networks (IJCNN) IEEE, Rio de Janeiro, Brazil, 2018. doi: 10.1109/IJCNN.2018.8489734
39. Bae, J.H.; Jung, H.; Kim, S.; Kim, S. & Kim, Y.D. Deep reinforcement learning-based air-to-air combat manoeuvre generation in a realistic environment. *IEEE Access*, 2023, **11**, 26427-26440. doi: 10.1109/ACCESS.2023.3257849
40. Wang, X.; Wang, Y.; Su, X.; Wang, L.; Lu, C.; Peng, H. & Liu, J. Deep reinforcement learning-based air combat manoeuvre decision-making: literature review, implementation tutorial and future direction. *Artificial Intell. Rev.* , 2024, **57**(1), 1. doi: 10.1007/ s10462-023-10620-2
41. Xi, Z.; Kou, Y.; Li, Z.; Lv, Y. & Li, Y. An air combat manoeuvre pattern extraction based on time series segmentation and clustering analysis. *Def. Technol.*, 2024, **36**, 149-162. doi: 10.1016/j.dt.2023.11.010
42. Wang, H.; Zhou, Z.; Jiang, J.; Deng, W. & Chen, X. Autonomous air combat manoeuvre decision-making based on PPO-BWDA, *IEEE Access*, 2024, **12**, 119116 – 119132. doi: 10.1109/ ACCESS.2024.3419889

CONTRIBUTORS

Dr Gyuejeong Lee obtained his PhD in Computer Science from Seoul National University. He currently serves as the leader of the AI/Data Laboratory at SAKAK Co., Ltd. His research interests include: Data science, machine learning, deep learning, and trustworthy artificial intelligence.
In this study, he conceptualised the research framework, conducted a comprehensive literature review, and developed the methodology.

Dr Yong-hwan Kim obtained his PhD in Defense Fusion Engineering from Yonsei University. He currently serves as the Air Force Program Manager at the Republic of Korea Embassy in the United States, bringing his experience as a fighter pilot to the role. His research focuses on: Artificial intelligence and data science applications in the defense sector.
In this study, he contributed to methodology development and was responsible for data collection and curation.

Prof Daeyoung Choi obtained his PhD in Computer Science from Seoul National University (2019). He is currently working as an Assistant Professor in the Division of Artificial Intelligence and Data Science at The Cyber University of Korea. His research interests include: Edge AI, federated learning, and data science.
In this study, he developed the methodology and performed machine learning modeling and data analysis.