

Original Article

<https://doi.org/10.12985/ksaa.2026.34.1.066>
ISSN 1225-9705(print) ISSN 2466-1791(online)

머신러닝 기반 항공기 사고 기체 손상 심각도에 영향을 미치는 요인 분석

이정열*, 전정환**

Analysis of Factors affecting Aircraft Damage Severity in Aircraft Accidents based on Machine Learning

Jeoungyeoul Lee*, Jeonghwan Jeon**

ABSTRACT

This study examines aircraft damage severity by framing it as a machine learning-based multiclass classification problem. Using NTSB accident records, we constructed a multidimensional dataset that integrates aircraft, pilot, airport, and meteorological information. XGBoost was applied under class-imbalanced conditions, with recall as the primary performance metric and SHAP analysis used to enhance model interpretability. The results indicate that aircraft damage severity is driven by multiple interacting factors, particularly spatial airport-related variables. Among these factors, airport elevation and spatial proximity measures were more influential than traditional predictors such as aircraft age or general weather conditions. These findings underscore the need for a shift toward proactive aviation safety management based on aircraft damage-oriented risk assessment.

Key Words : Aircraft Damage Severity(항공기 손상 심각도), Machine Learning(머신러닝), XGBoost(엑스지부스트), SHAP(샤프 분석), Aviation Safety(항공 안전), NTSB(국가교통안전위원회)

1. 서론

세계 항공운송 시장은 세계화, 항공 기술의 발전, 저비용항공사(LCC)의 성장, 항공 운임 인하 등의 요인으로 지속적인 성장세를 보이고 있다(Yoon and Son, 2023). 이에 따라 항공기 운항 횟수와 공항 이용 밀도 역시 꾸준히 증가하고 있으며, 특히 주요 국제공항을 중심으로 이착륙이 집중되는 시간대에는 활주로 포화

와 관제 부담이 가중되고 있다(Shin et al., 2024). 이러한 고밀도 운항 환경은 항공 안전 확보의 중요성을 더욱 부각하고 있으며, 항공사고 예방을 위한 체계적인 안전관리의 필요성이 제기되고 있다(Kim, 2020; Yoo et al., 2024).

항공사고는 단일 원인에 의해 발생하기도 하지만, 대부분 인적 요인, 환경 요인, 기계적 요인, 조직·시스템 요인 등이 복합적으로 작용한 결과로 나타난다(Gaur, 2005). 기존의 항공 안전 연구는 사고 발생 이후의 원인 규명과 통계적 분석을 중심으로 발전해 왔다. 특히 인명 피해 규모를 주요 지표로 활용하였으며, 상당수 연구는 사고의 결과를 인명 상해 등급(injury severity) 중심으로 분석하였다(Yoon, 2025). 인명 상해 등급을 포함한 인명 피해 분석은 항공사고 연구에서 가장 중요한 지표 중 하나이지만, 항공기 기체 손상

Received: 15. Jan. 2026, Revised: 25. Jan. 2026,

Accepted: 3. Mar. 2026

* 대한민국 공군 시험평가단

** 경상국립대학교 산업시스템공학과

연락처 E-mail: jhjeon@gnu.ac.kr

연락처 주소: 경상남도 진주시 진주대로 501 (52828)

심각도(aircraft damage severity) 역시 항공사의 직접적인 경제적 손실과 운항 중단에 따른 간접 비용 등으로 이어질 수 있으며, 항공 시스템 전반에도 중대한 영향을 미친다(Silagyi and Liu, 2023; Rosadi et al., 2024). 하지만 국내에서 기체 손상 심각도를 정량적으로 분석한 연구는 상대적으로 제한적이다.

항공기 기체 손상은 사고 당시의 운항 환경, 공항 접근 조건, 기체 성능 특성, 정비 이력, 기상 요인 등 다양한 요인의 상호작용에 의해 결정된다(Rosadi et al., 2024). 이러한 복합적 관계는 전통적인 통계 분석 기법만으로는 충분히 설명하기 어렵다. 최근 항공 안전 분야에서는 대규모 사고 데이터를 활용하여 비선형적 관계와 변수 간 상호작용을 효과적으로 포착할 수 있는 머신러닝 기법의 활용이 확대되고 있다(Silagyi and Liu, 2023; Rosadi et al., 2024; Yoon, 2025). 머신러닝은 사고 발생 가능성뿐만 아니라 사고 결과의 심각도를 예측하는 데에도 효과적으로 적용될 수 있으며, 데이터 기반의 예방적 안전관리 체계로의 전환을 가능하게 한다(Rosadi et al., 2024).

미국 국가교통안전위원회(National Transportation Safety Board, 이하 NTSB)에서는 항공기 사고와 준사고에 대한 체계적이고 신뢰성 높은 데이터를 제공하고 있으며, 이는 항공 안전 연구에서 가장 널리 활용되는 공공 데이터 중 하나이다(Kim et al., 2019). NTSB 데이터에는 사고 발생 위치, 기체 특성, 조종사 정보, 기상 조건, 공항 환경 등 사고와 관련된 다양한 변수가 포함되어 있어, 항공기 사고의 원인과 결과를 다차원적으로 분석하는 데 적합하다(Silagyi and Liu, 2023). 이러한 데이터는 머신러닝 기반 분석을 통해 기존 연구에서 충분히 조명되지 못했던 사고 결과의 구조적 특성을 규명할 수 있는 잠재력을 지닌다.

이에 본 연구는 NTSB 항공사고 데이터를 활용하여 항공기 사고 시 기체 손상 심각도에 영향을 미치는 주요 요인을 머신러닝 기반으로 분석하는 것을 목적으로 하였다. 이를 위해 비선형적 관계와 데이터 불균형 문제에 강점을 지닌 XGBoost(eXtreme Gradient Boosting) 알고리즘을 적용하고, SHAP(shapley additive explanations) 분석을 통해 예측 결과에 대한 해석 가능성을 확보하고자 하였다(Gwon et al., 2025). 본 연구는 인명 상해 등급 중심의 기존 연구를 확장하여 기체 손상 심각도를 결과 변수에 설정하여 항공기 사고 분석의 범위를 확장하여 항공 안전관리 개선을 위

한 실증적 근거의 기초자료를 제시하고자 하였다.

II. 이론적 배경

2.1 항공기 사고의 복합적 특성

항공사고는 단일 요인에 의해 발생하기보다는 인적 요인, 환경 요인, 기계적 요인, 그리고 조직-시스템 요인이 상호작용한 결과로 나타나는 복합적 사건이다(Hong, 2002). 조종사의 판단 오류 및 피로 누적과 같은 인적 요인, 기상 악화 및 공항 접근 환경 등의 외부 조건, 항공기의 성능 저하와 정비 상태 등 기계적 요인, 그리고 관제 체계 및 안전관리 절차의 한계와 같은 구조적 요인이 복합적으로 작용하여 사고로 이어지는 경우가 많다(Hong, 2005). 이러한 특성으로 인해 항공사고는 발생 여부뿐 아니라, 사고 이후 나타나는 결과 역시 단선적으로 설명하기 어렵다.

전통적인 항공 안전 연구에서는 사고의 결과를 주로 인명 피해 규모를 중심으로 평가해 왔다(Yoon, 2025). 인명 상해 등급은 사고의 사회적 파급력과 안전 수준을 직관적으로 보여주는 지표로 활용되어 왔으며, 항공 안전 정책과 규제 체계에서도 중요한 판단 기준으로 작용해 왔다(Kwon and Lee, 2023). 그러나 항공사고의 결과는 인명 피해에 국한되지 않으며, 기체 손상 수준 역시 사고의 심각성을 평가하는 핵심 요소 중 하나이다(Park et al., 2025).

항공기 기체 손상은 사고 이후 항공사의 운영 안정성, 재정적 부담, 공항 운영 효율성, 항공 교통 흐름 등에 장기적인 영향을 미친다(Bae et al., 2016). 경미한 손상은 단기간의 정비로 복구할 수 있지만, 중대 손상이나 기체 전손의 경우 항공기 운항 중단, 노선 축소, 보험 비용 증가 등으로 이어질 수 있다(Bae et al., 2016). 따라서 항공기 사고의 위험을 종합적으로 평가하기 위해서는 인명 상해 등급 분석과 함께 기체 손상 심각도에 대한 정량적 분석이 병행될 필요가 있다.

2.2 기체 손상 심각도

기체 손상 심각도는 항공기 사고 발생 시 기체가 입은 물리적 손상의 정도를 의미하며, 일반적으로 경미 손상(minor), 중대 손상(substantial), 전손(destroyed)의 범주로 구분된다(Amaral et al., 2023). 이러한 손상 수준은 사고 당시 충돌 강도, 항공기 구조 특성, 속도

및 증량, 지면 또는 장애물과의 상호작용, 기상 조건 등 다양한 요인의 영향을 받는다(Xu et al., 2024).

기체 손상 심각도는 사고의 즉각적인 위험뿐 아니라, 항공 시스템의 회복탄력성을 평가하는 지표로도 활용될 수 있다(Li et al., 2024). 유사한 사고 상황에서도 기체 손상 수준이 낮게 나타나는 경우, 이는 기체 설계, 운항 절차, 공항 환경 관리가 일정 수준의 안전 여유를 확보하고 있음을 시사한다(Xiong et al., 2024). 반대로 인명 피해가 없더라도 기체 손상이 심각한 사고는 향후 유사 사고 발생 시 잠재적 위험 증폭될 가능성을 내포한다(Li et al., 2024).

이러한 중요성에도 불구하고 기존 항공 안전 연구에서는 기체 손상 심각도를 주로 사고의 사후적 결과 또는 보조적 정보로 취급되어 왔다(Bae et al., 2016). 일부 연구에서는 착륙 사고나 활주로 이탈 사고를 대상으로 기체 손상 여부를 이진 변수로 분석하거나, 특정 사고 유형에 한정하여 손상 수준을 비교하는 데 제한되는 경향을 보였다(Si et al., 2024). 이러한 접근은 기체 손상 심각도의 결정 요인을 체계적으로 규명하는 데 한계를 지닌다. 따라서 기체 손상 심각도를 다범주 결과 변수로 설정하고, 이에 영향을 미치는 요인을 종합적으로 분석하는 연구가 필요하다.

2.3 항공 안전 연구의 데이터 기반 분석

항공 안전 연구는 오랫동안 사고 발생 빈도 분석, 로지스틱 회귀, 생존 분석 등 전통적 통계 기법에 기반하여 발전해 왔다(Amaral et al., 2023; Liu et al., 2025). 그러나 항공사고 데이터는 희소한 중대 사고 사례와 클래스 불균형, 변수 간 비선형적 관계를 포함하는 특성을 지니므로, 복잡한 상호작용을 충분히 반영하는 데 한계가 있다(Farhadpour et al., 2024). 특히 사고의 결과 심각도를 다범주로 분류하는 문제에서는 전통적인 통계 모형의 예측력이 제한적으로 나타나는 경우가 많다(Tu et al., 2025).

최근에는 대규모 사고 및 운항 데이터를 활용한 머신러닝 기반 분석이 항공 안전 연구의 새로운 대안으로 평가된다(Yoon, 2025). 머신러닝 기법은 변수 간 복잡한 상호작용과 비선형적 관계를 효과적으로 학습할 수 있으며, 다수의 설명 변수를 동시에 고려한 예측이 가능하다. Random forest, gradient boosting, XGBoost 등 앙상블 기반 알고리즘은 항공사고 예측, 위험 요인 중요도 분석, 이상 징후 탐지 등 다양한 분

야에서 활용되고 있다(Yoon, 2025).

특히 XGBoost는 과적합 제어, 결측치에 대한 내재적 처리 메커니즘, 데이터 불균형 대응 등에서 강점을 지니며, 항공기 사고와 같이 관측 빈도가 낮고 클래스 분포가 불균형한 데이터에 적합한 알고리즘으로 평가되고 있다(Benfaress et al., 2025). 더불어 SHAP와 같은 설명 가능한 인공지능 기법을 적용할 경우, 단순한 예측 정확도를 넘어 각 변수의 기여도를 정량적으로 해석할 수 있어 정책적·실무적 활용 가능성을 높일 수 있다(Benfaress et al., 2025).

선행연구들은 항공사고의 발생 원인과 인명 피해 심각도에 영향을 미치는 요인을 중심으로 다양한 분석을 수행해 왔다(Table 1). 다수의 연구에서 항공기 누적 운항시간, 정비 이력, 기상 조건, 조종사 특성 등이 사고 위험 및 인명 피해 수준과 관련이 있음을 보고하였다. 그러나 이러한 연구들은 주로 인명 피해를 결과 변수로 설정함으로써, 기체 손상과 관련된 구조적 요인을 충분히 조명하지 못했다는 한계를 지닌다.

머신러닝 기반 능동형 모델 개선 피드백 기술을 적용한 보안관제 모델을 제안한 선행연구에서는 운영 중 예측된 유사 이벤트를 군집화하고, 피드백이 필요한 군집을 생성하는 성능 개선 프로세스를 제안하였다(Shin and Jo, 2021). 해당 연구는 세 가지 시나리오 실험을 수행하여 이상 탐지 및 분류 모델의 성능을 검증하였다. 그 결과 탐지율 향상과 오탐률 감소에 따른 대응 시간 단축 효과를 확인하였으며, 불확실한 데이터를 선별함으로써 30% 이상의 성능 향상을 보고하였다.

다수의 의사결정나무를 앙상블로 구성한 머신러닝 모형인 그래디언트 부스팅 분류기 (gradient boosting classifier, 이하 GBC)와 랜덤 포레스트(random forest, 이하 RF)를 적용해 항공기 사고 데이터에서의 인명 피해 등급과 사고 요인 간의 상관관계를 파악한 선행연구에서는 두 모델의 정확도, 재현율(recall), F1-score 등의 성능 지표를 검증하였다(Yoon, 2025). GBC는 과소 표집 또는 과대 표집 없이 불균형 데이터인 항공기 사고 데이터에서 높은 분류 성능을 보고하였다(Yoon, 2025).

최근 일부 연구에서는 착륙 사고나 특정 사고 유형을 대상으로 기체 손상 분류를 시도하고 있으나(Si et al., 2024), 분석 범위가 제한적이거나 단일 요인 중심의 접근에 머무르는 경우가 많다(Rosadi et al., 2024). 또한 기체 손상 심각도를 종속변수로 설정하

Table 1. Summary of previous studies

선행연구	연구 내용	연구 방법
Yoon (2025)	Injury severity	GBC, RF
Shin & Jo (2021)	System performance improvement model	System performance optimization
Liu et al. (2025)	Dynamic determinants of aviation accidents	Random parameter bivariate probit approach
Silagyi & Liu (2023)	Prediction of severity of aircraft landing accidents	Support vector machine model
Rosadi et al. (2024)	Prediction of classification of aircraft damage	Analysis of algorithms

라도, 변수 중요도에 대한 해석 가능성을 충분히 확보하지 못한 연구도 존재한다(Silagyi and Liu, 2023).

이에 본 연구는 신뢰도 높은 NTSB 항공사고 데이터를 활용하여, 기존의 인명 피해 중심 분석에서 기체 손상 심각도를 포함한 다각적 분석을 수행함으로써 기존 연구의 한계를 보완하고 분석 범위를 확장하는 데 목적이 있다.

III. 연구 설계

3.1 연구 절차

본 연구의 전체 분석 절차는 자료 수집 및 전처리, 변수 정의, 머신러닝 기반 예측 모형 설계, 모형 성능 평가, 변수 중요도 해석의 단계로 구성되었다(Fig. 1). 먼저 항공사고 데이터의 신뢰성과 활용 가능성을 고려하여 신뢰도 높은 공공 데이터를 수집하고, 분석 목적에 부합하지 않는 변수를 제거하고, 결측치를 포함한 관측 변수를 정제하였다(Yoon, 2025). 이후 항공기 기체 손상 심각도를 종속변수로 설정하고, 사고 발생 당시의 기체·조종사·환경·공항 관련 변수를 독립변수로 구성하였다(Rosadi et al., 2024). 머신러닝 모형으로는 XGBoost, GBC, RF 알고리즘을 성능 비교를 통해 최적 모형을 선정하였다. 또한 예측 결과의 해석 가능성을 확보하기 위해 SHAP 분석을 적용하여 각 변수의 기여도를 정량적으로 해석하였다(Benfaress et al., 2025).

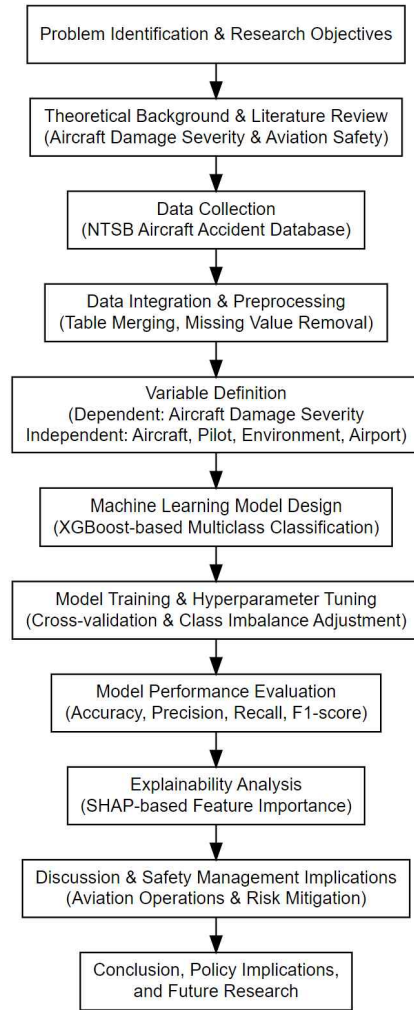


Fig. 1. Research process flowchart

3.2 자료 수집

항공기 사고의 원인을 체계적으로 분석할 수 있는 데이터는 상대적으로 제한적이다. 이러한 제약에도 불구하고, 본 연구에서는 신뢰성과 활용도가 높은 NTSB의 항공기 사고 데이터를 모형 구축 및 항공기 사고 요인 분석에 활용하였다. 본 연구에서 활용한 NTSB 데이터는 2008년 1월부터 2025년 12월까지 항공기 사고(incident, 이하 ACC) 및 준사고(incident, 이하 INC)와 관련하여 사고 요인이 될 수 있는 다양한 변수와 사고별 특성을 포함하고 있다. NTSB에서 제공하는 avall.mdb 파일은 20개의 테이블로 구성되며, 이 가운데 항공기 사고와 관련이 깊은 것으로 평가되는 3개

의 테이블인 events, flight_crew, aircraft를 병합하여 하나의 통합 데이터로 정리하여 분석하였다(Kim et al., 2019).

전처리 전 통합 데이터에는 총 36,742건의 항공기 사고 기록이 있었다. 이 데이터셋의 ev_type에는 사고에 해당하는 ACC가 34,256건(93.2%), 준사고에 해당하는 INC가 2,486건(6.8%) 포함되어 있다. 준사고(INC)는 기체 손상이 경미한 경우로, 다범주 손상 심각도 분류 모형의 학습 및 평가 과정에서 클래스 분류 왜곡을 초래할 가능성이 있다(Paek et al., 2022). 이에 본 연구에서는 준사고는 제외하고 전체 사건의 93.2%를 차지하는 사고(ACC) 34,256건을 최종 분석을 위한 기본 데이터셋으로 선정하였다.

3.3 변수 정의 및 전처리

본 연구는 NTSB 데이터에서 기체 손상 수준 정보인 damage 변수를 기반으로 경미 손상(MINR), 중대 손상(SUBS), 전손(DEST)의 3가지 등급의 기체 손상 심각도를 종속변수로 정의하였다(Rosadi et al., 2024).

먼저 해당 사고의 구분을 위한 식별자 ev_id를 기준으로 events, flight_crew, aircraft를 통합하였다. 통합 파일은 모두 191개의 변수를 포함한다. 이 변수들 가운데에는 상당한 결측치를 포함하는 변수들이 존재한다. 결측 비율이 50%를 초과하는 경우, 정보 대비 노이즈가 급격히 증가해 무리한 보간을 하는 경우 잘못된 패턴을 학습할 위험이 크다(Ahmad et al., 2024). 본 연구에서도 결측값이 50% 이상인 변수를 제거하였다. 그 결과 변수는 128개로 정제되었다.

다음으로 기체 손상 심각도 예측과의 관련성이 낮거나 분산이 거의 없어 과적합과 잡음을 유발할 가능성이 큰 변수를 제거하였다. ev_tmzn과 wx_obs_tmzn은 UTC가 100%, acft_missing은 N가 99.9%, site_seeing은 N가 98.6%, certs_held는 Y가 100%, air_medical은 N가 99.1%, commercial_space_flight는 F가 100%, unmanned는 F가 99.8%이었다. 또한 해당 사고의 구분을 위한 식별자 ev_id, ntsb_no, aircraft_key, ev_type은 항공기 사고 예측과 상관이 없으므로 제외하였다. ev_city는 좌표(dec_latitude/dec_longitude)와 ev_state가 있어 대체 가능하므로 제거하고 117개의 변수가 남았다.

충돌 에너지, 비행 단계, 기상 조건, 기체 구조 및 항공기 기령과 같이 기체 손상 심각도에 직접적인 영

향을 미치는 요인들을 중심으로 분석을 수행하였으며, ev_date, ev_dow, 의학 인증등급 (med_certf), 의학 인증 유효성(med_crtf_vldty), 승무원 부상 수준(crew_inj_level), 좌석 위치(seat_occ_pic), home-built, light_condition, sky_cond, acft_category, crew_tox_perf 등 직접적인 인과 경로가 낮을 것으로 판단되는 시간 관련 변수들은 예비 SHAP 분석을 통해 실제 기여도가 0에 근접하게 나타나는 것을 확인한 후 제거하였다.

그리고 항공기 사고에서 기체 손상 심각도는 인명 피해 등급과 높은 상관 가능성이 존재한다. 그러나 이 관계는 단순한 상관을 넘어 시간적·인과적 구조상 매우 밀접하다. 일반적으로 항공사고가 발생하면 그 충격으로 기체가 손상되고 인명 피해가 발생하게 된다. 즉 인명 피해는 기체 손상의 결과이거나 최소한 동시 발생 결과인 경우가 많다. 인명 피해 등급 변수를 설명 변수로 포함할 경우, 결과 변수로부터 설명 변수를 예측하는 인과 역전(causal inversion) 문제가 발생할 수 있다. 본 연구는 기체 손상 심각도에 영향을 미치는 운항·기체·환경 요인을 규명하고, 사고 예방 관점의 사전 위험 요인 분석이 목적이므로 인명 피해 등급 변수를 분석에서 제외하였다.

분석모형의 목표 특성으로 설정한 항공기 사고 당시의 기체 손상 심각도에서 UNK(unknown)은 손상 수준이 정의되지 않은 범주로, 다범주 분류의 레이블 불확실성을 유발하므로 분석모형의 목표 특성으로 활용하기에는 부적절하여 관측치에서 제거하였다. 결측값을 제거한 후 최종 7,914건의 데이터 기준으로 damage의 빈도 분석 결과, SUBS(중대 손상)가 6,857건(86.7%)으로 가장 높은 비중을 차지하였으며, 다음으로 DEST(전손) 974건(12.3%), MINR(경미 손상) 83건(1.0%) 순으로 나타났다. 따라서 본 연구에서는 기체 손상 심각도 등급을 나타내는 damage를 예측 모형의 목표 특성으로 설정하였다. damage의 심각성 분포는 Fig. 2에 제시하였다.

모델의 예측대상은 사고 당시 기체 손상 심각도이므로 사후 결과에 해당하는 피해에 해당하는 기체 화재, 기체 폭발 변수는 제외하였다. 다음의 27개의 변수를 최종 분석 변수로 확정하였다. 기체/성능은 acft_year, afm_hrs, afm_hrs_since, cert_max_gr_wt, elt_install, far_part, fixed_retractable, ifr_equipped_cert, num_eng, total_seats의 10개의 변수를 사용하

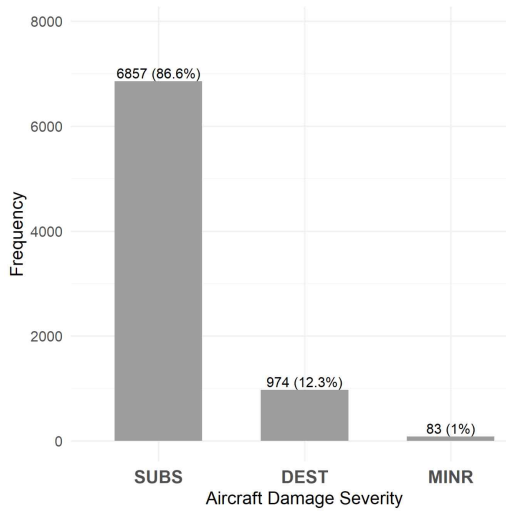


Fig. 2. Distribution of aircraft damage severity (N, %)

였다. 위치/환경은 altimeter, apt_dist, apt_elev, dec_latitude, dec_longitude, gust_ind, gust_kts, vis_sm, wind_dir_deg, wind_vel_kts, wind_cond_basic, wx_dew_pt, wx_temp의 13개의 변수를 사용하였다. 조종사 관련은 crew_age, second_pilot의 2개의 변수를, 그리고 기타 ev_year, ev_month, 의 2개 변수를 사용하였다.

3.4 머신러닝 모델 설계

기체 손상 심각도는 다범주 분류 문제이며, 클래스 간 분포가 불균형한 특성을 가진다(Farhadpour et al., 2024). 이러한 데이터 특성을 고려하여 본 연구에서는 비선형적 관계와 변수 간 상호작용을 효과적으로 학습할 수 있는 XGBoost 알고리즘을 예측 모형으로 선정하였다(Yu and Yang, 2025).

NTSB 데이터를 활용하여 결측 행과 일부 특성을 제거한 뒤, 항공기 기체 손상 심각도를 목표 특성으로 설정하고 27개의 설명 변수를 입력 특성으로 하여 항공기 사고 시 기체 손상 심각도 등급을 예측하는 머신러닝 모형을 구축하였다. 전체 7,914건의 데이터를 훈련 데이터(70%)와 테스트 데이터(30%)로 분할하고, 교차 검증은 훈련 데이터 내부에서 수행하였다(Tu et al., 2025).

모형 학습 과정에서는 교차 검증을 통해 주요 하이퍼파라미터를 탐색하였다(Yu and Yang, 2025). 트리 수는 최대 1,000개까지 학습을 시도하였으며, 조기 중

료(early stopping)를 적용하여 최적 반복 수를 자동으로 선택하였다(Yu and Yang, 2025). 트리 깊이(max_depth)는 과적합을 방지하면서도 충분한 비선형성을 확보하기 위해 4~10 범위에서 탐색하였으며, 교차검증 결과 트리 깊이는 6으로 결정하였다. 리프 노드의 최소 가중치 합(min_child_weight)은 기본값인 1로 설정하였다(Benfaress et al., 2025). 또한 학습률 η 를 변화시키며 최적의 값을 탐색하여 적용하였고, 클래스 가중치 조정 기법을 적용하여 소수 클래스의 학습 비중을 보정하였다(Farhadpour et al., 2024).

모형 성능 평가는 accuracy, precision, recall, F1-score 등의 지표를 활용하여 수행하였다. 특히 다범주 분류 문제의 특성을 고려하여 단순 정확도뿐만 아니라 소수 클래스에 대한 예측 성능을 종합적으로 평가하였다. 예측 결과의 해석을 위해 SHAP 분석을 적용하여 각 설명 변수가 항공기 손상 심각도 예측에 기여하는 정도를 정량적으로 산출하였다(Benfaress et al., 2025).

본 연구는 머신러닝 모델 선정을 위해, 트리 기반 앙상블 모델인 random forest(RF)와 gradient boosting classifier(GBC), XGBoost를 동일 조건에서 성능을 정량적으로 평가하여, 최적의 모델을 선정하였다.

IV. 분석 결과

4.1 후보 모델 성능 비교 및 최종 모형 선정

본 연구는 기체 손상 심각도를 다범주 분류 문제로 설정하고, 트리 기반 앙상블 모형인 random forest(RF), gradient boosting classifier(GBC), XGBoost의 예측 성능을 동일 조건에서 비교하였다(Rosadi et al., 2024). 모든 모형은 동일한 학습/검증 데이터 분할(70%/30%)과 동일한 27개 입력 변수를 사용하였으며, 다범주-클래스 불균형 특성을 고려하여 성능을 평가하였다.

모형 비교 결과는 Table 2에 제시된 바와 같다. Accuracy는 XGBoost(0.8959)가 가장 높았으며, precision(0.7149), recall(0.6990), F1-score(0.7009) 역시 세 모델 중 가장 우수하게 나타났다. 특히 본 연구에서 핵심 평가 지표로 설정한 재현율(recall) 측면에서 XGBoost가 RF(0.5995) 및 GBC(0.6456) 대비 뚜렷한 향상을 보였다.

이는 항공 안전과 같은 고위험 영역에서 중요한 중대 손상 사례를 놓치지 않는 탐지 능력 측면에서

Table 2. Performance comparison of XGBoost, random forest, and GBC models

Metric	RF	GBC	XGBoost
Accuracy	0.8839	0.8863	0.8959
F1	0.6885	0.7002	0.7009
Precision	0.6404	0.6765	0.7149
Recall	0.5995	0.6456	0.6990

XGBoost가 상대적으로 안정적임을 의미한다. 반면 RF와 GBC는 일부 클래스에서 보수적인 예측 경향을 보이며, 소수 클래스 탐지 성능에서 한계를 나타냈다.

결과적으로 XGBoost는 전체 성능 지표의 균형성 및 재현을 중심 평가 기준을 동시에 충족하는 최적 모델로 판단되었다. 이에 따라 이후의 학습률(η) 조정, PR/ROC 분석, 혼동행렬 및 SHAP 기반 변수 중요도 분석은 모두 XGBoost 모형을 중심으로 수행하였다.

4.2 재현을 기반 학습률(η) 선택 결과

학습률 변화에 따른 정확도(accuracy), 정밀도(precision), 재현율(recall), F1-score의 검증 성능은 Fig. 3에 제시하였다. 기체 손상 심각도 등급인 'MINR', 'SUBS', 'DEST'에 대해서는 클래스 간 표본 수의 불균형을 고려하여 정밀도, 재현율, F1-score를 가중 평균(weighted average) 방식으로 산출하였다(Farhadpour et al., 2024). 본 연구는 사고 데이터의 물리적·현실적 특성을 고려하여, 데이터 분포를 인위적으로 변

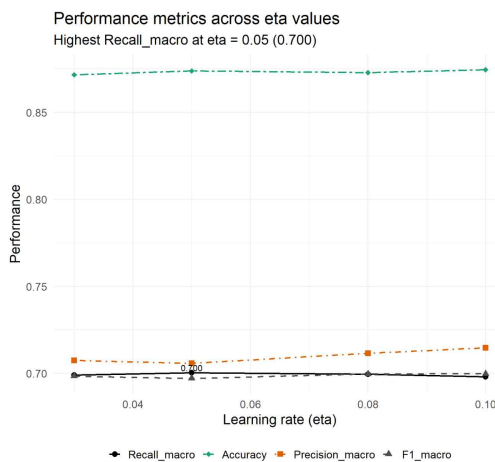


Fig. 3. XGBoost performance metrics by learning rate (accuracy, F1, precision, recall)

화시키는 증강 기법보다는 예측 안정성과 해석 가능성을 우선한 비용 민감(cost-sensitive) 접근법을 채택하였다(Tu et al., 2025). 분석 결과, 정확도는 데이터 불균형의 영향으로 학습률 변화에 따른 유의한 변동을 보이지 않았다.

오탐(false positive)은 실제로 중대한 기체 손상이 발생하지 않았음에도 이를 고위험 사례로 분류하는 경우로, 추가적인 점검이나 정비 절차가 요구되어 운영 비용이 증가할 수 있다(Kim and Choi, 2025). 그러나 이러한 비용은 예방적 조치의 하나로 사전에 관리·통제 가능한 수준의 비용에 해당한다. 반면, 미탐(false negative)은 실제로 중대한 기체 손상이 존재함에도 이를 식별하지 못하는 경우로, 정비 및 운항 의사결정의 지연을 초래하여 사고 재발 위험을 증대시킬 수 있다(Kwon and Lee, 2023). 이러한 비용 구조의 비대칭성을 고려할 때, 본 연구에서는 항공사고와 같이 위험 민감도가 높은 분야의 특성을 고려하여, 오탐 최소화보다 미탐을 줄이는 것이 더 중요하다고 판단하여 재현율(recall)을 주요 최적화 기준으로 설정하였다(Tu et al., 2025). 즉, 중대 기체 손상 사례를 놓치는 경우 정비 및 운항 의사결정이 지연되고, 사고 재발 방지를 위한 후속 조치가 충분히 이루어지지 않을 가능성이 있다. 따라서 고위험 사례를 최대한 포착하는 능력이 모델 평가의 핵심 기준이 된다.

이에 따라 학습률 η 에 따른 성능 비교에서는 재현율과 F1-score를 중심으로 지표 간 상충 관계(trade-off)를 평가하였으며, 임계값 변화에 따른 분류 성능의 안정성을 보조적으로 검증하기 위해 PR curve와 ROC curve를 함께 제시하였다(Tu et al., 2025). 그 결과, 기체 손상 심각도 발생 사례를 식별하는 재현율은 학습률 $\eta=0.05$ 에서 가장 큰 값을 보였다. 다만 본 연구에서 도출된 학습률 설정은 특정 데이터셋에서 재현율을 최대화한 경험적으로 도출된 최적값에 기반한 것으로, 다양한 항공 운영 환경에 대한 일반화를 위해서는 추가적인 외부 검증이 필요하다.

4.3 Precision-Recall 곡선 기반 성능 평가

본 연구에서 클래스별 average precision(AP)은 one-vs-rest(OVR) 전략을 기반으로 산출하였다(Chen and Lin, 2020). 각 손상 등급을 양성 클래스(positive class)로 설정하고, 나머지 등급을 모두 음성 클래스(negative class)로 통합한 이진 분류 문제로 변

환한 후, 해당 클래스에 대한 예측 확률을 사용하여 precision - recall(PR) 곡선을 계산하였다(Psaltakis et al., 2024).

PR 곡선은 분류 임계값을 0에서 1까지 연속적으로 변화시키며 산출된 재현율(recall)과 정밀도(precision)의 관계를 나타낸다(Farhadpour et al., 2024). AP는 PR 곡선 아래 면적을 적분하여 산출한 값으로 정의되며, 재현율 증가에 따른 정밀도의 변화를 종합적으로 반영한다(Psaltakis et al., 2024).

구체적으로, AP는 다음 식 (1)과 같이 정의된다.

$$AP = \sum_n (R_n - R_{n-1}) P_n \quad (1)$$

여기서 R_n 은 n 번째 임계값에서의 재현율, P_n 은 해당 지점에서의 정밀도를 의미한다. 즉, AP는 재현율 축을 기준으로 한 계단형 적분(step-wise integration)을 통해 계산되며, 이는 PR 곡선의 전체 형태를 하나의 지표로 요약한다(Davis and Goadrich, 2006).

AP는 클래스 불균형 상황에서도 안정적인 성능 비교가 가능한 지표로, 특히 고위험-희소 클래스의 판별 성능을 평가하는 데 적합하다(Gaudreault and Branco, 2024). 정확도나 ROC - AUC와 달리, AP는 양성 클래스에 대한 예측 정밀도 저하를 직접적으로 반영하므로, 항공기 기체 손상 심각도와 같이 안전상 중요한 사건 분류 문제에서 적합한 평가 지표로 평가된다(Psaltakis et al., 2024).

Fig. 4에 나타난 바와 같이, 클래스별 precision - recall(PR) 곡선의 평균 정밀도(AP)를 비교한 결과, SUBS 등급은 AP=0.959로 매우 우수한 PR 성능을 보여 매우 우수한 분류 성능을 나타냈다. 이는 다양한 분류 임계값 전반에서 정밀도와 재현율이 동시에 높은 수준으로 유지됨을 의미하며, 모델이 SUBS 등급의 손상 패턴을 안정적이고 명확하게 학습하였음을 시사한다.

MINR 등급은 AP=0.701로 중간 수준의 성능을 보였으며, 재현율이 증가함에 따라 정밀도가 점진적으로 감소하는 비교적 완만한 PR 곡선 형태를 나타냈다. 이는 경미 손상 사례에 대해 일정 수준의 판별력은 확보하였으나, 다른 손상 등급과의 경계 영역에서 분류 불확실성이 존재함을 의미한다.

반면, DEST 등급은 AP=0.580으로 가장 낮은 값을 기록하였다. DEST 등급의 PR 곡선은 재현율이 증가

함에 따라 정밀도가 급격히 저하되는 양상을 보였는데, 이는 해당 클래스가 데이터 내에서 상대적으로 희소할 뿐 아니라, SUBS 및 MINR 등급과 특성 공간에서 중첩되는 사례가 많아 분류 난이도가 가장 높은 범주임을 시사한다.

이러한 결과는 항공기 기체 손상 심각도 분류가 본질적으로 안전 중대 영역(high-risk domain)에 속하는 문제임을 보여주며, 특히 DEST 등급에서의 낮은 AP는 고위험 사고를 정확히 식별하는 데 구조적 한계와 불확실성이 정량적인 지표를 통해 확인되었다(Tu et al., 2025). 이는 향후 고위험 손상 등급에 대해 비용 민감 학습(cost-sensitive learning)이나 임계값 조정 전략의 필요성을 시사한다.

4.4 ROC 곡선을 통한 분류 성능 검증

One-vs-Rest 설정에서 산출된 ROC 곡선과 AUC (area under the curve) 값은 기체 손상 심각도 각 등급에 대한 분류기의 판별 성능을 정량적으로 평가하였다(Fig. 5). 분석 결과, 세 클래스 모두 AUC가 0.8 이상으로 나타나(DEST=0.839, SUBS=0.835, MINR=0.810), 본 모델이 무작위 분류 수준(AUC=0.5)을 충분히 상회하는 양호한 판별력을 확보하고 있음을 확인할 수 있다.

클래스별로는 DEST의 AUC가 0.839로 가장 높아, 중대 손상 여부를 다른 손상 등급과 구분하는 가장 높

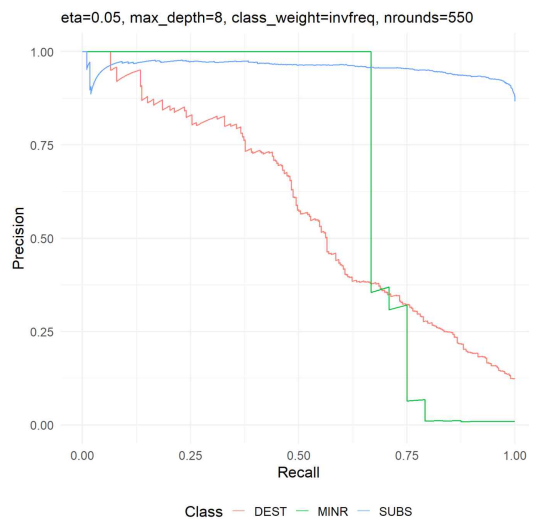


Fig. 4. PR curves and AP for aircraft damage severity classes, highlighting lower separability for the high-risk DEST class

은 분리 성능을 보였다. SUBS(중대 손상) 역시 AUC가 0.835로 DEST와 매우 근접하여, 두 클래스가 테스트 세트 전반에서 유사한 수준의 평균적 판별 성능을 보임을 시사한다.

반면 MINR(경미 손상)은 AUC가 0.810으로 상대적으로 낮았으나, 여전히 0.8을 상회하여 통계적으로 의미 있는 판별력을 유지하였다.

다만 AUC는 모든 임계값(threshold)에 걸친 평균적 분리 성능을 요약한 지표이므로, 실제 운용 관점에서 중요한 낮은 거짓양성률(FPR) 구간에서의 성능 특성을 추가로 고려할 필요가 있다. 이에 낮은 FPR 영역에서의 진양성률(TPR)을 분석한 결과, 클래스별 성능 양상은 AUC 순위와는 다른 특징을 보였다.

FPR=0.05 기준에서의 TPR은 MINR=0.750, DEST=0.497, SUBS=0.276으로 나타나, 매우 보수적인 임계값 설정 하에서도 MINR 등급이 가장 높은 민감도를 확보하고 있음을 확인할 수 있다. FPR=0.10에서도 유사한 경향이 관찰되어, MINR의 TPR은 0.750으로 유지되었으나 DEST는 0.586, SUBS는 0.445로 나타났다. 이러한 결과는 낮은 FPR 구간에서는 DEST 및 SUBS 등급에 대한 TPR이 상대적으로 감소하는 상보적 특성이 존재함을 보여준다.

한편, FPR=0.25로 허용 오차 범위를 확장한 경우, 세 클래스의 TPR은 MINR=0.792, SUBS=0.767, DEST=0.774로 나타나, 클래스 간 탐지 성능 차이가 상당 부분 축소되었음을 확인하였다. 이는 오차 허용 수준이 증가함에 따라 모델이 손상 심각도 전반에 대해 보다 균형적인 식별 능력을 보이게 됨을 시사한다.

종합하면, 본 모델은 전체 임계값 범위에 걸친 평균적 분리 성능(AUC) 측면에서는 DEST와 SUBS 등급에서 상대적으로 우수한 성능을 보이는 반면, 낮은 FPR 구간에서는 MINR 등급이 가장 높은 민감도를 확보하는 상보적인 성능 특성을 나타낸다. 이러한 결과는 손상 심각도별 분류 성능이 단일 지표(AUC)만으로는 충분히 설명되기 어렵고, 실제 운용 환경에서 요구되는 오차 허용 수준에 따라 서로 다른 강점을 가질 수 있음을 시사한다.

4.5 혼동행렬을 통한 분류 결과 분석

혼동행렬(confusion matrix)은 XGBoost 모형의 예측 결과를 실제 값과 비교하여 분류 성능을 직관적으로 파악할 수 있도록 시각화한 분석 도구이다

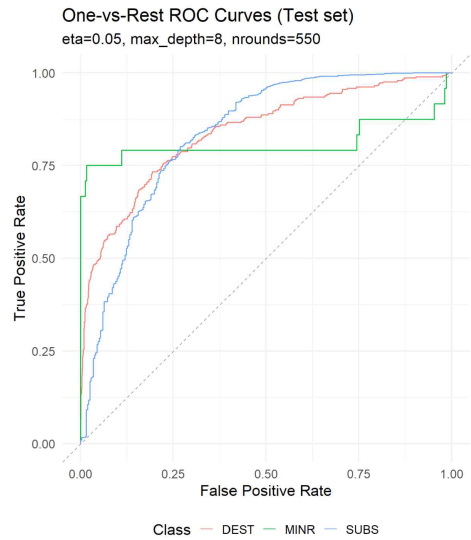


Fig. 5. Receiver operating characteristic (ROC) curves for aircraft damage severity classes (MINR, SUBS, DEST) under a one-vs-rest setting

(Powers, 2011). 혼동행렬을 히트맵(heatmap) 형태로 제시함으로써, 정확도(accuracy), 정밀도(precision), 재현율(recall), F1-score와 같은 주요 성능 지표를 종합적으로 평가할 수 있으며, 범주별로 올바르게 분류된 사례와 오분류된 사례를 구체적으로 확인할 수 있다(Powers, 2011). 이러한 특성으로 인해 혼동행렬은 머신러닝 분류 모형의 성능을 진단하는 데 효과적인 분석 기법으로 활용된다.

Fig. 6에 나타낸 혼동행렬은 각 손상 심각도 등급에 대한 모델의 실제 분류 행태를 구체적으로 보여준다. 전반적으로 모델은 SUBS(중대 손상) 등급에 대한 예측 비중이 상대적으로 높으며, 다른 두 등급(MINR, DEST)에 대해서는 보다 보수적인 예측 경향을 보였다.

먼저 SUBS 등급의 경우, 실제 SUBS 사례 2,057건 중 1,909건이 정확히 SUBS로 분류되어 약 92.8%의 높은 재현율을 나타냈다. 이는 SUBS 등급이 데이터 내에서 가장 높은 빈도를 차지하고 있으며, 모델이 해당 클래스의 전반적 패턴을 안정적으로 학습하였음을 시사한다. 다만 실제 SUBS 사례 중 일부는 DEST(143건) 또는 MINR(5건)로 오분류되어, 중간 손상 등급이 다른 손상 수준과 일정 부분 특성 공간을 공유하고 있음을 시사한다.

DEST(중대 손상) 등급의 경우, 실제 DEST 사례 292건 중 165건이 DEST로 정확히 분류되어 재현율은 56.5%로 나타났다. 반면 127건(43.5%)은 SUBS로 오

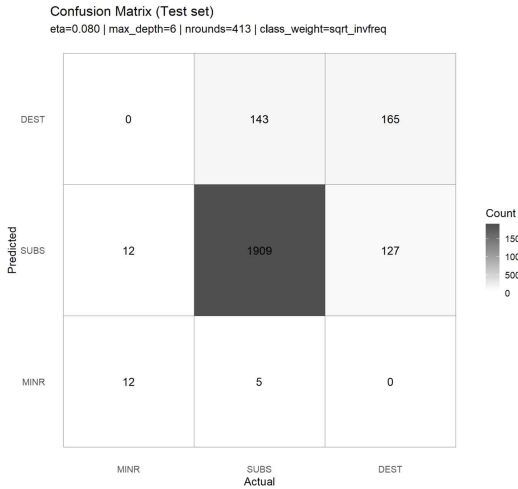


Fig. 6. Confusion matrix of the XGBoost model

분류되었다. 이는 중대 손상 사례가 데이터 상에서 상대적으로 희소하며, 중간 손상(SUBS)과 특성 분포가 증척되는 경향으로 인해 모델이 DEST에 대해 보수적인 예측을 수행하고 있음을 시사한다. 이러한 결과는 ROC 분석에서 DEST의 AUC가 상대적으로 높게 나타났음에도 불구하고, 실제 분류 임계값 하에서는 중대 손상에 대한 민감도가 제한될 수 있음을 시사한다.

MINR(경미 손상) 등급의 경우, 실제 MINR 사례 24건 중 12건이 MINR로 정확히 분류되어 재현율은 50.0%로 나타났다. 나머지 12건(50.0%)은 SUBS로 오분류되었으며, 이는 경미 손상 사례가 중간 손상과 구분 경계가 제한적임을 시사한다. 이러한 분류 패턴은 모델이 낮은 오탐을 유지하기 위해 MINR에 대해 상대적으로 보수적인 의사결정을 수행하고 있음을 반영한다.

종합하면, 본 모델은 SUBS 등급에 대해 상당히 안정적인 분류 성능을 보였으며, MINR 및 DEST 등급에 대해서는 오탐을 최소화하는 방향으로 예측이 SUBS 등급으로 상대적으로 집중되는 경향을 보였다. 이러한 결과는 앞서 제시한 ROC 분석과도 일관되게, 낮은 거짓양성을 환경에서 손상 심각도 간 분류 경계가 보수적으로 형성됨을 시사한다. 즉, 본 분류기는 손상 심각도 전반을 균등하게 구분하기보다는, 중간 손상(SUBS)을 기준 축으로 한 안정적인 분류 구조를 형성하고 있음이 확인되었다.

4.6 변수 중요도 분석

XGBoost 모형의 SHAP 중요도 분석 결과, Table 3

Table 3. SHAP feature importance ranking (top 20)

Rank	Variable	Mean (SHAP)
1	apt_dist	0.406
2	apt_elev	0.302
3	afm_hrs	0.279
4	cert_max_gr_wt	0.266
5	afm_hrs_since	0.247
6	wx_dew_pt	0.239
7	dec_latitude	0.233
8	ev_year	0.194
9	wind_vel_kts	0.181
10	wind_dir_deg	0.175
11	acft_year	0.157
12	dec_longitude	0.154
13	crew_age	0.151
14	altimeter	0.141
15	fixed_retractable.RETR	0.115
16	total_seats	0.097
17	wx_temp	0.092
18	ev_month	0.090
19	ifr_equipped_cert	0.077
20	gust_kts	0.073

과 Fig. 7에 제시된 바와 같이 항공기 기체 손상 심각도는 공항 환경(고도·거리), 기체의 구조적·물리적 특성(중량·좌석 수), 운항 및 정비 이력, 그리고 접근·착륙 단계의 기상 조건이 복합적으로 작용하는 경향을 보였다. 다만 SHAP 값은 각 변수가 모델의 예측 결과에 기여한 상대적 중요도를 의미하며, 개별 변수의 인과적 효과를 직접적으로 의미하지는 않는다.

분석 결과, 공항 및 비행 환경과 관련된 공간·지리적 변수가 가장 높은 중요도를 보였다. 특히 사고 발생 지점과 가장 가까운 공항까지의 거리(apt_dist, MeanAbsSHAP=0.406)와 공항 기준점의 해발 높이인 공항표고(apt_elev, MeanAbsSHAP=0.302)는 전체 변수 중 가장 높은 기여도를 나타내어, 사고 발생 시 비행 환경의 물리적·지형적 조건이 기체 손상 심각도를 결정하는 핵심 요인으로 작용할 가능성을 시사한다. 이는

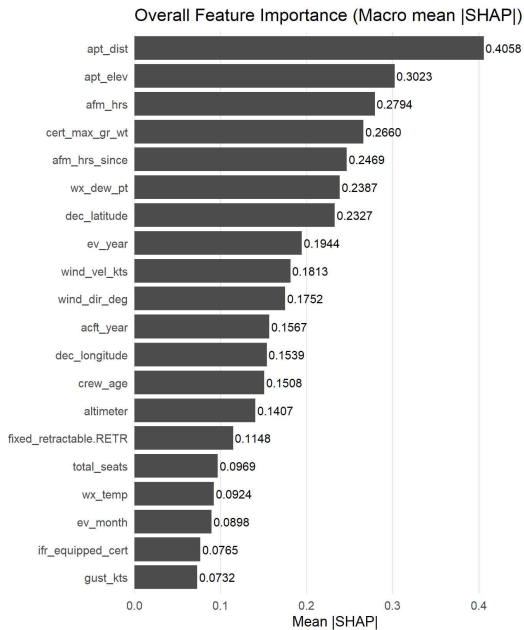


Fig. 7. Feature importance by SHAP

고지대 공항에서의 이착륙 과정이 항공기 성능 부담을 증가시키고, 사고 발생 시 손상 수준을 증가시킬 가능성을 시사하는 결과로 해석할 수 있다.

다음으로 항공기 성능 및 운항 이력과 관련된 변수들이 높은 중요도를 보였다. 총 비행시간(*afm_hrs*, 0.279), 최대 인중 중량(*cert_max_gr_wt*, 0.266), 기체 정비 이후 경과 비행시간(*afm_hrs_since*, 0.247)은 기체의 사용 강도, 구조적 피로, 운용 조건을 간접적으로 반영하는 지표로서, 사고 결과의 심각도에 유의한 영향을 미치는 것으로 나타났다. 이러한 결과는 항공기 노후도와 운항 이력이 사고의 결과 단계에 중요한 역할을 한다는 기존 연구와도 일관된다.

또한 기상 및 대기 조건 변수 역시 상위 중요 변수로 확인되었다. 이슬점(*wx_dew_pt*, 0.239), 풍속(*wind_vel_kts*, 0.181), 풍향(*wind_dir_deg*, 0.175), 기압계(*altimeter*, 0.141), 기온(*wx_temp*, 0.092), 돌풍(*gust_kts*, 0.073)은 사고 당시의 기상 불안정성과 조종 난이도를 반영하며, 기체 손상 심각도를 증폭시키는 환경적 요인으로 작용함을 시사한다. 이는 악천후 조건에서 발생한 사고가 단순 발생 여부를 넘어 결과의 중대성까지 영향을 미친다는 점을 정량적으로 확인되었다.

한편 시간적·공간적 맥락 변수인 위도(*dec_latitude*,

0.233), 경도(*dec_longitude*, 0.154), 사고 연도(*ev_year*, 0.194), 사고 월(*ev_month*, 0.090)은 사고가 발생한 지역적·시계열적 특성이 손상 심각도 분포와 연관 가능성을 시사한다. 이는 특정 지역의 지형, 기후대, 항공 인프라 특성이나 제도·기술 변화가 사고 결과에 구조적으로 반영될 가능성을 시사한다.

마지막으로 조종사와 항공기 구성 요인도 의미 있는 기여를 보였다. 조종사 나이(*crew_age*, 0.151), 항공기 제작 연도(*acft_year*, 0.157), 착륙장치 유형(*fixed_retractable.RETR*, 0.115), 좌석 수(*total_seats*, 0.097), 계기비행장비 여부(*ifr_equipped_cert*, 0.077)는 인적 요인과 기체 설계 특성이 사고 결과의 심각도를 조절하는 요인으로 작용함을 시사한다. 이는 사고 발생 이후 결과 단계에서 인적·기체적 요인이 환경 요인과 상호작용하며 복합적인 영향을 미친다는 점을 보여준다.

종합하면, 종합적인 SHAP 분석 결과는 기체 손상 심각도가 단일 요인에 의해 결정되기보다는, 공항·지형 조건, 항공기 성능 및 운항 이력, 기상 환경, 시공간적 맥락, 인적 요인이 다층적으로 결합된 결과 변수임을 보여준다. 이러한 결과는 항공사고의 결과 심각도를 설명하는 데 있어 단순 발생 원인 중심 접근을 넘어, 사고 발생 시점의 환경·운용 조건을 통합적으로 고려해야 함을 시사한다.

V. 결 론

본 연구는 항공기 사고 발생 시 기체 손상 심각도에 영향을 미치는 핵심 요인을 머신러닝 기반으로 규명하고, 이를 통해 항공 안전관리 및 관제·운항 체계 개선을 위한 실증적 시사점을 도출하는 것을 목적으로 수행하였다. 기존 항공사고 연구가 인명 피해 중심의 사고 결과 분석에 주로 초점을 맞추어 온 것과 달리, 본 연구는 사고 이후 항공 시스템 전반에 장기적 영향을 미치는 기체 손상이라는 결과 변수에 주목함으로써 항공 안전 분석의 범위를 확장하였다.

이를 위해 NTSB가 제공하는 대규모 항공사고 데이터를 활용하여 사고·기체·조종사·환경 요인을 통합한 다차원 데이터셋을 구축하였으며, 비선형적 관계와 변수 간 상호작용, 데이터 불균형 문제에 강점을 지닌 XGBoost 알고리즘을 적용하였다. 분석 과정에서는 결측치 및 불필요한 변수를 제거한 후, *damage*를 목표 특성으로 설정하고 나머지 27개 변수를 입력 특성으로

활용하였다. 분석 결과, 데이터 불균형으로 인해 전체 정확도에는 큰 변화가 없었으나, 항공기 사고에서 기체 손상 심각도를 탐지하는 재현율(Recall) 측면에서 학습률(η)의 변화가 큰 영향을 미치는 것으로 나타났다. 특히 학습률이 0.05일 때 재현율이 가장 높게 도출되어, 항공 안전관리에서는 중대 사고와 같은 소수 클래스 탐지 성능을 증시하는 유효한 접근 방식이 될 가능성을 시사하였다. 또한 SHAP 분석을 병행함으로써 모델 예측 결과의 해석 가능성을 확보하고, 각 요인이 기체 손상 심각도에 기여하는 상대적 중요도를 정량적으로 제시하였다.

분석 결과, 기체 손상 심각도는 단일 요인에 의해 결정되기보다는 공항 환경 및 공간적 요인, 기체의 구조적·물리적 특성, 운항 및 정비 이력, 그리고 접근·착륙 단계의 기상 조건이 복합적으로 작용한 결과로 나타났다. 특히 공항 고도와 공항까지의 거리와 같은 공간적 환경 요인이 가장 높은 중요도를 보였는데, 이는 기체 손상이 주로 이착륙 및 접근 단계와 같은 고위험 비행 국면에서 발생하며, 해당 국면이 공항의 물리적·환경적 특성과 밀접하게 연관되어 있음을 시사한다. 이는 기존 연구에서 상대적으로 간과되어왔던 공간적 맥락이 기체 손상 심각도 분석에서 핵심적 역할을 할 가능성을 실증적으로 보여주는 결과로 확인되었다.

아울러 기체의 최대 인증 중량, 좌석 수, 제작 연도와 같은 구조적·물리적 특성 변수와 최근 점검 이후 운항 시간, 누적 운항 시간 등 운항·정비 이력 변수 역시 기체 손상 심각도에 유의한 영향을 미치는 것으로 확인되었다. 이는 기체의 사용 강도와 노후화, 정비 주기 관리가 사고 발생 이후 손상 규모를 결정하는 중요한 요인임을 의미한다. 더 나아가 기온, 이슬점, 풍속 및 돌풍 여부, 시계비행조건(VMC)과 같은 기상 변수들은 접근 및 착륙 단계에서 조종 난이도와 기체 제어 안정성을 저하해 손상 위험 증가와 관련된 요인으로 나타났다.

본 연구의 결과는 항공 안전관리 관점을 사고 발생 여부나 인명 피해 중심의 사후적 평가에서, 기체 손상 기반의 위험 평가 및 예방 중심 관리 체계로 확장할 필요성을 시사한다. 기체 손상은 항공사의 직접적 복구 비용뿐만 아니라 운항 중단, 공항 운영 차질, 항공 교통 흐름의 불안정성으로 이어질 수 있다는 점에서, 항공 시스템 전반의 안전성과 효율성을 동시에 고려한 관리 지표로 활용될 수 있다. 특히 본 연구에서 제시한

머신러닝 기반 예측 모델은 향후 FOQA(flight operational quality assurance) 데이터나 실시간 운항·기상 정보와 결합될 경우, 사고 발생 이후의 분석 도구를 넘어 기체 손상 위험을 사전에 탐지하는 예방적 안전 관리 시스템으로 발전 가능성을 시사한다.

다만 본 연구는 NTSB 공개 데이터에 기반하여 분석을 수행하였다는 점에서, 일부 변수의 결측이나 보고 편차가 존재할 수 있으며, 사고의 세부 기계적 손상 메커니즘까지 충분히 반영하지 못하였다. 향후 연구에서는 항공기 제조사 정비 데이터, FOQA, 고해상도 기상 자료 등을 결합한 보다 정밀한 분석이 필요할 것이다. 또한 본 연구에서 활용한 다중 분류 모델을 위험 수준에 기반한 연속형 예측 문제로 확장하거나, 시간적 순서를 고려한 시계열 기반 접근을 적용하는 것도 의미 있는 연구 방향이 될 수 있다.

종합하면, 본 연구는 머신러닝 기법을 활용하여 항공기 사고 시 기체 손상 심각도의 결정 요인을 다각적으로 규명함으로써, 항공사고 분석의 이론적 지평을 확장하고 실무적 안전관리 전략 수립에 기여할 수 있는 실증적 근거를 제시하였다.

사 사

“이 논문은 2026년도 정부(산업통상자원부)의 재원으로 한국산업기술진흥원의 지원을 받아 수행된 연구임(RS-2025-02303051, 2026년 산업혁신인재성장지원사업).”

References

1. Yoon, M. G. and Son, W. M., "A study on the long-term forecasting of the international passengers based on the changes of the air travel demand structure in Korea", Journal of the Korean Society for Aviation Management, 21(4), 2023, pp.57-80.
2. Shin, J., Kim, S., Baik, H. and Lee, J., "Study on airport operation characteristics analysis and capacity enhancement strategies using the capacity coverage chart (CCO)", Journal of the Korean Society for Aeronautical Science and Flight Operations, 32(3), 2024,

- pp.35-43.
3. Kim, D. H., "A study on the accident model from the system safety perspective: focused on aircraft accident", *Journal of the Korean Society for Aviation and Aeronautics*, 28(2), 2020, pp.63-70.
 4. Yoo, J. H., Lee, J. H., and Kim, H. D., "Study on aviation safety management analysis and improvement strategies for aviation services using text mining", *J. Korean Soc. Aviat. Aeronaut.*, 32(4), 2024, pp.8-15.
 5. Gaur, D. "Human factors analysis and classification system applied to civil aircraft accidents in India", *Aviation, Space, and Environmental Medicine*, 76(5), 2005, pp.501-505.
 6. Yoon, H. S., "Machine learning based prediction model and factor appraisal of aircraft accident", *Journal of Digital Industry and Information*, 21(1), 2025, pp.15-26.
 7. Rosadi, I., Franciscus, F., and Widanto, M. H. "A machine learning-based method for predicting the classification of aircraft damage", *EAI Endorsed Transactions on Internet of Things*, 10. 2024, pp.1-7.
 8. Silagyi II, D. V., and Liu, D. "Prediction of severity of aviation landing accidents using support vector machine models", *Accident Analysis & Prevention*, 187, 2023, p.107043.
 9. Kim, J. H., Kim, T. U., and Yoo, D. H. "Analysis of the factors and patterns associated with death in aircraft accidents and incidents using data mining techniques", *Journal of Digital Convergence*, 17(9), 2019, pp.79-88.
 10. Gwon, H., Lee, J., Kim, J. G., Kim, N. S., Kim, J., and Kwon, M. S., "Temporal changes in the contributions of determinants for DO and TOC using XGBoost prediction and SHAP analysis: A case study in the saemangeum estuarine system", *Korean Soc. Mar. Environ. Saf.*, 31(6), 2025, pp.824-836.
 11. Hong, S. J., "A study on human factors and organizational culture from the perspective of aviation safety", *The Journal of Aerospace Industry*, 2002, pp.88-100.
 12. Hong, S. J., "Historical review on Human Factors of Aviation Safety", *Journal of the Korean Air Transport Association*, 1, 2005, pp.115-128.
 13. Kwon, M. and Lee, J. R., "Analysis of aviation safety management issues using text mining", *Journal of the Korean Society for Aviation and Aeronautics*, 31(4), 2023, pp.19-27.
 14. Park, T., Suh, C., and Jin, S., "A study of damage tolerance substantiation for airframe structure", *The Society for Aerospace system Engineering 2025 Spring Conference*, 2025, p.239.
 15. Bae, J. H., Lee, S. B., and Kim, C. Y., "A study on the effect of aging aircraft on the accident", *Korean Society of Aviation Operations: Proceedings of the Conference*, 2016, pp.209-212.
 16. Amaral, Y., Santos, L. F., Valério, D., Melicio, R., and Barqueira, A., "Probabilistic and statistical analysis of aviation accidents", *Journal of Physics: Conference Series*, 2526(1), 2023, p. 012107.
 17. Xu, J., Chen, L., Xing, H., and Tian, W., "Causation correlation analysis of aviation accidents: A knowledge graph-based approach", *Applied Sciences*, 14(16), 2024, p.6887.
 18. Li, W. C., Nichanian, A., Lin, J., and Braithwaite, G., "What can we learn from severity index on flight data monitoring? Analysis of safety resilience in flight operations during COVID-19 disruptions", *Ergonomics*, 67(8), 2024, pp.1164-1176.
 19. Xiong, M., Hou, Z., Wang, H., Che, C., and Luo, R., (2024). "An aviation accidents prediction method based on MTCNN and Bayesian optimization", *Knowledge and*

- Information Systems, 66(10), 2024, pp. 6079-6100.
20. Si, Q., Zhao, Y., Li, J., Wang, H., and Zhai, K., "Risk cause analysis of runway excursion accidents in the aircraft landing stage through integrated FTA-BN method", *Journal of Engineering Science & Technology Review*, 17(1), 2024, pp.178-186.
 21. Farhadpour, S., Warner, T. A., and Maxwell, A. E., "Selecting and interpreting multiclass loss and accuracy assessment metrics for classifications with class imbalance: Guidance and best practices", *Remote Sensing*, 16(3), 2024, p.533.
 22. Tu, M., Guo, J., Shi, T., Gruyer, D., Li, Y., and Luo, X., "Traffic accident prediction considering imbalanced data: An interpretable machine learning model", *International Journal of Control, Automation and Systems*, 23(10), 2025, pp.3182-3196.
 23. Benfaress, I., Bouhoute, A., and Zinedine, A., "Explainable aircraft crash severity prediction using XGBoost and SHAP", In 2025 5th International Conference on Innovative Research in Applied Science, Engineering and Technology (IRASET), 2025, pp. 1-7.
 24. Shin, Y. S. and Jo, I. J., "SIEM system performance enhancement mechanism using active model improvement feedback technology", *Journal of the Korea Contents Association*, 21(12), 2021, pp.896-905.
 25. Liu, Q., Song, P. and Li, F., "Exploring the dynamic determinants of general aviation accidents across flight phases and time: A random parameter bivariate probit approach with heterogeneity in means", *Analytic Methods in Accident Research*, 2025, p.100386.
 26. Ahmad, A.F., Sayeed, M.S., Alshammari, K., and Ahmed, I. "Impact of missing values in machine learning: A comprehensive analysis", *arXiv preprint arXiv:2410.08295*, 2024.
 27. Yu, D. and Yang J., "Heart disease risk prediction and phenotyping using XGBoost-based K-means clustering analysis", *Journal of the Korea Institute of Information, Electronics, and Communication Technology*, 18(3), 2025, pp.172-182.
 28. Kim, J. and Choi, D., "Detection of aircraft defects in walk around inspection(WAI) using YOLO-based deep learning and image processing", *Journal of the KIMST*, 28(6), 2025, pp.564-574.
 29. Chen, K. T., and Lin, C. J. "Revisiting one-versus-one and one-versus-rest: Insights into imbalanced multi-class classification", In *Proceedings of the IEEE International Conference on Data Mining (ICDM 2025)*. IEEE.
 30. Psaltakis, G., Rogdakis, K., Loizos, M., and Kymakis, E. "One-vs-one, one-vs-rest, and a novel outcome-driven one-vs-one binary classifiers enabled by optoelectronic memristors towards overcoming hardware limitations in multiclass classification", *Discover Materials*, 4(1), 2024, p.7.
 31. Paek, H., Kim, J. H., Lim, J. J., Jeon, S., and Choi, Y. J., "Quantitative safety risk assessment using aviation safety data", *Journal of the Korean Society for Aviation and Aeronautics*, 30(4), 2022, pp.145-158.
 32. Davis, J., and Goadrich, M., "The relationship between precision-recall and ROC curves", In *Proceedings of the 23rd international conference on Machine Learning*, 2006, pp.233-240.
 33. Gaudreault, J. G., and Branco, P., "Empirical analysis of performance assessment for imbalanced classification", *Machine Learning*, 113(8), 2024, pp.5533-5575.
 34. Powers, D. M. W., "Evaluation: From precision, recall and F-measure to ROC, informedness, markedness and correlation", *Journal of Machine Learning Technologies*, 2(1), 2011, pp.37-63.