

기계학습 기반 조류 충돌 항공기 손상 예측 모델 개발

박민서¹, 박준태¹, 김은수¹, 박재우², 박현우³, 조석현*

¹충남대학교, ²계명대학교, ³조선대학교, *University of California, San Diego (UCSD)

hiyoukou@gmail.com, park.juntae6942@gmail.com, eunsukim22@gmail.com, pgo8177@gmail.com, phu0985@gmail.com, *justinshcho@gmail.com

Development of a Machine Learning-based Aircraft Damage Prediction Model for Bird Strikes

Minseo Park¹, Juntae Park¹, Eunsu Kim¹, JaeWoo Park², HyeonWoo Park³, Seokheon Cho*

¹Chungnam National University, ²Keimyung University, ³Chosun University,

*University of California San Diego (UCSD)

요약

Bird strikes present a major threat to aviation safety, highlighting the need for models that can rapidly assess potential aircraft damage following a bird strike. This study proposes a machine learning-based aircraft damage prediction model using a comprehensive dataset from the U.S. Federal Aviation Administration (FAA) containing detailed records of bird strike incidents. Given the pronounced class imbalance in the dataset, the G-Mean metric is adopted for performance evaluation, and a Random Forest algorithm is employed to construct the prediction model. To enhance predictive performance, the SMOTENC technique is applied to mitigate data imbalance, and a new feature representing weighted damage occurrence is introduced. The proposed model achieves performance improvement, supporting rapid return-to-service decisions for low-risk aircraft and enabling targeted in-depth inspections for high-risk cases, thereby enhancing operational safety as well as reducing maintenance costs.

I. 서 론

조류 충돌 (bird strike)은 항공기 운항 중 조류와 충돌하는 것으로 항공기의 엔진 결함 또는 기체 구조 손상으로 이어질 수 있다 [1]. 조류 충돌은 사고 발생 자체보다 사고 직후의 항공기의 점검 및 정비가 중요하므로 항공기 손상 발생 여부를 정확하고 신속하게 예측할 수 있는 모델이 요구된다 [2]. 하지만, 기존 연구는 비용 추정에 집중한 경우가 많기 때문에 운항 안전 관점에서 물리적 손상 자체를 직접 예측하는 접근은 상대적으로 제한적이라는 한계점이 있다 [3]. 이에 본 연구는 실제 조류 충돌 관련 정보들을 이용하여 Random Forest 알고리즘 기반 항공기 손상 여부 예측 모델을 개발하고자 한다.

II. 데이터세트 구성 및 데이터 전처리

본 연구는 미국 연방항공청 (FAA)이 제공하는 플로리다주에서 발생한 조류 충돌 데이터를 사용하였다 [4]. 원본 데이터는 101개 독립 변수들을 포함하지만 연구 목적과의 관련성이 낮은 변수, 결측이 과도한 변수 그리고 수리 비용과 같이 조류 충돌 손상 정도를 알 수 있는 변수들을 제외하여 최종 16개의 독립 변수들을 선정하였다. 종속 변수는 항공기 손상 유무이며, 0과 1의 값은 각각 무손상과 손상을 의미한다.

그림 1은 종속 변수의 클래스별 분포를 나타낸다. 조류가 충돌하였을지라도 실제로 항공기에 손상이 발생하는 건수는 약 743건으로 전체 데이터의 9.8%를 차지하여 극심한 데이터 불균형을 보여주고 있다.

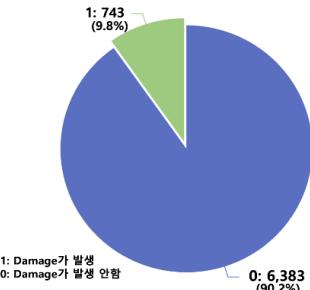


그림 1. 미국 연방항공청 조류 충돌 항공기 데이터세트의 클래스별 분포

또한, 비행 고도에 대한 결측값은 비행 단계별 중위값으로 대체하였고, 강수 상태 결측치는 '맑음'으로 처리하였다. 이외 결측값을 포함한 데이터 샘플들은 제거하여 최종적으로 6,604 개 샘플을 확보하였다.

조류 충돌에 항공기 부위 (N)를 총 17 개 영역으로 구분하고 있다. STR_{ji} 는 j 번째 항공기의 i 번째 부위 손상 여부 ($STRike$)를 의미한다. 항공기 손상 예측 모델의 성능을 향상시키기 위해서 가중치 손상 여부 (weighted strike: WSTR)이라는 새로운 독립 변수를 정의한다. 항공기 j 번째에 대한 가중치 손상 여부는 다음과 같다.

$$WSTR_j = \sum_{i=1}^N w_i \cdot STR_{ji} \quad (1)$$

여기에서, w_i 는 i 번째 부위의 상대 손상 정도를 반영하는 가중치이다. 이러한 w_i 는 1990년부터 2023년까지 미국 연방항공청에서 제공하는 조류 충돌 데이터를 이용하여 다음과 같이 도출하였다.

$$w_i = \frac{p_i}{\sum_{i=1}^N p_i} \quad (2)$$

여기에서, p_i 는 조류가 i 번째 부위에 충돌하였을 때 실제로 손상이 일어난 확률을 의미한다.

III. 조류 충돌 항공기 손상예측 모델 분석

3.1 항공기 손상 예측 모델 학습 및 성능 평가 지표

본 연구는 Random Forest (RF) 알고리즘을 기반으로 항공기 손상 예측 모델을 구축하였다. 6,604 개의 전체 데이터 샘플들을 학습 데이터세트와 테스트 데이터세트를 8:2 비율로 분할하였다.

또한, 데이터 불균형 완화를 위해 오버샘플링 기법의 하나인 SMOTENC (Synthetic Minority Over-sampling Technique for Nominal and Continuous features)를 학습 데이터에만 적용하였다. 고려하는 데이터세트에 포함된 독립 변수들이 대부분 범주형 변수이기 때문에 SMOTENC를 사용하였다.

극심한 데이터 불균형으로 인해 모델 성능 평가 지표로서 민감도 (sensitivity)와 특이도 (specificity)를 같이 반영하는 G-Mean을 사용하였고 다음처럼 정의한다.

$$G - Mean = \sqrt{Sensitivity \cdot Specificity} \quad (3)$$

3.2 항공기 손상 예측 모델 결과

그림 2는 SMOTENC를 사용 여부에 따른 RF 알고리즘 기반 항공기 손상 예측 모델의 성능 결과를 보여준다. SMOTENC를 학습 데이터에만 적용하여 소수 클래스 데이터인 손상된 항공기 정보를 더욱 학습함으로써 G-Mean이 0.5471에서 0.7891로 상승함을 확인할 수 있다. 또한, 정확도 (accuracy)는 미약하게 감소했지만 실제로 손상된 항공기를 정확하게 예측함을 보여주는 지표인 재현율 (Recall)이 0.37 정도로 많이 상승하였다.

IV. 결론

본 연구에서는 미국 연방항공청에서 제공하는 플로리다주에서 발생한 항공기 조류 충돌 관련 데이터를 사용하여 항공기 손상 여부를 예측하는 모델을 개발하였다.

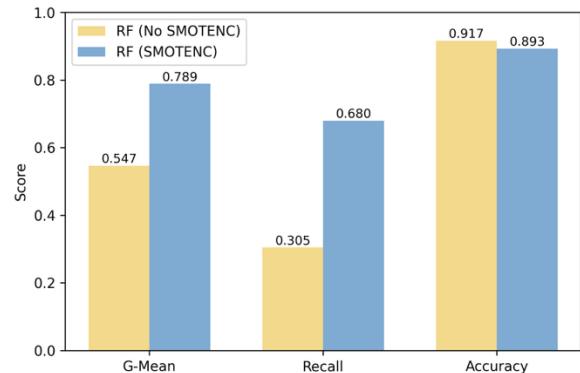


그림 2. 조류 충돌 항공기 손상 예측 모델 성능

고려하는 데이터세트는 극심한 불균형 데이터이기 때문에 범주형 변수들도 오버샘플링할 수 있는 SMOTENC 기법을 학습 데이터세트에 적용하였다. 또한, 항공기 손상 여부를 분류하기 위해 Random Forest (RF) 알고리즘을 사용하였고 모델 성능 평가 지표로 G-Mean을 채택하였다. 마지막으로 예측 모델 성능을 향상시키기 위해서 과거 데이터를 고려한 가중치 손상 여부 (WSTR)라는 새로운 변수를 생성하여 주요 독립 변수로 사용하였다. 이로써 SMOTENC 사용 시 RF 기반 항공기 손상 예측 모델의 G-Mean과 재현율 (Recall)이 각각 약 0.789와 0.680으로 향상된 결과를 보여주었다. 향후 연구에서는 조류 충돌 지역을 플로리다주가 아닌 미국 전역으로 확장하여 보다 일반적인 항공기 손상 예측 모델을 개발하고자 한다.

ACKNOWLEDGMENT

This work was supported by the Software-Oriented University Support Program through the Ministry of Science and ICT (MSIT) and the Institute of Information & Communications Technology Planning & Evaluation (IITP) in 2025 (2021-0-01435) & (2024-0-00062); and by the Specialized University of Information Security through the Korea Internet & Security Agency (KISA) funded by the Korean government (MSIT) in 2025. Following are results of a study on the “Leaders in Industry-university Cooperation 3.0” Project, supported by the Ministry of Education and National Research Foundation of Korea.

참 고 문 헌

- [1] M. Bae, J. Yoon, and I. Ahn, “Study on Bird Strike Prevention by Using Cries of Falcon in The Birds of Prey,” Korean Society of Computer Information Conference, Jan. 2016.
- [2] R. A. Dolbeer, M. J. Begier, P. R. Miller, J. R. Weller, and A. L. Anderson, “Wildlife Strikes to Civil Aircraft in The United States, 1990–2022”, Washington, DC, USA: U.S. Dept. of Transportation, Federal Aviation Administration, 2023.
- [3] S. Misra, I. Toppo, and F. A. C. Mendonca, “Assessment of Aircraft Damage Due to Bird Strikes: A Machine Learning Approach,” International Journal of Sustainable Aviation, vol. 8, no. 1, pp. 1–17, 2022.
- [4] U.S. Department of Transportation, Federal Aviation Administration, “Wildlife Strikes to Civil Aircraft in The United States, 1990–2023”, Serial Report No. 30, 2024.