Check for updates

Journal of the Korean Institute of Industrial Engineers
Vol. 52, No. 5, pp. 470-480, October 2026.
ISSN 1225-0988 | EISSN 2234-6457
https://doi.org/10.7232/JKIIE.2026.52.5.470
© 2026 KIIE

함정 추진체계 상태기반정비를 위한 가스터빈 열화 예측: 외삽 일반화 평가와 비용기반 의사결정 관점

손 창 호†

육군3사관학교 인공지능・시스템과학과

Gas Turbine Decay Prediction for Condition-Based Maintenance of Naval Propulsion Systems: An Extrapolative Generalization Assessment and a Cost-based Decision Perspective

Chang-Ho Son

Department of Artificial Intelligence and Systems Science, Korea Army Academy at Yeongcheon

As naval maintenance shifts to condition-based maintenance (CBM), estimating gas-turbine degradation from sensor data is essential. Using the public CODLAG frigate dataset, this study moves beyond model comparison and contributes along three axes. First, because the data is a deterministic full-factorial simulator grid, conventional random cross-validation overstates accuracy: six nonlinear models score $R^2 \geq 0.97$, yet all models degrade sharply under extrapolation to unseen operating points or decay levels; repeated runs over five seeds show a multilayer perceptron (MLP) robustly retains a positive R^2 only under decay-level extrapolation (0.94 ± 0.03 for the compressor decay coefficient), while under operating-point extrapolation even its R^2 is unstable across seeds (mean -0.08 ± 0.51)—and rankings reshuffle. Second, we link prediction to decisions through cost-based model selection, sensor-portfolio optimization, conformal prediction intervals, and inspection-policy optimization; a policy designed under optimistic (interpolation) error costs about 46.6% more under realistic error. Third, beyond Gaussian noise we assess structural noise, showing MLP is fragile to outliers and missingness while random forests are robust. Significance is confirmed by Friedman and Nemenyi tests. The study recalibrates achievable CBM performance and offers decision-oriented guidance.

Keywords: Condition-Based Maintenance, Naval Propulsion System, Extrapolative Generalization, Cost-based Decision, Conformal Prediction

1. 서 론

무기체계의 복잡화, 고가화에 따라 정비 비용은 무기체계 총 수명주기비용의 상당 부분을 차지하며, 효율적 정비정책의 수립은 군 전투준비태세와 직결된다. 전통적 정비 방식인 사후정비(corrective maintenance)는 고장 발생 후 조치로 인한 가동률 저하를, 계획정비(preventive maintenance)는 잔존 수명이 남은 부품의 조기 교체로 인한 과잉정비를 초래한다. 이러한 한계를 극복하기 위해 미 국방부는 장비 상태정보의 실시간 수집, 분석에 기반하여 필요한 시점에 필요한 정비만을 수행하는 상태기반정비(Condition-Based Maintenance, CBM) 및 이를 신뢰성 분석과 결합한 CBM+를 정책적으로 추진해 왔으며, 우리 군도 국방혁신과 연계하여 한국형 CBM+ 도입을 진행하고 있다(Jeong, 2021; Son and Kim, 2021).

이 성과는 충성대연구소와 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(NRF-2022R1F1A1062959와 RS-2026-25480141).
† 연락저자 : 손창호 교수, 38900 경북 영천시 고경면 호국로 495 육군3사관학교 인공지능・시스템과학과, Tel : 054-330-4821, E-mail : c13981@kaay.ac.kr
2026년 7월 3일 접수; 2026년 7월 17일, 2026년 7월 20일 수정본 접수; 2026년 7월 20일 게재 확정.

함정 추진체계는 CBM 적용 효과가 큰 대표적 체계이다. 함정용 가스터빈은 해염 입자 흡입에 따른 압축기 오염(fouling)과 고온부 열화 등으로 성능이 점진적으로 저하되며, 정확한 진단을 위해서는 분해점검이 필요하나 이는 장기간의 전력 공백과 높은 비용을 수반한다. 따라서 운용 중 상시 수집되는 센서 데이터로부터 압축기, 터빈의 열화 수준을 추정할 수 있다면, 분해점검 주기를 상태 기반으로 최적화하고 정비소요를 사전에 예측하는 것이 가능해진다.

데이터 기반 열화 예측에 관한 선행연구는 적지 않으나, 대다수가 공개 벤치마크 데이터에 다양한 회귀 알고리즘을 적용하여 정확도를 비교하는 데 머물러 왔다. 이러한 접근은 두 가지 근본적 한계를 갖는다. 첫째, 평가 설계의 문제이다. 본 연구가 사용하는 CODLAG 데이터를 포함한 다수의 시뮬레이터 기반 데이터는 운용변수와 열화변수를 완전요인(full-factorial) 격자로 조합하여 결정론적으로 생성된다. 이러한 데이터에 무작위 교차검증을 적용하면, 임의의 시험 표본 근방에 거의 동일한 운용조건, 인접 열화수준의 학습 표본이 존재하므로 모형은 일반화가 아니라 조밀한 격자 위의 내삽(interpolation)을 수행하게 된다. 그 결과 보고되는 높은 결정계수는 실제 운용 일반화 성능이 아니라 실험설계의 산물일 수 있다. 둘째, 의사결정 연계의 부재이다. 정비 현장에서 의미 있는 것은 RMSE 자체가 아니라, 예측으로부터 도출되는 점검 여부 결정의 비용, 위험, 어떤 센서를 어느 예산으로 확보할지의 선택, 그리고 예측의 불확실성을 반영한 의사결정이다. 예측 정확도 비교만으로는 이러한 산업공학적 의사결정 문제에 답할 수 없다.

본 연구는 이 두 한계를 정면으로 다룬다. 연구질문은 본 연구의 세 축에 대응하는 세 가지 핵심 질문으로 요약된다. (RQ1=Q1, 평가 타당성) 관례적 무작위 교차검증은 시뮬레이터 격자 데이터에서 열화 예측 성능을 어느 정도 과대평가하며, 미관측 운용점, 열화수준에 대한 외삽 성능과 모형 우열은 어떻게 달라지는가? (RQ2, 예측 - 의사결정 연계) 예측을 정비비용, 센서 구성, 예측 불확실성, 검사・정비 정책과 같은 실제 의사결정에 연결하면 어떤 함의가 도출되는가? 세부적으로 (Q2) RMSE 최적 모형과 기대 정비비용 최적 모형은 일치하는가? (Q3) 정확도 - 센서수 사이의 파레토 최적 센서 부분집합은 무엇인가? (Q4) 등각예측 구간은 목표 신뢰수준을 만족하는가? (Q6) 예측 불확실성을 반영한 검사, 정비 정책 최적화에서, 평가 타당성은 의사결정 비용에 어떤 영향을 주는가? (RQ3=Q5, 실환경 이전성) 가우시안 백색잡음을 넘어선 구조적 잡음(드리프트, 결측, 이상치)에 대해 모형은 얼마나 강건한가? 각 질문에 대응하는 분석 방법과 핵심 결론은 4장 서두에 요약한다.

본 연구의 기여는 다음과 같다. 첫째(방법론적 독창성), 완전요인 시뮬레이터 데이터에서 무작위 교차검증이 야기하는 내삽 과대평가를 정량화하고, 운용점 외삽(leave-one-operating-condition-out)과 열화수준 외삽이라는 현실적 일반화 프로토콜을 제안하여, 동일 데이터를 사용한 기존 벤치마크 연구(Coraddu *et al.*, 2016; Cipollini *et al.*, 2018a)가 다루지 않은 평가 타당성 문제를 제기한다. 둘째(의사결정 연계), 예측을 비용기반 모형선택, 센서 포트폴리오 최적화, 불확실성 기반 점검 규칙, 나아가 장기 비용률을 최소화하는 상태기반 검사, 정비 정책 최적화문제로 구성하여, 예측 모형을 정비 의사결정 시스템의 구성요소로 통합한다. 셋째(실환경 이전성), 구조적 잡음에 대한 모형별 강건성의 상보적 구조를 규명함으로써 시뮬레이터 결과의 실선 적용 가능성을 정량적으로 진단한다. 넷째(재현성), 모든 분석을 공개 데이터와 공개 절차(스크립트 일체)로 수행하여 후속 연구의 검증 가능한 기준을 제공한다.

본 논문의 구성은 다음과 같다. 제2장에서 관련 연구를 검토하고, 제3장에서 데이터와 연구 방법을 기술한다. 제4장에서 실험 결과를 제시, 분석하며, 제5장에서 결론과 향후 연구 방향을 논의한다.

2. 관련 연구

2.1 함정 추진체계 열화 예측 연구

함정 가스터빈 추진체계의 데이터 기반 상태진단 연구는 Coraddu *et al.*(2016)이 실측 검증된 수치 시뮬레이터로 CODLAG 데이터를 생성, 공개하면서 본격화되었다. 이들은 커널 기반 기법으로 열화계수 예측 가능성을 보였고, Cipollini *et al.*(2018a)은 신경망・가우시안 프로세스・앙상블 등을 비교해 비선형 모형의 우수성을, Cipollini *et al.*(2018b)은 최소 피드백만으로 열화를 추정하는 준지도 접근을 제안하였다. 이후 유전 프로그래밍 기반 기호회귀(Andelic *et al.*, 2020) 등 다양한 기법이 이 데이터셋에 적용되어 왔으나, 공통적으로 무작위 분할 평가를 채택하였을 뿐 완전요인 격자 데이터의 내삽적 성격과 외삽 일반화 성능의 괴리는 검토되지 않았다. 본 연구는 이 지점을 출발점으로 삼는다.

2.2 평가 타당성: 데이터 누수와 외삽 일반화

기계학습 평가에서 학습, 시험 분할이 데이터의 종속구조를 무시할 때 성능이 낙관적으로 편향된다는 문제는 여러 응용분야에서 보고되어 왔으며, 시계열, 공간, 그룹 구조 데이터에서는 무작위 분할 대신 블록(group) 기반 분할이 권고된다 (Roberts *et al.*, 2017; Kapoor and Narayanan, 2023). 완전요인 시뮬레이터 데이터는 설계상 운용조건의 모든 조합이 조밀하게 채워져 있어, 무작위 분할 시 시험점이 사실상 학습점들의 볼록조합 내부에 위치한다. 이때 측정되는 것은 운용 일반화가 아니라 격자 내삽 능력이다. 따라서 미관측 운용점, 미관측 열화수준에 대한 그룹기반, 외삽기반 분할이 실제 배치 성능의 보수적 추정치를 제공한다. 본 연구는 이러한 평가 타당성 논의를 함정 CBM 열화 예측에 처음으로 적용한다. 이러한 내삽 - 외삽 격차는 CODLAG 데이터셋에 국한되지 않고, 결정론적 시뮬레이터 기반 격자 데이터 전반에 내재한 구조적 위험

이다(Kapoor and Narayanan, 2023). 다만 그 크기는 데이터셋마다 다를 수 있어(3.1절), 본 연구의 배율(최대 24배)은 CODLAG 데이터의 사례값이다.

2.3 예측 – 의사결정 연계와 불확실성 정량화

예지정비(predictive maintenance) 문헌에서는 예측 모형의 가치가 궁극적으로 정비 의사결정의 비용·위험 절감으로 환산되어야 한다는 관점이 강조되어 왔고(Jardine *et al.*, 2006; Lei *et al.*, 2018; Achouch *et al.*, 2022), 해운·함정 분야에서도 AI 기반 예지정비가 주목받고 있다(Simion *et al.*, 2024). 체계적 문헌고찰(Carvalho *et al.*, 2019) 역시 예측 정확도를 넘어선 의사결정·비용 관점 평가의 미흡함을 지적하며, 비용민감(cost-sensitive) 평가, 센서 예산 제약 하의 변수선택, 불확실성 정량화가 중요하게 다루어진다. 특히 등각예측(conformal prediction; Vovk *et al.*, 2005; Shafer and Vovk, 2008; Angelopoulos and Bates, 2023)은 분포가정 없이 목표 커버리지를 보장하는 예측구간을 제공해 안전이 중요한 의사결정에 적합하며, 잔존수명 추정에도 적용된 바 있다(Javanmardi and Hüllermeier, 2023). 국내 함정 CBM 연구에서 이러한 의사결정 연계·불확실성 정량화를 통합한 사례는 찾아보기 어렵다.

2.4 국내 국방 CBM 연구

국내에서는 CBM의 정책적 도입 방안에 관한 연구가 주를 이루어 왔다. Jeong(2021)은 국내외 CBM 적용사례를 조사해 제도·교육·기술 발전방향을 제시하였고, Son and Kim (2021)은 상태기반정비 요구도의 국방규격 반영 방안을, Lee *et al.*(2024)은 획득단계별 한국형 CBM+ 적용 업무절차를 제시하였다. 함정 분야에서는 Oh(2016)가 한국형 CBM 체계 개발을 다루었고, 한국 해군은 광개토-III급 이후 건조 함정에 통합상태평가체계(ICAS)를 적용해 왔다(Jeong, 2021). 그러나 공개 데이터로 함정 추진체계 열화 예측 모형을 정량적으로 비교하고 평가 타당성·의사결정 연계까지 통합적으로 다룬 국내 연구는 전무하며, 본 연구는 이 간극을 메우고자 한다.

3. 데이터 및 연구 방법

3.1 데이터

본 연구는 UCI Machine Learning Repository에 공개된 'Condition Based Maintenance of Naval Propulsion Plants' 데이터셋(CC BY 4.0)을 사용하였다(UCI Machine Learning Repository, 2014). 이 데이터는 실측 자료로 검증된 CODLAG 호위함 추진체계 수치 시뮬레이터로부터 생성된 것으로, 가스터빈, 감속기어, 가변피치 프로펠러 등으로 구성된 추진체계의 정상상태 운용을 모사한다. 데이터는 추진레버 위치 9수준, 압축기 열화계수 k_{Mc} 51수준(0.95~1.00), 터빈 열화계수 k_{Mt} 26수준(0.975~1.00)의 완전요인 설계로 구성되어 총 11,934개(9×51×26) 관측치를 포함한다. 열화계수 1.0은 신품 상태를, 값이 작을수록 진행된 열화를 의미한다.

각 관측치는 18개 변수로 구성되며, 이 중 16개가 센서 측정변수, 2개가 예측 대상인 열화계수이다. 탐색적 분석 결과 압축기 입구 온도(T1)와 입구 압력(P1)은 전 관측치에서 상수였고, 좌현, 우현 프로펠러 토크(Ts, Tp)는 완전히 동일하였다. 이에 T1, P1, Tp를 제거하고 <Table 1>의 13개 변수를 입력으로 사용하였다.

Table 1. Input variables

Variable	Description	Unit
lp	Lever position	-
v	Ship speed	knots
GTT	Gas turbine shaft torque	kN·m
GTn	Gas turbine rate of revolutions	rpm
GGn	Gas generator rate of revolutions	rpm
Ts	Propeller torque	kN
T48	HP turbine exit temperature	°C
T2	Compressor outlet air temperature	°C
P48	HP turbine exit pressure	bar
P2	Compressor outlet air pressure	bar
Pexh	Exhaust gas pressure	bar
TIC	Turbine injection control	%
mf	Fuel flow	kg/s

주목할 점은 개별 센서 변수와 열화계수 간의 주변 상관관계가 매우 약하다는 것이다. 압축기 열화계수(k_{Mc})와의 피어슨 상관계수는 절대값 기준 최대 0.047(T2), 터빈 열화계수(k_{Mt})와는 최대 0.038(T48)에 불과하였다. 이는 센서 측정값의 분산 대부분이 열화가 아닌 운용조건(레버 위치, 함속)에 의해 설명되기 때문으로, 열화 신호를 추출하기 위해서는 운용조건을 조건부로 고려하는 비선형 모형이 필요함을 시사한다. 동시에 이 사실은 평가 프로토콜의 중요성을 부각한다. 운용조건이 동일한 학습 표본이 시험 표본 근방에 존재할 때(무작위 분할)와, 시험 운용조건이 학습에서 완전히 배제될 때(외삽 분할)의 성능은 본질적으로 다를 수밖에 없다.

이 약한 상관관계가 외삽에서 큰 성능 붕괴로 이어지는 원인을 확인하기 위해, 핵심 센서(압축기: T2, P48, P2; 터빈: P2, GTT, GGn)를 운용조건(lp)과 열화계수의 주효과 및 교호작용으로 분산분해하는 이원분산분석(Type II ANOVA)을 수행하였다. 그 결과 lp 주효과가 센서 분산의 99.5~99.9%를 설명한 반면 열화계수 주효과는 0.01~0.22%에 불과했고, lp×열화계수 교호작용은 통계적으로 유의하나(n=11,934, $p<0.001$) 설명분산은 0.03% 이하로 무시할 수준이었다. 즉 외삽 붕괴는 비선형 교호작용이 아니라 열화 신호의 극단적으로 작은 분산 비중

(약 1/500~1/10,000)에서 비롯되며, 무작위 분할에서 존재하던 근접 학습표본의 내삽 단서가 운용점 외삽(P2)에서는 사라지기 때문이다.

3.2 평가 프로토콜: 내삽과 외삽

본 연구는 세 가지 평가 프로토콜을 정의하고 그 결과를 직접 대비한다. (P1) 무작위 5-겹 교차검증(내삽적 평가): 전체 관측치를 무작위로 5개 겹으로 분할하며, 선행연구의 관례적 평가이자 과대평가의 상한으로 활용한다. (P2) 운용점 외삽 − Leave-One-Lever-Out(LOLO): 추진레버 위치(lp) 9수준을 그룹으로 하여 8개 운용점으로 학습하고 남은 1개 운용점을 시험한다(9-겹 그룹교차검증). 학습 시 관측되지 않은 운용조건에서의 일반화를 평가한다. (P3) 열화수준 외삽: 각 열화계수의 상위 60% 구간(경미한 열화)으로 학습하고 하위 40% 구간(심한 열화)을 시험한다. 운용 초기의 경미한 열화 데이터만으로 미래의 심한 열화를 예측해야 하는 실제 CBM 시나리오를 직접 반영하며, 가장 어려운 일반화 과제이다. 세 프로토콜 모두 동일 난수시드로 분할을 고정하여 모형 간 공정 비교를 보장하였다.

3.3 비교 모형

선형 모형(OLS, 릿지회귀), 사례기반 모형(kNN), 트리 모형(CART), 배깅 앙상블(랜덤포레스트; Breiman, 2001), 부스팅 앙상블(XGBoost; Chen and Guestrin, 2016, LightGBM; Ke *et al.*, 2017), 신경망 모형(MLP)의 8개를 비교하였다. OLS, 릿지, kNN, CART, 랜덤포레스트, MLP는 scikit-learn(Pedregosa *et al.*, 2011)으로, XGBoost와 LightGBM은 각각 xgboost, lightgbm 라이브러리로 구현하였다. 압축기, 터빈 열화계수 각각에 대해 별도의 단일출력 회귀모형을 구축하였다. 선형, 사례기반, 신경망 모형에는 z-점수 표준화를 적용하였고, 트리 계열은 원척도를 사용하였다. 릿지 벌점계수 1.0, kNN 이웃 수 5, CART 최대깊이 16, 잎 최소표본 3, 랜덤포레스트 트리 30개, XGBoost, LightGBM 트리 600개, 학습률 0.05, 부표본 0.9, MLP 은닉층 (64, 32), ReLU, Adam(학습률 0.001; Kingma and Ba, 2015), 미니배치 256, 최대 400 에포크, 조기종료(40 에포크)로 설정하였다. 이 값들은 저자의 예비실험에 기반한 고정값으로(전수 탐색 아님), 세 프로토콜 간 공정 비교를 위해 의도적으로 고정하였다. RandomizedSearchCV 점검 결과 개선폭은 3.5~25.5%로 내삽 - 외삽 격차(최대 24배)에 비해 미미해 정성적 결론에 영향이 없었다. MLP는 목표변수도 표준화 후 역변환하였다. 결과의 확률적 변동성을 확인하기 위해, P1(무작위 5-겹 교차검증)은 서로 다른 5개의 분할 시드(42, 1, 2, 3, 4)로, P2(운용점 외삽)와 P3(열화수준 외삽)은 분할이 그룹・분위수 기준으로 결정론적이므로 모형 초기화 시드를 각각 5회 달리하여 반복하고, <Table 2>・<Table 3>에는 5회 반복의 평균과 표준편차를 함께 보고한다.

3.4 의사결정 연계 방법

(D1) 비용기반 모형선택: 예측 열화계수 $\hat{k}$가 임계 d^*(하위 25% 분위수) 미만이면 분해점검 지시로 하는 이진의사결정 문제로 정의한다. 미탐(n_{FN}, 고장위험 C_{FN})과 오탐(N_{FP}, 불필요점검 C_{FP})으로 오분류를 구분하고, 기대비용 $C = C_{FN} n_{FN} + C_{FP} n_{FP}$를 비율 $C_{FN} : C_{FP} \in \{1, 5, 10, 20\}$에서 산출한다. (D2) 센서 포트폴리오 최적화: 탐욕적 전진선택으로 센서를 1개씩 추가하며 3-겹 교차검증 RMSE를 측정하여 정확도 - 센서수 파레토 프론티어를 산출한다. (D3) 등각예측 기반 불확실성: 분할 등각예측으로 보정셋 잔차의 $(1-\alpha)$ 분위수 q로 예측구간 $[\hat{k}-q, \hat{k}+q]$를 구성한다($\alpha = 0.1$). 경험적 커버리지와 평균 구간폭을 보고한다. (D4) 상태기반 검사, 정비 정책 최적화: 장기 평균 비용률을 최소화하는 최적화 문제를 구성한다. 열화계수 k는 신품(1.0)에서 운용시간에 따라 단조 감소하여 고장임계 $k_f = 0.95$에 이르며, 엔진별 열화율은 신품 → 고장 도달시간 $\sim U(40, 80)$ 개월로 이질적이라 가정한다. 결정변수는 검사주기 τ(개월)와 정비임계 d^*이다. 검사 시 추정오차 $\varepsilon \sim N(0, \sigma)$의 σ를 본 연구가 측정한 예측오차(내삽 σ_{opt} vs 외삽 σ_{real})로 설정하여 예측 평가와 의사결정을 연결한다. 비용은 검사 $c_{sp} = 1$, 오버홀 $c_{oh} = 10$, 고장 $c_{fail} = 100$으로 두고 장기 비용률을 몬테카를로로 추정하여 (τ, d^*) 격자에서 최적 정책을 탐색한다. 본 연구가 사용한 비용 파라미터(c_{sp}, c_{oh}, c_{fail})와 열화수명분포는 "검사비용 ≪ 오버홀비용 ≪ 고장비용"이라는 CBM 문헌에서 통용되는 정성적 비용 순서(Jardine *et al.*, 2006)를 반영한 예시적 상대값으로, 특정 함정급의 실측 정비비용 장부로 보정된 값이 아니다. 절대 비용률 수치는 실제 정비예산의 예측치가 아니라 상대적 비교를 위한 프레임워크로 해석되어야 하며, 실제 적용 시에는 해당 함정급의 실측 검사・오버홀・고장 비용과 열화이력 데이터로 재보정이 필요하다.

3.5 강건성 및 통계적 유의성

실운용 센서는 (i) 가우시안 백색잡음뿐 아니라 (ii) 드리프트(센서별 일정 바이어스), (iii) 결측(무작위 결측 후 열평균 대치), (iv) 이상치(대형 스파이크) 등의 구조적 교란을 겪는다.

각 교란을 표준편차의 1, 2, 5%로 주입하여 무잡음 학습 모형의 시험 성능 저하를 측정한다. 변수중요도는 순열 중요도(MLP 기준)와 SHAP(XGBoost 기준; Lundberg and Lee, 2017; 트리 모형에는 TreeSHAP, Lundberg *et al.*, 2020)으로 교차검증하고, 모형 간 성능차의 유의성은 겹별 RMSE에 대한 Friedman 검정과 Nemenyi 사후분석(임계차 CD)으로 판정한다(Demšar, 2006). 다만 교차검증 겹은 완전히 독립이 아니므로, 겹 단위 Friedman, Nemenyi 검정 결과는 보수적으로 해석한다.

4. 실험 결과 및 분석

서론에서 제기한 연구질문에 대한 분석과 결론을 요약하면, RQ1(평가 타당성, Q1)은 4.1~4.2절에서 외삽 시 성능이 급격히 저하됨을, RQ2(예측 - 의사결정 연계, Q2 · Q3 · Q4 · Q6)는 4.4~4.6, 4.8절에서 낙관적 평가가 정비비용 증가로 이어짐을, RQ3(실환경 이전성, Q5)은 4.7절에서 모형별 상보적 강건성을 각각 확인한다. 4.3절 변수중요도 분석은 해석적 토대를 제공한다.

4.1 평가 프로토콜에 따른 예측 성능: 내삽 vs 외삽

<Table 2>는 무작위 5-겹 교차검증(P1)을 5개 시드로 반복한 결과이다. 두 열화계수 모두에서 MLP가 가장 우수하여 압축기 열화계수(k_{Mc})에 대해 RMSE 0.000362±0.000040(R^2=0.9994±0.0001), 터빈 열화계수(k_{Mt})에 대해 RMSE 0.000307±0.000054(R^2=0.9983±0.0006)를 기록하였고, 부스팅, 배깅이 그 뒤를 이었다. 선형 모형은 R^2=0.78~0.91로 현저히 낮았다. 모든 모형에서 5회 반복에 따른 표준편차는 평균의 10% 이내로 작아, 이하의 정성적 결론이 분할 · 초기화의 확률적 변동성에 기인한 것이 아님을 확인하였다.

그러나 이 수치는 완전요인 격자와 무작위 분할이 결합하여 산출된 내삽 성능이다. <Table 3>은 동일 모형을 운용점 외삽(P2)과 열화수준 외삽(P3)으로 평가한 RMSE와, 괄호 안에 외삽 R^2를 함께 제시한 것이다.

<Table 3>은 본 연구의 핵심 발견을 드러낸다. 첫째, 모든 모형이 외삽 조건에서 큰 폭으로 악화되었다. 최우수 모형 MLP의 k_{Mc} RMSE는 무작위 0.000370에서 운용점 외삽 0.008975로 약 24배, kNN은 0.001620에서 0.025825로 약 16배 증가하였다. 둘째, 더 결정적으로, 무작위 평가에서 $R^2 \geq 0.97$을 보이던 6개 비선형 모형을 포함한 8개 모형 중 MLP를 제외한 전부가 압축기 열화계수 외삽에서 R^2가 음수로 추락하였다(상수 예측보다 못함). 셋째, 5개 시드 반복실험 결과(3.3절), MLP의 부분적 일반화는 외삽 유형에 따라 갈렸다. 열화수준 외삽(P3)에서는 두 열화계수 모두 5개 시드 전부 안정적으로 양의 R^2를 유지하였으나(압축기 0.939±0.029, 터빈 0.726±0.128), 운용점 외삽(P2)에서는 R^2가 시드마다 크게 요동해 평균이 음수에 가까웠다(압축기 −0.08±0.51, 터빈 −1.79±1.22). 즉 원고에 보고된 단일 시드(seed=42) 기준 운용점외삽 R^2(압축기 0.49, 터빈 −0.32)는 반복실험 중 가장 우수했던 값일 뿐이며, 운용점 외삽에서는 MLP를 포함한 사실상 모든 모형이 일반화에 실패하고 열화수준 외삽에서만 MLP가 안정적으로 성공한다는 것이 더 정확한 결론이다. 넷째, 모형 간 우열 순위가 재편된다(4.2절 참조). 무작위에서 5위였던 kNN은 외삽에서 최하위권으로 내려앉고, 무작위 최하위였던 Ridge는 운용점 외삽에서 상위권으로 부상하였다. 격자 내삽에 최적화된 사례기반, 단일트리 모형일수록 외삽에 취약하고, 강한 정규화(Ridge)나 매끄러운 전역 근사(MLP)가 외삽에 상대적으로 강건하였다. 강한 정규화(Ridge)나 매끄러운 전역 근사(MLP)가 외삽에 상대적으로 강건하였다. 이러한 붕괴의 구조적 원인은 3.1절의 분산분해 분석으로 설명되는데, 열화 신호가 운용조건 대비 극히 작은 분산(약 1/500~1/10,000)만을 차지하기 때문이며, 무작위 분할에서만 존재하던 내삽 단서가 외삽에서 사라지는 것이 붕괴의 주된 요인이다.

Table 2. Five-fold cross-validation (P1) performance, mean±SD over 5 repeated seeds

Target	Model	MAE	RMSE	R^2
k_{Mc}	OLS	0.004588±0.000001	0.005840±0.000001	0.8425±0.0001
k_{Mc}	Ridge	0.005607±0.000001	0.006742±0.000001	0.7901±0.0001
k_{Mc}	kNN	0.000753±0.000005	0.001636±0.000020	0.9876±0.0003
k_{Mc}	CART	0.001332±0.000021	0.002331±0.000112	0.9747±0.0024
k_{Mc}	RF	0.000512±0.000003	0.000992±0.000010	0.9954±0.0001
k_{Mc}	XGBoost	0.000702±0.000006	0.000993±0.000005	0.9954±0.0000
k_{Mc}	LightGBM	0.000649±0.000002	0.000907±0.000007	0.9962±0.0001
k_{Mc}	MLP	0.000248±0.000025	0.000362±0.000040	0.9994±0.0001
k_{Mt}	OLS	0.001686±0.000000	0.002240±0.000001	0.9107±0.0001
k_{Mt}	Ridge	0.002752±0.000001	0.003519±0.000001	0.7797±0.0001
k_{Mt}	kNN	0.000652±0.000004	0.001434±0.000016	0.9633±0.0008
k_{Mt}	CART	0.000933±0.000013	0.001950±0.000044	0.9319±0.0030
k_{Mt}	RF	0.000354±0.000002	0.000733±0.000009	0.9904±0.0002
k_{Mt}	XGBoost	0.000498±0.000002	0.000761±0.000013	0.9897±0.0004
k_{Mt}	LightGBM	0.000458±0.000003	0.000698±0.000010	0.9913±0.0003
k_{Mt}	MLP	0.000183±0.000022	0.000307±0.000054	0.9983±0.0006

Table 3. RMSE by evaluation protocol, mean±SD over 5 repeated seeds

Target	Model	P1 Random	P2 LOLO (R^2)	P3 Decay (R^2)
k_{Mc}	OLS	0.005840±0.000001	0.034114±0.000000 (-8.52±0.00)	0.010095±0.000000 (-2.06±0.00)
k_{Mc}	Ridge	0.006742±0.000001	0.014467±0.000000 (-0.30±0.00)	0.013702±0.000000 (-4.65±0.00)
k_{Mc}	kNN	0.001636±0.000020	0.025825±0.000000 (-2.16±0.00)	0.015721±0.000000 (-6.43±0.00)
k_{Mc}	CART	0.002331±0.000112	0.025972±0.000025 (-2.13±0.01)	0.014613±0.000015 (-5.42±0.01)
k_{Mc}	RF	0.000992±0.000010	0.019639±0.001573 (-0.86±0.28)	0.014517±0.000133 (-5.34±0.12)
k_{Mc}	XGBoost	0.000993±0.000005	0.018057±0.000152 (-0.58±0.03)	0.012999±0.000094 (-4.08±0.07)
k_{Mc}	LightGBM	0.000907±0.000007	0.017707±0.000000 (-0.57±0.00)	0.012711±0.000000 (-3.86±0.00)
k_{Mc}	MLP	0.000362±0.000040	0.012390±0.002542 (-0.08±0.51)	0.001378±0.000379 (0.94±0.03)
k_{Mt}	OLS	0.002240±0.000001	0.016599±0.000000 (-16.51±0.00)	0.004186±0.000000 (-1.12±0.00)
k_{Mt}	Ridge	0.003519±0.000001	0.007173±0.000000 (-0.91±0.00)	0.008699±0.000000 (-8.17±0.00)
k_{Mt}	kNN	0.001434±0.000016	0.013627±0.000000 (-2.36±0.00)	0.006977±0.000000 (-4.90±0.00)
k_{Mt}	CART	0.001950±0.000044	0.013230±0.000190 (-2.17±0.07)	0.008255±0.000007 (-7.26±0.01)
k_{Mt}	RF	0.000733±0.000009	0.012483±0.000132 (-1.84±0.05)	0.007139±0.000064 (-5.18±0.11)
k_{Mt}	XGBoost	0.000761±0.000013	0.011526±0.000149 (-1.44±0.06)	0.006881±0.000034 (-4.74±0.06)
k_{Mt}	LightGBM	0.000698±0.000010	0.011682±0.000000 (-1.51±0.00)	0.006742±0.000000 (-4.51±0.00)
k_{Mt}	MLP	0.000307±0.000054	0.010876±0.002384 (-1.79±1.22)	0.001475±0.000331 (0.73±0.13)

함의는 분명하다. 함정 CBM 체계의 성능 요구도는 무작위 평가치가 아닌 외삽 평가치를 기준으로 설정되어야 하며, 모형 선정 시에도 내삽 정확도가 아닌 외삽 강건성을 우선 고려해야 한다. 본 연구는 함정 열화 예측의 달성 가능 성능을 현실적 일반화 조건에서 재정립함으로써, 시뮬레이터 성능치를 그대로 실배치 성능으로 해석해 온 관행에 정량적 교정을 제공한다.

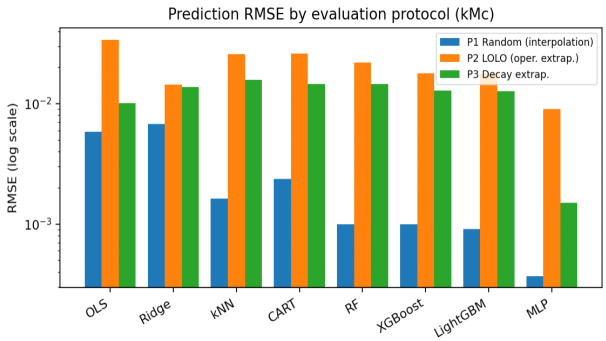


Figure 1. Prediction RMSE by evaluation protocol (k_{Mc}, log scale). For every model a large gap separates P1 Random (interpolation) from the extrapolation protocols (P2, P3); MLP attains the lowest error throughout.

4.2 통계적 유의성 검정

<Table 4>는 겹별 RMSE에 대한 Friedman 검정과 Nemenyi 사후분석 결과이다. 두 프로토콜 모두에서 모형 간 차이는 통계적으로 유의하였다($p < 0.001$).

순위표는 4.1절의 정성적 관찰을 정량적으로 뒷받침한다. Ridge는 무작위에서 8위(최하위)였으나 운용점 외삽에서 2위로 급상승하였고, 반대로 kNN은 평균순위 5.0에서 6.33으로, CART는 6.0에서 6.67로 하락하였다. 운용점 외삽 기준 임계차(CD=3.50)를 적용하면, 최상위 MLP(2.11), Ridge(2.89)는 최하위 CART(6.67), kNN(6.33)과 통계적으로 유의하게 구별된다. 모형 평가, 선정의 결론이 평가 프로토콜에 따라 통계적으로 유의하게 달라짐을 확인한 것이다.

4.3 변수중요도 분석: 순열중요도와 SHAP의 교차검증

압축기 열화계수 예측에서 순열 중요도는 압축기 출구 온도(T2)와 출구 압력(P2)을, SHAP은 T2와 고압터빈 출구 압력(P48)을 상위 변수로 식별하여, T2가 두 기법 모두에서 가장 중요한 변수로 일관되게 나타났다. 가스발생기 회전수(GGn)와 P48도 양쪽에서 공통 상위권이었다. 이는 압축기 오염이 압축

Table 4. Friedman/Nemenyi analysis (kMc; lower average rank is better)

Protocol	Friedman χ^2 (df=7)	Average rank (best → worst)	Nemenyi CD(0.05)
P1 Random	34.60	MLP 1.0, LightGBM 2.0, RF 3.4, XGBoost 3.6, kNN 5.0, CART 6.0, OLS 7.0, Ridge 8.0	4.70
P2 LOLO	27.96	MLP 2.11, Ridge 2.89, LightGBM 3.56, XGBoost 4.0, RF 5.0, OLS 5.44, kNN 6.33, CART 6.67	3.50

효율 저하를 통해 출구 온도 상승과 압력 변화로 발현된다는 가스터빈 성능 이론과 부합한다. 터빈 열화계수 예측에서는 압축기 출구 압력(P2)이 두 기법 모두에서 최상위였고, 가스터빈 축 토크(GTT)와 GGn이 공통 상위권이었다. 서로 다른 모형(MLP/XGBoost)과 서로 다른 원리(순열/Shapley 분해)에 기반한 두 분석이 핵심 변수에서 수렴함으로써, 변수중요도 해석이 특정 모형의 인공물이 아니라 데이터의 구조적 특성임을 확인하였다. 이 결과는 4.5절 센서 최적화에서 다시 교차검증된다.

4.4 비용기반 모형선택

<Table 5>는 정비 의사결정을 이진 결정 문제로 설정하고 (d^*=하위 25% 분위수, 80/20 무작위 홀드아웃) 비용비율별 기대비용을 산출한 결과이다.

Table 5. Misdetections, false alarms, and expected maintenance cost by cost ratio (k_{Mc})

Model	FN	FP	1:1	5:1	10:1	20:1
MLP	0	14	14	14	14	14
LightGBM	4	41	45	61	81	121
RF	7	30	37	65	100	170
XGBoost	8	39	47	79	119	199
kNN	10	30	40	80	130	230
CART	19	50	69	145	240	430
OLS	158	59	217	849	1,639	3,219
Ridge	284	61	345	1,481	2,901	5,741

내삽 홀드아웃에서는 MLP가 미탐 0건으로 모든 비용비율에서 최소 기대비용을 보였다. 주목할 점은 미탐비용이 클수록(고장위험이 큰 함정 정비 환경) 모형 간 비용 격차가 급격히 확대된다는 것이다. 1:1에서 RF(37)와 kNN(40)은 근소하나, 20:1에서는 RF(170) 대비 kNN(230), CART(430)로 벌어진다. 즉 정확도가 비슷해도 미탐 성향이 다른 모형은 고비용 환경에서 전혀 다른 선택지가 된다. 다만 이 우열은 내삽 조건의 것이며, 4.1절에서 kNN 등이 외삽에 취약했음을 고려하면 실배치 환경에서는 외삽 강건성과 비용 구조를 함께 반영한 모형 선정이 필요하다.

4.5 센서 포트폴리오 최적화

<Table 6>은 랜덤포레스트 기준 탐욕적 전진선택의 정확도 - 센서수 파레토 프론티어이다.

세 가지 시사점이 도출된다. 첫째, 전진선택이 가장 먼저 선택한 세 센서는 P48, T2, P2로, 4.3절의 순열중요도, SHAP이 식별한 상위 변수와 일치한다. 서로 다른 방법론이 동일한 핵심 센서로 수렴하여 센서 우선순위 해석의 신뢰성이 재확인된다. 둘째, RMSE는 센서 3개(P48, T2, P2)에서 최소(0.008203)에 도달하며, 그 이후 센서를 추가하면 오히려 성능이 단조 악화된다. 이는 다중공선성, 잡음 변수 추가에 따른 과적합으로, 무조건적 센서 증설이 최적이 아님을 분명히 보여준다. 셋째, 단 3개의 핵심 센서로 전체 13개 사용 시보다 더 나은 성능을 달성하므로, 함정 CBM 센서 구성 시 계측 예산을 압축기, 고압터빈 출구 상태량(P48, T2, P2)에 집중하는 것이 정확도, 비용 양면에서 합리적이다.

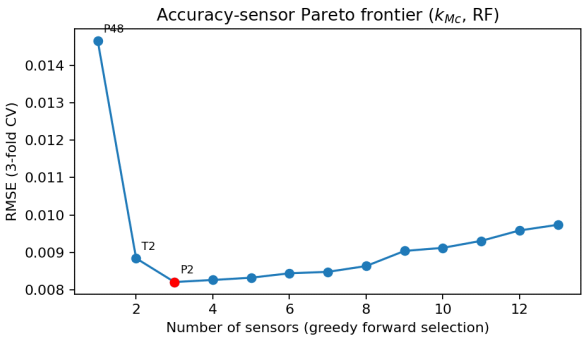


Figure 2. Accuracy – sensor Pareto frontier (k_{Mc}, RF). RMSE is minimized with three sensors (P48, T2, P2; red), and adding further sensors degrades performance.

4.6 불확실성 정량화와 점검 의사결정

<Table 7>은 최우수 모형 MLP에 대한 분할 등각예측(α=0.1, 목표 커버리지 90%) 결과이다.

Table 6. Accuracy – sensor Pareto frontier (k_{Mc}, RF, 3-fold CV RMSE)

No. of sensors	Added sensor	RMSE	No. of sensors	Added sensor	RMSE
1	P48	0.014663	8	GTT	0.008631
2	T2	0.008839	9	Pexh	0.009036
3	P2	0.008203	10	GGn	0.009119
4	mf	0.008258	11	GTn	0.009305
5	T48	0.008319	12	TIC	0.009584
6	lp	0.008437	13	Ts	0.009736
7	v	0.008476			

Table 7. Conformal prediction interval coverage and width (k_{Mc}, MLP)

Model	Target coverage	Empirical coverage	Mean width	vs. domain (0.05)	Quantile q
MLP	0.90	0.906	0.00136	2.7%	0.000681

경험적 커버리지가 목표 90%에 부합하여(0.906) 분포가정 없이 목표 신뢰수준을 달성하는 등각예측의 타당성이 확인되었다. 평균 구간폭은 0.00136으로 열화계수 정의역(0.05)의 2.7%에 불과하여, MLP 예측이 좁고 정보량이 큰 신뢰구간을 제공함을 보였다. 실무적으로 이 예측구간은 구간 상한이 임계 미만일 때만 점검지시하는 보수적 규칙으로 활용되어, 점예측만으로는 불가능한 미탐 위험의 명시적 통제를 가능하게 한다. 다만 본 구간폭은 내삽 조건의 추정치이므로, 외삽 환경에서는 4.1절의 큰 오차를 반영하여 구간이 넓어짐에 유의해야 한다.

4.7 잡음 강건성: 가우시안과 구조적 잡음

무잡음 데이터로 학습한 모형은 가우시안 잡음에 민감하여, 변수 표준편차의 1% 잡음에도 MLP의 RMSE는 0.000304에서 0.00337로 약 11배 증가하였다(시험만 오염). Table 8은 본 연구가 새로 추가한 구조적 잡음(드리프트, 결측, 이상치) 실험 결과로, 가우시안 백색잡음과 함께 MLP, RF를 비교한 것이다(시험만 오염, 10회 반복 평균±표준편차).

결과는 백색잡음 가정의 위험성과 모형별 강건성의 상보적 구조를 동시에 보여준다. 첫째, MLP는 1% 이상치에 RMSE가 0.00030→0.15079로 약 500배, 1% 결측에 약 72배 폭증하였다. 즉 내삽 성능 최우수 모형이 구조적 잡음에는 가장 취약하다. 둘째, 반면 RF는 이상치, 결측에 강건(각각 0.00360, 0.00323에 그침)하나 가우시안, 드리프트에는 상대적으로 더 민감하였다. 이는 트리 기반 모형이 표 형식(tabular) 데이터에서 신경망 대비 강건하다는 최근 벤치마크 결과(Grinsztajn *et al.*, 2022)와도 부합한다. 이는 트리 분할이 개별 이상치, 결측을 국소화하여 흡수하는 반면, 매끄러운 전역 함수인 MLP는 입력 교란을 전역적으로 증폭하기 때문으로 해석된다. 실무적 함의는 명확하다. 함정 CBM 실선 적용 시 이상치, 결측이 우려되는 환경에서는 트리 앙상블을, 정밀 계측이 보장되는 환경에서는 신경망을 채택하거나 양자를 결합하는 전략이 필요하며, 무엇보다 이상치 탐지, 결측 대치, 센서 보정 등 전처리 강건성이 예측 모형과 동등하게 중요하다. 정리하면, 정밀 계측 환경(결측·이상치 드묾)은 MLP(최저 RMSE, <Table 2>), 결측·이상치 빈발 환경은 RF(1% 이상치에서 RMSE 약 40분의 1, <Table 8>)를 우선 채택하고, 불확실한 환경에서는 두 모형을 앙상블로 결합하거나 전처리를 병행하며, 어떤 경우든 외삽 프로토콜(4.1절)로 재검증해야 한다.

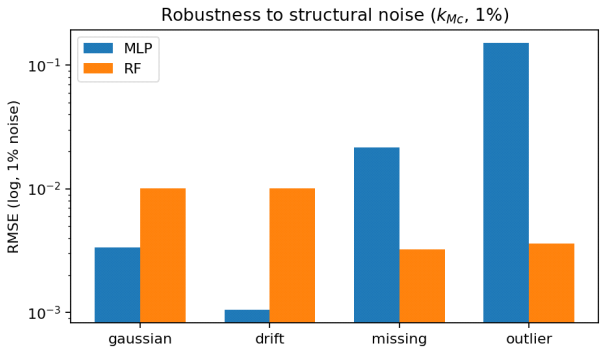


Figure 3. Robustness to structural noise (k_{Mc}, 1% level, log scale). MLP is extremely fragile to outliers and missingness, whereas RF is robust to them.

4.8 예측 - 의사결정 통합: 상태기반 검사, 정비 정책 최적화

<Table 9>는 D4의 정책 최적화 결과로, 본 연구의 예측 - 평가 - 의사결정을 하나로 잇는 종합 분석이다. 추정오차 σ는 4.1절에서 측정한 MLP의 실측 RMSE에 직접 연동하여, 내삽(낙관) σopt(무작위 CV), 외삽(현실) σreal(운용점 외삽)을 사용하였다. 비용, 열화수명 파라미터는 3.4절에서 밝힌 바와 같이 정

Table 8. RMSE under noise types and levels, mean±SD over 10 repetitions

Model	Noise	clean	1%	2%	5%
MLP	Gaussian	0.000304	0.003426±0.000061	0.006753±0.000132	0.016219±0.000218
MLP	Drift	0.000304	0.001058±0.000000	0.001997±0.000000	0.004778±0.000000
MLP	Missing	0.000304	0.020564±0.001084	0.028854±0.001615	0.044667±0.001536
MLP	Outlier	0.000304	0.161928±0.010143	0.220732±0.007912	0.350686±0.008596
RF	Gaussian	0.000954	0.010027±0.000156	0.011541±0.000184	0.015326±0.000173
RF	Drift	0.000954	0.010137±0.000000	0.011056±0.000000	0.018619±0.000000
RF	Missing	0.000954	0.003456±0.000185	0.004626±0.000388	0.007492±0.000179
RF	Outlier	0.000954	0.003786±0.000370	0.005246±0.000338	0.007885±0.000271

Table 9. Inspection/maintenance policy optimization (Monte Carlo, long-run cost rate; csp=1, coh=10, cfail=100; life ~ U(40, 80) months)

Scenario	Optimal interval τ*	Optimal threshold d*	Long-run cost rate
Perfect information (σ→0) lower bound	18 months	0.967	0.260
Realistic error (σreal=0.00897) optimum	18 months	0.976	0.367
Optimistic (σopt=0.00037) design, deployed	18 months	0.967	0.538

성적 비용 순서를 반영한 예시값이므로, 이하 제시하는 비용률의 절대 수준이 아니라 낙관설계 대비 현실배치의 상대적 페널티(+46.6%)와 그 부호(양(+))에 연구의 함의가 있다.

세 가지 시사점이 도출된다. 첫째, 추정오차가 커질수록 최적 정비임계 가 상향된다(0.967→0.976). 센서 추정이 부정확할수록 더 보수적으로(더 이른 시점에) 오버홀해야 고장위험을 통제할 수 있으며, 이는 예측 불확실성이 최적 정비 정책의 구조 자체를 바꾼다는 의사결정론적 통찰이다. 둘째, 가장 중요하게는, 낙관적 오차(σopt)로 설계한 정책(τ*=18, d*=0.967)을 현실 환경에 배치하면 비용률이 0.538로 현실최적(0.367) 대비 약 +46.6% 증가한다. 무작위 CV가 야기하는 성능 과대평가가 단순한 평가상의 문제에 그치지 않고 실제 정비비용 증가로 직접 전이됨을 정량적으로 보인 것이다. 셋째, 완전정보하한(0.260)과 현실최적(0.367)의 격차는 센서 추정오차가 유발하는 불가피한 비용으로, 4.5절의 센서 보강, 4.7절의 전처리 강건성이 이 격차를 줄이는 직접적 수단이다.

이 결론의 견고성을 확인하기 위해, 비용비율, 외삽오차 수준, 열화 수명분포를 변화시킨 민감도 분석을 수행하였다(Table 10). 낙관설계의 추가비용은 모든 조건에서 양(+)으로 유지되었으며, 특히 고장/오버홀 비용비율이 커질수록(즉 안전이 중요할수록) 페널티가 급격히 증가(+14.5% → +246.6%)하였다. 외삽오차 수준과 열화 수명분포 변화에 대해서도 페널티는 +32~49% 범위에서 안정적이었다. 즉 낙관적 내삽 평가로 정비정책을 설계하면 실환경 비용이 증가한다는 본 연구의 핵심 결론은 특정 파라미터 선택의 산물이 아니라 광범위한 가정에 걸쳐 견고하며, 고장비용이 큰 함정 추진체계와 같은 안전중요 체계일수록 그 함의가 더 커진다.

이 분석은 본 연구의 세 축을 통합한다(핵심 논리사슬은 5장에서 재정리한다). 함정 CBM+ 체계의 가치는 궁극적으로 정비비용, 가동률로 환산되므로, 예측 정확도 경쟁이 아니라 이 통합적 관점에서 연구가 설계되어야 한다는 것이 본 연구의 핵심 주장이다. 본 정책 분석은 명시한 비용, 열화 가정에 기반한 프레임워크이며, Table 10의 민감도 분석으로 결론의 견고성을 확인하였다.

5. 결 론

본 연구는 함정 추진체계 상태기반정비를 위한 가스터빈 성능저하 예측 문제를, 단순 모형 비교를 넘어 평가 방법론, 의사결정 연계, 실환경 이전성의 세 측면에서 재구성하였다. 본 연구를 관통하는 논리는 다음과 같다: 평가 타당성(외삽 평가)이 확보되어야 예측오차를 현실적으로 추정할 수 있고, 이 예측오차라야 정비 의사결정 비용을 올바르게 계산할 수 있다. 이 논리사슬을 따라 얻은 주요 결과는 다음과 같다.

첫째, 완전요인 시뮬레이터 격자라는 구조적 특성으로 인해 관례적 무작위 교차검증이 열화 예측 성능을 체계적으로 과대평가함을 실측으로 보였다. 운용점 외삽에서 RMSE는 무작위 평가 대비 최대 약 24배 증가하였고, 무작위에서 $R^2 \geq 0.97$이던 6개 비선형 모형을 포함한 8개 모형 중 MLP를 제외한 전부가

Table 10. Sensitivity of the policy conclusion
(extra cost of an optimistic design deployed in reality, %)

Factor	Setting	Realistic-optimal cost rate	Extra cost of optimistic design
Cost ratio	5:1	0.337	+14.5%
	10:1 (base)	0.367	+46.6%
	20:1	0.395	+113.6%
	40:1	0.420	+246.6%
Extrapolation error	0.005	0.309	+32.1%
	0.00897 (base)	0.367	+46.6%
	0.013	0.415	+48.8%
	0.020	0.466	+47.1%
Degradation life	U(30, 60)	0.489	+48.2%
	U(40, 80) (base)	0.367	+46.6%
	U(60, 120)	0.248	+41.5%

압축기 열화계수 외삽에서 R^2가 음수로 추락하였다. 5개 시드 반복실험 결과 MLP조차 운용점 외삽에서는 R^2가 불안정했고(압축기 −0.08±0.51, 터빈 −1.79±1.22), 안정적인 양의 R^2는 열화수준 외삽에서만 확인되어(압축기 0.939±0.029, 터빈 0.726±0.128) 운용점 외삽에서는 사실상 모든 모형이 일반화에 실패함을 보였다. Friedman, Nemenyi 검정 결과 모형 순위도 유의하게 재편되었다(Ridge 8위→2위, kNN 5위→하위권).

둘째, 예측을 정비 의사결정으로 연계함으로써 비용기반 모형선택, 정확도 - 센서 파레토(핵심 3개 센서 P48, T2, P2로 최적), 등각예측 기반 신뢰구간(경험적 커버리지 0.906)을 제시하고, 나아가 상태기반 검사, 정비 정책을 장기 비용률 최소화 문제로 구성하였다. 특히 낙관적(내삽) 오차로 설계한 정책을 현실(외삽) 환경에 배치하면 정비비용이 약 46.6% 증가하며, 민감도 분석 결과 이 페널티는 광범위한 비용, 열화 가정에서 견고하고(+14.5~246.6%) 고장비용이 클수록 급증함을 보여, 평가 타당성이 의사결정 비용으로 직접 전이됨을 정량화하였다.

셋째, 구조적 잡음 분석에서 동일 1% 수준의 이상치가 MLP의 RMSE를 약 500배, 결측이 약 72배 증가시켜 가우시안 잡음을 압도하는 반면, 트리 앙상블(RF)은 이상치, 결측에 강건한 상보적 구조를 규명하였다(4.7절). 이는 백색잡음 가정만으로 강건성을 평가하는 것이 실배치 위험을 크게 과소평가하며, 환경에 따른 모형 선택과 전처리 강건성이 필수임을 보여준다.

본 연구의 한계와 향후 과제는 다음과 같다. 사용 데이터는 시뮬레이터 기반으로 과도상태(transient)가 배제되어 있어 실제 함정 데이터 확보와 검증이 필요하다. 아울러 본 연구가 주입한 구조적 잡음(3.5절)은 통계적으로 단순화한 모형이라 실제 센서의 상관된 다중 이상, 비무작위적 결측 등은 반영하지 못하므로, 강건성 수치는 정성적・상대적 비교로 해석되어야 하며 실제 함정 데이터(예: 한국 해군 ICAS)를 통한 재검증이 후속 과제이다. 또한 본 연구는 단일 시점 정상상태 관측치 기반의 열화 수준 추정을 다루었으며, 시계열 관점의 열화 진행 예측과 잔존수명(RUL) 추정(Si *et al.*, 2011; Saxena *et al.*, 2008; Ferreira and Gonçalves, 2022; Kim *et al.*, 2017), 다중출력 동시 추정으로의 확장이 요구된다. 정책 최적화(4.8절)는 비용, 열화 가정에 대한 민감도 분석으로 결론의 견고성을 확인하였으나, 실제 함정 정비비용, 열화 이력 데이터를 활용한 파라미터 보정은 후속 과제이다. 이러한 확장에도 불구하고, 본 연구는 함정 열화 예측의 달성 가능 성능을 현실적 일반화 조건에서 재정립하고 예측 - 의사결정 연계의 정량적 틀을 제시함으로써 한국형 함정 CBM+ 연구의 방법론적 출발점을 제공한다.

생성형 AI 활용 공개

본 논문의 작성 과정에서 저자는 Claude(Anthropic)를 데이터 전처리, 분석 코드 작성, 실험 수행 보조 및 원고 초안 작성을 위해 사용하였다. 저자는 해당 도구 사용 후 그 내용을 검토 및 수정하였으며, 논문의 모든 내용에 대한 책임은 저자에게 있다.

참고문헌

Achouch, M., Dimitrova, M., Ziane, K., Sattarpanah Karganroudi, S., Dhouib, R., Ibrahim, H., and Adda, M. (2022), On predictive maintenance in Industry 4.0: Overview, models, and challenges, *Applied Sciences*, **12**(16), 8081.

Andelic, N., Baressi Segota, S., Lorencin, I., and Car, Z. (2020), Estimation of gas turbine shaft torque and fuel flow of a CODLAG propulsion system using genetic programming algorithm, *Pomorstvo*, **34**(2), 323-337.

Angelopoulos, A. N. and Bates, S. (2023), Conformal prediction: A gentle introduction, *Foundations and Trends in Machine Learning*, **16**(4), 494-591.

Breiman, L. (2001), Random forests, *Machine Learning*, **45**(1), 5-32.

Carvalho, T. P., Soares, F. A. A. M. N., Vita, R., Francisco, R. da P., Basto, J. P., and Alcalá, S. G. S. (2019), A systematic literature review of machine learning methods applied to predictive maintenance, *Computers & Industrial Engineering*, **137**, 106024.

Chen, T. and Guestrin, C. (2016), XGBoost: A scalable tree boosting system, *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785-794.

Cipollini, F., Oneto, L., Coraddu, A., Murphy, A. J., and Anguita, D. (2018a), Condition-based maintenance of naval propulsion systems with supervised data analysis, *Ocean Engineering*, **149**, 268-278.

Cipollini, F., Oneto, L., Coraddu, A., Murphy, A. J., and Anguita, D. (2018b), Condition-based maintenance of naval propulsion systems: Data analysis with minimal feedback, *Reliability Engineering and System Safety*, **177**, 12-23.

Coraddu, A., Oneto, L., Ghio, A., Savio, S., Anguita, D., and Figari, M. (2016), Machine learning approaches for improving condition-based maintenance of naval propulsion plants, Proceedings of the Institution of Mechanical Engineers, *Part M: Journal of Engineering for the Maritime Environment*, **230**(1), 136-153.

Demšar, J. (2006), Statistical comparisons of classifiers over multiple data sets, *Journal of Machine Learning Research*, **7**, 1-30.

Ferreira, C. and Gonçalves, G. (2022), Remaining useful life prediction and challenges: A literature review on the use of machine learning methods, *Journal of Manufacturing Systems*, **63**, 550-562.

Grinsztajn, L., Oyallon, E., and Varoquaux, G. (2022), Why do tree-based models still outperform deep learning on typical tabular data?, *Advances in Neural Information Processing Systems*, **35**, 507-520.

Jardine, A. K. S., Lin, D., and Banjevic, D. (2006), A review on machinery diagnostics and prognostics implementing condition-based maintenance, *Mechanical Systems and Signal Processing*, **20**(7), 1483-1510.

Javanmardi, A. and Hüllermeier, E. (2023), Conformal prediction intervals for remaining useful lifetime estimation, *International Journal of Prognostics and Health Management*, **14**(2), 3417.

Jeong, H. S. (2021), A study on the condition based maintenance development for the expansion of CBM adoption in weapon systems, *Journal of the Korea Academia-Industrial cooperation Society*, **22**(8), 631-638.

Kapoor, S. and Narayanan, A. (2023), Leakage and the reproducibility crisis in machine-learning-based science, *Patterns*, **4**(9), 100804.

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., and Liu, T.-Y. (2017), LightGBM: A highly efficient gradient boosting decision tree, *Advances in Neural Information Processing Systems*, **30**, 3146-3154.

Kim, N.-H., An, D., and Choi, J.-H. (2017), *Prognostics and Health Management of Engineering Systems: An Introduction*, Springer, Cham.

Kingma, D. P. and Ba, J. (2015), Adam: A method for stochastic optimization, *Proceedings of the 3rd International Conference on Learning Representations (ICLR)*.

Lee, S.-W., Lee, S.-U., Park, G.-J., and Hur, J.-W. (2024), A study on the procedures for the application of Korean weapon system condition based maintenance (CBM+), *Journal of the Korea Academia-Industrial cooperation Society*, **25**(5), 365-372.

Lei, Y., Li, N., Guo, L., Li, N., Yan, T., and Lin, J. (2018), Machinery health prognostics: A systematic review from data acquisition to remaining useful life prediction, *Mechanical Systems and Signal Processing*, **104**, 799-834.

Lundberg, S. M. and Lee, S.-I. (2017), A unified approach to interpreting model predictions, *Advances in Neural Information Processing Systems*, **30**, 4765-4774.

Lundberg, S. M., Erion, G., Chen, H., DeGrave, A., Prutkin, J. M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N., and Lee, S.-I. (2020), From local explanations to global understanding with explainable AI for trees, *Nature Machine Intelligence*, **2**(1), 56-67.

Oh, K. (2016), Development of Korean condition based maintenance system to monitor naval weapon systems, *Journal of the Society of Aerospace System Engineering*, **10**(4), 67-74.

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, É. (2011), Scikit-learn: Machine learning in Python, *Journal of Machine Learning Research*, **12**, 2825-2830.

Roberts, D. R., Bahn, V., Ciuti, S., Boyce, M. S., Elith, J., Guillera-Arroita, G., Hauenstein, S., Lahoz-Monfort, J. J., Schröder, B., Thuiller, W., Warton, D. I., Wintle, B. A., Hartig, F., and Dormann, C. F. (2017), Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure, *Ecography*, **40**(8), 913-929.

Saxena, A., Goebel, K., Simon, D., and Eklund, N. (2008), Damage propagation modeling for aircraft engine run-to-failure simulation, *Proceedings of the International Conference on Prognostics and Health Management (PHM)*, 1-9.

Shafer, G. and Vovk, V. (2008), A tutorial on conformal prediction, *Journal of Machine Learning Research*, **9**, 371-421.

Si, X.-S., Wang, W., Hu, C.-H., and Zhou, D.-H. (2011), Remaining useful life estimation – A review on the statistical data driven approaches, *European Journal of Operational Research*, **213**(1), 1-14.

Simion, D., Postolache, F., Fleacă, B., and Fleacă, E. (2024), AI-driven predictive maintenance in modern maritime transport—Enhancing operational efficiency and reliability, *Applied Sciences*, **14**(20), 9439.

Son, M. and Kim, Y.-G. (2021), A study on the reflection of condition-based maintenance requirement in the defense specification, *Journal of the Korean Society for Quality Management*, **49**(3), 269-279.

UCI Machine Learning Repository (2014), Condition Based Maintenance of Naval Propulsion Plants Data Set, https://archive.ics.uci.edu/dataset/316/.

Vovk, V., Gammerman, A., and Shafer, G. (2005), *Algorithmic Learning in a Random World*, Springer, New York.

저자소개

손창호: 육군사관학교를 졸업하고 2002년 임관하였으며, 노스캐롤라이나주립대학교에서 산업공학 석사, 서울대학교에서 산업공학 박사학위를 취득하였다. 현재 육군3사관학교 인공지능・시스템과학과 정교수로 재직하고 있다. 연구분야는 국방데이터분석, 시스템 최적화, 인공지능이다.