

# 센서 데이터의 시계열 특성을 고려한 선박 연료 소모량 예측 프레임워크

이현하<sup>1</sup> · 김도희<sup>2</sup> · 추정호<sup>3</sup> · 조상민<sup>3</sup> · 배혜림<sup>4\*</sup>

<sup>1</sup>부산대학교 산업공학과 / <sup>2</sup>부산대학교 인간 중심-탄소 중립 글로벌 공급망 연구센터  
<sup>3</sup>부산대학교 산업공학과 산업데이터공학융합전공 / <sup>4</sup>부산대학교 데이터사이언스전문대학원

## A Forecasting Framework of Ship Fuel Consumption using Temporal Patterns in Sensor Data

Hyun Ha Lee<sup>1</sup> · Dohee Kim<sup>2</sup> · Jung-Ho Choo<sup>3</sup> · Sangmin Jo<sup>3</sup> · Hyerim Bae<sup>4</sup>

<sup>1</sup>Department of Industrial Engineering, Pusan National University

<sup>2</sup>Safe & Clean Supply Chain Research Center, Pusan National University

<sup>3</sup>Major in Industrial Data Science & Engineering, Department of Industrial Engineering, Pusan National University

<sup>4</sup>Department of Data Science, Graduate School of Data Science, Pusan National University

The maritime industry plays a central role in global trade but also contributes significantly to environmental degradation due to high fuel consumption. Accordingly, the International Maritime Organization (IMO) has introduced environmental regulations, including the Carbon Intensity Indicator (CII). Complying with these regulations requires accurate fuel consumption forecasting to support optimal voyage planning. However, ship operation data frequently contain missing and anomalous values due to sensor failures, which reduce the accuracy of forecasts that rely on time-series characteristics. Although previous studies have applied machine learning and time-series models, they often fail to employ temporal dependencies or maintain data continuity. This study proposes a time-series forecasting framework that separates movement and anchorage states and performs imputation using environmental variables and similar trajectory group. The proposed method effectively leverages sensor data, despite frequent missing and anomalous values, and improves forecasting performance. The proposed framework facilitates voyage optimization by enabling accurate speed estimation and ensuring compliance with CII regulations.

**Keywords:** Ship Fuel Consumption, CII, Ship Sensor Data, Time-Series Forecasting, Missing Data Imputation

### 1. 서론

환경 문제에 대한 국제 사회의 인식이 높아지면서 전 세계적으로 탄소 중립 실현에 대한 요구가 커지고 있다(Rissman *et al.*, 2020). 이에 따라 해운 산업 또한 탈탄소화라는 시대적 과제에 직면하게

되었다(Dong *et al.*, 2022). 이러한 변화에 대응하고자 2020년 국제 해사 기구(International Maritime Organization, IMO)는 'IMO 2020' 규제의 일환으로 탄소집약도지수(Carbon Intensity Indicator, CII) 규제를 발효하였다(Bergstrom *et al.*, 2023). CII는 선박의 탄소 배출 효율성을 평가하는 지표로 연간 연료

이 논문은 부산대학교 기본연구지원사업(2년)에 의하여 연구되었음

\* 연락저자 : 배혜림 교수, 46241 부산광역시 금정구 부산대학로63번길 2 (장전동) 제10공학관 10623호, Tel : 051-510-2733, Fax : 051-512-7603, E-mail : hrbae@pusan.ac.kr

2025년 5월 21일 접수; 2025년 7월 15일 수정본 접수; 2025년 9월 30일 게재 확정.

소모량을 기반으로 산출되며, 해당 값에 따라 A부터 E까지의 등급이 선박에 부여된다(Kim *et al.*, 2024). 기준을 충족하지 못한 선박(D등급 이하)의 경우 운항 제한 등의 시정 조치가 취해지므로 선사들은 규제를 엄격히 준수해야 한다(Grapsas, 2023). 더욱이 감축률이 매년 강화됨에 따라 CII 허용값(C등급 중간값)이 지속적으로 하향 조정되고 있어, 선사들은 규제 준수를 위한 운영 부담이 가중되고 있다(Trakakis *et al.*, 2023). CII 기준을 충족하기 위해서는 연간 연료 소모량을 효율적으로 관리해야 한다. 연료 소모량은 선박의 항구 대기 시간 및 속도와 항로 선택에 따라 크게 달라진다(Shin *et al.*, 2021). 따라서 사전에 연료 소모량을 정확하게 예측하고, 이를 기반으로 항구 대기 시간 단축을 고려한 최적 속도 및 항로를 설정하는 것이 CII 규제 준수를 위한 핵심 전략이다(Rauca and Batrinca, 2023).

정확한 연료 소모량 예측을 위해서는 실시간 위치, 속도 등 운항 중 선박 상태를 추적하는 Automatic Identification System (AIS) 데이터의 활용이 필수적이다(Kim *et al.*, 2023). 또한, 해양 기상 데이터, 엔진 센서 데이터를 함께 활용할 경우 예측 정확도를 더욱 높일 수 있다(Yang *et al.*, 2023; Sim *et al.*, 2022). 그러나 이러한 데이터는 운항 중 실시간으로 선박에서 수집되는 센서 데이터로, 통신 장애나 센서 오류 등의 요인으로 인해 결측값이 빈번하게 발생하는 문제가 존재한다(Perera *et al.*, 2017). 이는 연료 소모량 예측 시 연속성을 고려해야 하는 특성을 갖는 선박 센서 데이터(Liu *et al.*, 2024)의 시간적 패턴을 효과적으로 활용하는 데 어려움을 유발하는 주요 요인 중 하나로 지적된다.

따라서 센서 데이터를 활용하여 연료 소모량을 정확히 예측하기 위해서는 센서 오류나 통신 문제로 인한 결측값을 적절히 처리하고, 데이터의 연속성을 보존하여 시계열 특성을 효과적으로 활용해야 한다. 다수의 선행 연구는 이러한 결측 문제를 보완하기 위해 머신러닝 모델을 적용하였으나, 시계열 정보를 충분히 반영하지 못하여 연료 소모량의 시간적 변화 패턴을 학습하는 데 한계를 보였다(Hu *et al.*, 2021; Melo *et al.*, 2024; Yan *et al.*, 2020). 이를 개선하고자 일부 연구에서는 Long Short Term Memory (LSTM)과 같은 시계열 모델을 적용하였지만 항해 상태의 정확한 식별이나 데이터 연속성에 대한 고려가 부족하였다(Lei *et al.*, 2021; Yuan *et al.*, 2021; Shi *et al.*,

2024). 특히, 시계열 데이터의 결측값을 단순 삭제하는 기존 방식은 데이터의 연속성을 훼손할 수 있으며(Barzi, 2004), 선형 및 곡선 보간은 랜덤하게 발생한 결측값에는 효과적이지지만, 연속적인 결측 구간에서는 성능이 저하되는 한계가 존재한다(Shi *et al.*, 2020).

본 연구는 이러한 한계점을 해결하기 위하여 선박 센서 데이터의 시계열 특성을 고려한 선박 연료 소모량 시계열 예측 프레임워크를 제안한다. 본 연구의 주요 기여는 다음과 같다.

- (1) 도메인 지식 기반 변수 구성과 실시간으로 운항 중 수집되는 다양한 시계열 센서 데이터의 통합을 통해 항적별 연료 소모량 예측 정확도를 제고한다.
- (2) 실제 선박의 운항 특성을 기반으로 정박 및 항해 상태를 구분하여 학습 노이즈를 제거하고 유의미한 연료 소모 패턴에 집중한다.
- (3) 변수 상관성을 고려한 유사 항적 기반 보간을 수행하여 시계열 데이터의 시간적 연속성을 보존하여 결측값을 처리한다.

결과적으로 본 연구는 보다 정확한 선박 연료 소모량 예측을 위한 접근 방안을 제안함으로써, 선사의 CII 규제 준수 대응을 위한 운항 계획 수립에서의 활용과 더불어 해운업의 탈탄소화 촉진에 기여할 것으로 기대된다.

본 논문의 구성은 다음과 같다. 제2장에서는 선박 연료 소모량 예측 관련 기존 연구를 검토하고, 제3장에서는 본 연구에서 제안하는 방법론을 소개한다. 제4장에서는 실험 결과 및 분석을 통해 제안 방법론의 성능을 평가한다. 마지막으로, 제5장에서는 본 연구의 결론을 정리하고 향후 연구 방향을 제시한다.

## 2. 관련 연구

선박 연료 소모량 예측 관련 연구는 크게 세 가지 방향으로 진행되어 왔다. 2.1절에서는 머신러닝 기법을 활용한 예측 연구를, 2.2절에서는 데이터의 시간적 연속성을 고려한 시계열 예측 모델 활용 연구를, 2.3절에서는 외부 환경 변수를 고려한 선박 연료 소모량 연구를 정리한다. 관련 연구의 주요 내용은 <Table 1>에 요약하여 정리하였다.

**Table 1.** Summary of Related Studies

Author(Year)	Utilized Data	Utilized Method	Feature Engineering	Missing Value Imputation
Yan <i>et al.</i> (2020)	Noon report	RandomForest	O	-
Zhou <i>et al.</i> (2022)	Noon report	SVR(Support Vector Regressor), RandomForest, etc.	O	-
Ma <i>et al.</i> (2023)	Sensor data	CatBoost, Gradient Boosting, etc.	O	-
Kim <i>et al.</i> (2023)	Sensor data	XGBoost	O	-
Lei <i>et al.</i> (2021)	Sensor data	LSTM	O	X
Zhang <i>et al.</i> (2024)	Sensor data	Bidirectional-LSTM (Bi-LSTM)	O	X
Shi <i>et al.</i> (2024)	Sensor data	LSTM	X	X
Ours	Sensor data	LSTM	O	O

**2.1 머신러닝 모델을 활용한 연료 소모량 예측 연구**

인공지능 및 정보 기술의 발전으로 선박 속도, 흘수, 풍속, 파고 등 다양한 운항 데이터를 활용한 연료 소모량 예측 연구가 활발히 진행되고 있다. 기존의 물리 기반 모델은 연료 소모량과 운항 변수 간 관계를 이론적으로 정립하였으나, 실제 환경에서의 비선형적 상호작용을 충분히 반영하지 못하는 한계가 있었다(Cepowski and Drozd, 2023). 이를 해결하기 위해 XGboost와 SVR와 같은 머신러닝 기법이 연료 소모량 예측을 위하여 도입되었다.

Zhou *et al.*(2022)은 SVR, RandomForest 등의 다양한 머신러닝 모델을 비교하여 모델에 따른 성능 차이를 비교하였다. Yan *et al.*(2020)은 RandomForest 모델을 활용한 예측 결과를 기반으로 속도 최적화를 수행하여 기존 방식 대비 연료 소비를 2~7% 절감할 수 있음을 확인하였다. 또한, Ma *et al.*(2023)은 여러 머신러닝 모델을 적용한 하이브리드 모델을 제안하여 예측 성능을 향상시켰으며 해양환경 관련 변수의 결합이 성능 개선에 기여함을 보였다.

이처럼 머신러닝 기반 연구는 변수 간의 비선형 관계를 학습하는 데 유용하지만, <Table 1>과 같이 많은 연구가 정오 보고서(Noon report)에 기반한다는 한계가 있다. 정오 보고서는 항해 중 하루 1회 수기로 작성되는 기록으로, 시간 해상도가 낮고 연속성이 부족하여 선박 운항 중 시간의 흐름에 따른 데이터의 패턴을 충분히 활용하기 어렵다(Smith *et al.*, 2013). 한편, 일부 연구는 시간 해상도가 높고 연속적인 센서 데이터를 사용했음에도 불구하고, 기존 머신러닝 모델은 개별 시점의 데이터를 독립적으로 처리하기 때문에 시간적 연속성을 효과적으로 반영하지 못한다. 이러한 문제를 해결하기 위해서는 시계열 모델 기반 접근법이 필요하며, Karagiannidis and Themelis(2021)는 연속적인 센서 데이터와 시간 흐름을 반영한 학습이 선박 연료 소모량 예측의 정확도 향상에 중요하다고 강조하였다.

**2.2 시계열 예측 모델을 활용한 연료 소모량 예측 연구**

선박 연료 소모량 데이터가 시간에 따른 변화 패턴을 보이지만 기존 머신러닝 모델은 이를 활용하지 못한다는 한계가 지적되었다(Lei *et al.*, 2021; Yuan *et al.*, 2021; Shi *et al.*, 2024). Lei *et al.*(2021)는 딥러닝 기반의 LSTM이 과거 데이터의 영향을 효과적으로 학습하여 연료 소모량 예측에 유용함을 입증하였다. Wang *et al.*(2023)은 유전 알고리즘을 활용한 하이퍼파라미터 최적화 기법을 적용하여 LSTM 모델의 예측 오류를 감소시켰다. 또한, Zhang *et al.*(2024)은 Bi-LSTM과 Attention 메커니즘을 결합한 모델을 제안하여 복잡한 데이터 패턴을 보다 효과적으로 처리하였다. Liu *et al.*(2024)은 Temporal Convolution Network-Gated Recurrent Unit-Multi Head Self Attention (TCN-GRU-MHSA) 신경망 모델을 개발하여 센서 데이터의 자기상관성을 분석하고 과거 시점이 후속 시점에 미치는 영향을 규명하였다. Chen *et al.*(2024)은 단일 스텝 예측 위주의 기존 연구가 장기 예측 시 성능 저하 문제

를 겪는다고 지적하며, 다단계 예측의 오차 누적 문제를 완화하기 위해 Ensemble Empirical Mode Decomposition-LSTM(EEMD-LSTM) 및 Bi-LSTM을 적용하였다.

시계열 모델을 활용한 연구들은 기존 머신러닝 모델의 한계를 극복하며 연료 소모량 예측의 정확도를 향상시키는 데 기여하였다. 그러나 다음과 같은 몇 가지 한계점이 존재한다. 첫째, 결측값 및 이상치를 단순 삭제하는 방식은 데이터의 시간적 연속성을 훼손할 수 있다(Lei *et al.*, 2021; Shi *et al.*, 2024; Wang *et al.*, 2023). 특히 선형 및 곡선 보간 기법은 무작위 결측에는 효과적일 수 있으나 연속적인 결측이 발생할 경우 데이터 패턴을 과도하게 단순화하여 예측 성능 저하를 초래할 수 있다(Shi *et al.*, 2020). 따라서 데이터의 시간적 연속성 보존을 위해서는 적절한 전처리 방법이 요구되며, 본 연구는 데이터의 특성을 고려하여 변수 간의 상관성을 반영한 유사 항적 기반 보간을 수행한다. 둘째, 연료 소모량은 항해 상태에 따라 뚜렷한 패턴 차이를 보이며, 이는 시계열 예측의 정확도에 직접적인 영향을 미친다. 그러나 기존 연구의 대부분은 정오 보고서 기반 항적 정보를 활용한 예측을 수행하였다. 이는 수기 입력에 의존하기 때문에 인적 오류(human error)가 개입될 수 있고 항적 정보가 실제 항해 상태와 일치하지 못한다는 한계가 존재한다. 따라서 본 연구는 실제 운항 데이터를 기반으로 항해 및 정박 상태 구분을 통한 항적 분류 방법론을 새롭게 제안한다. 셋째, 시계열 모델의 예측 성능에 중요한 영향을 미치는 도메인 지식 기반 파생 변수 생성과 변수 선택이 충분히 고려되지 않았다. 특히 해양 환경 변수는 연료 소모량과 밀접한 관련이 있으므로 예측 성능 향상을 위해서는 이를 효과적으로 활용하여야 한다(Wang *et al.*, 2023).

**2.3 외부 환경 변수를 고려한 선박 연료 소모량 연구**

선박의 연료 소모량은 해양 환경의 영향을 크게 받으며 특히 풍속, 파고, 해류 등은 선박 저항에 직접 작용하여 연료 소비를 의미 있게 변화시킨다. 이에 따라, 최근 연구들은 다양한 변수들이 미치는 영향을 정량적으로 분석하고 규명하였다(Zhou *et al.*, 2023).

Shin *et al.*(2021)은 AIS, 해양 기상, 엔진 센서 데이터를 활용한 분석을 통해 유속, 파랑력, 파고, 풍속이 연료 소모량에 영향을 크게 미친다고 보고하였으며, Sun *et al.*(2022)은 강한 바람과 높은 파도가 선박 저항을 증가시켜 연료 소비를 크게 늘린다는 점을 실증하였다. Ren *et al.*(2022)은 환경 변수를 포함할 때 연료 소모량 예측 정확도가 향상됨을 보였고, Wang *et al.*(2023)은 특히 태풍이나 해일과 같은 극한 기상 이벤트의 고려가 예측 성능 향상에 중요하다고 강조하였다. 한편, Gkerekos *et al.*(2019)은 동일한 선박이라도 운항 해역에 따라 연료 소모량이 달라질 수 있음을 보였다. 또한, Venkatesan *et al.*(2025)과 Xiao *et al.*(2023)은 풍향과 해류 방향과 같은 상대적인 변수들이 연료 소비에 미치는 복합적인 영향을 분석하였다. 이에 따라 단순한 변수 선택을 넘어 도메인 지식을 반영한 파생 변수의 중요성이 제기되었다. Zhou *et al.*(2023)과 Kim *et al.*(2023)

은 SHapley Additive exPlanations (SHAP) 분석을 통해 상대 풍속과 유속과 같은 파생 변수가 예측 결과에 미치는 영향력을 정량적으로 보였다.

이처럼 외부 환경 변수는 단순히 입력 변수로 사용하기보다는 선박의 운항 조건과 변수의 특성을 반영한 전처리 및 변수 가공이 선행되어야 한다. 특히 풍속, 유속, 파고 등은 방향성이나 상대적 영향력을 고려한 가공이 수반되어야 연료 소모에 미치는 효과를 정확히 반영할 수 있다. 따라서 본 연구에서는 도메인 지식을 반영한 외부 환경 기반 파생 변수를 생성하고, 주요 변수를 선별한 후 모델의 입력 변수로 활용하여 연료 소모량 예측 정확도를 제고한다.

### 3. 제안 방법론

본 연구에서 제안하는 방법론은 <Figure 1>과 같다. 먼저, 다양한 데이터를 수집 및 통합한 후 항해 및 정박 상태를 구분하여 항적을 분리한다. 이후 정합성이 검증된 항해 상태의 항적에 대해 결측치 처리 및 변수 가공을 포함한 데이터 전처리를 수행한다. 마지막으로, 시계열 예측 모델을 적용하여 항적별 선박 연료 소모량을 예측한다.

#### 3.1 데이터 수집

본 연구에서는 연료 소모량 예측을 위해 선박 운항 중 실시

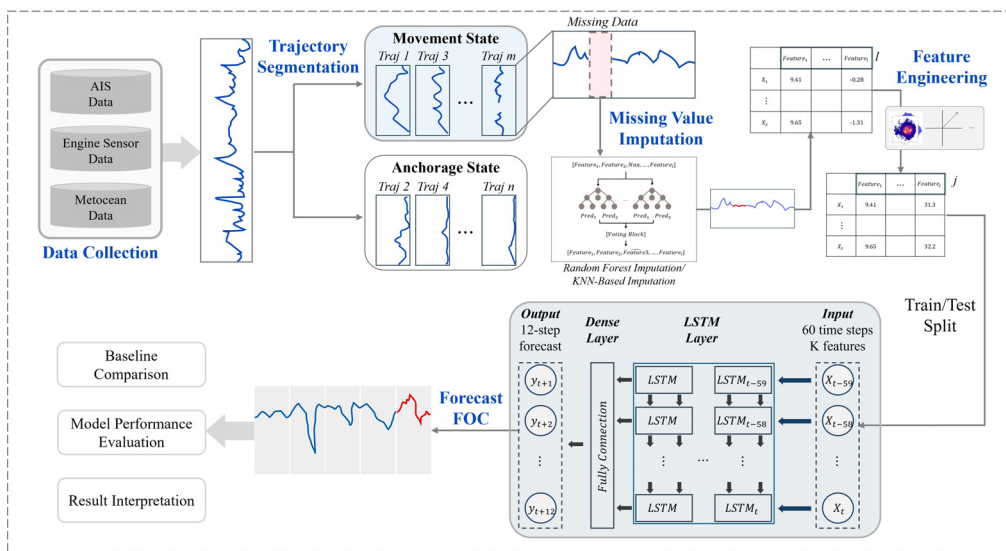


Figure 1. Overall Framework of the Proposed Method

Table 2. Description of the Dataset

No.	Type	Name	Description	Unit
1	AIS data	SOG	Speed over ground	knots
2	AIS data	COG	Direction of which the vessel is moving	degree
3	AIS data	LAT	Latitude	degree
4	AIS data	LON	Longitude	degree
5	AIS data	SPEED_VG	Speed through water	knots
6	AIS data	SPEED_LW	Speed relative to long wave or water flow	knots
7	AIS data	SHIP_HEADING	Direction the vessel's bow is pointing	degree
8	AIS data	DRAFT_FORE	Vertical distance between waterline and bottom of the hull measured at the perpendicular of the bow	meter
9	AIS data	DRAFT_AFT	Vertical distance between waterline and bottom of the hull measured at the perpendicular of the stem	meter
10	AIS data	RUDDER_ANGLE	Angle of rudder	degree
11	Engine Sensor data	ME1_FOC	No 1. main engine fuel oil consumption	kg/h
12	Metocean data	SEA_SURFACE_SALINITY	Salt concentration at the ocean surface	PSU (Practical Salinity Unit)
13	Metocean data	SEA_SURFACE_TEMP	Temperature of the ocean surface	°C
14	Metocean data	WAVE_PERIOD	Time between successive wave crests	sec
15	Metocean data	WAVE_DIRECTION	Direction from which the wave is coming	degree
16	Metocean data	WAVE_HEIGHT	Vertical distance from wave trough to crest	meter

간으로 수집된 AIS 데이터, 엔진 센서 데이터, 해양 기상 데이터(Metoccean data)의 총 세 가지 데이터를 활용하였다. 데이터는 컨테이너선 A를 대상으로 2021년 1월 1일부터 2023년 6월 30일까지 10분 주기로 수집되었으며, 결측값은 -9,999로 표기되어 있다. 통합된 데이터셋의 총 100,121개의 행과 28개 중 16개 변수를 분석에 사용하였으며, 사용한 변수는 <Table 2>와 같다.

### 3.2 데이터 전처리

#### (1) 항해 및 정박 상태 구분을 통한 항적 분리

연료 소모량은 항해 상태의 선박 속도를 기반으로 산출되므로, 연료 소모량 예측 모델 학습을 위해서는 실제 운항 데이터를 기반으로 선박의 항해 상태를 정확하게 식별하는 것이 중요하다. 본 연구는 이를 위해 실시간 센서 데이터를 활용하여 시공간적 임계값 기준에 따른 항적 분리 방식을 제안한다. 해당 방식은 연속된 시점 간의 Haversine 거리(Maria *et al.*, 2020)를 계산하고, 거리가 사전에 설정한 거리 임계값 이하일 경우 정지 상태로 판단한다. 이때 정지 상태가 연속적으로 시간 임계값을 초과하여 지속되면 정박 상태(Anchorage State)로 분류된다. 반대로, 일정 거리 이상을 연속적으로 이동한 구간은 항해 상태(Movement State)로 구분된다. 절차는 Algorithm 1과 같이 상태 전이가 발생할 때마다 새로운 항적 구간을 생성하는 방식으로 진행된다. 이는 단순한 속도 또는 거리 기반의 상태 구분에서 발생할 수 있는 일시적인 정지나 센서 오류로 인한 오분류 문제를 방지할 수 있다. 또한, 분석 목적에 따라 임계값을 조정하여 유연한 상태 구분이 가능하여 다양한 운항 조건에서의 적용 가능성이 높다는 이점을 가진다. 특히 정박 상태는 통신 단절이나 센서 결함 등으로 인해 데이터 품질이 저하되는 경우가 많기 때문에 해당 구간을 학습 대상에서 명시적으로 제외함으로써 시계열 데이터의 구조를 왜곡하는 불필요한 노이즈를 제거하고, 연료 소모 패턴의 학습에 집중할 수 있도록 한다. 더불어 정박 구간에서 비정상적으로 높은 연료 소모가 관측되는 경우가 다수 존재한다는 점에서, 해당 구간을 별도로 분리함으로써 이상치를 효과적으로 식별하고 제거하는 데이터 정제 과정에서도 활용 가능하다.

Algorithm 1을 적용한 선박 운항 상태 구분 방식의 타당성은 연료 소모량의 분포 및 항적별 연료 소모 패턴 분석을 통해 검증할 수 있다. <Figure 2>는 정박 상태(빨간색)와 항해 상태(파란색)로 구분된 각 구간의 연료 소모량 분포를 비교한 것으로, 정박 상태에서는 대부분의 값이 0 또는 0에 근접하지만 항해 상태에서는 1,000 ~ 3,000kg/h 범위의 높은 구간에 분포가 밀집되어 있어 뚜렷한 차이를 보인다. <Figure 3>에는 Trajectory 4-5 및 Trajectory 51-52에 대한 운항 경로와 해당 구간의 연료 소모 분포를 제시하였다. <Figure 2>와 유사하게 정박 구간에서는 연료 소모량이 대부분 0에 수렴하거나 일정하게 유지되는 반면, 항해 구간에서는 일정 수준의 연료 소모가 지속적으로 발생하는 경향을 띤다.

Algorithm 1: Trajectory Segmentation Based on Operation Status

---

Input: Sensor data  $D = \sum_{i=1}^T (t_i, \phi_i, \lambda_i)$ , Parameters:  $\psi, \epsilon$

Output: Labeled Sensor data  $D$  with  $s \in \{s_a, s_m\}$  and  $\tau$

- 1: Initialize  $\tau \leftarrow -1, start \leftarrow -1, t_{cont} \leftarrow 0$
- 2: For  $i=2$  to  $T$  do
- 3:  $\Delta\phi \leftarrow \phi_i - \phi_{i-1}, \Delta\lambda \leftarrow \lambda_i - \lambda_{i-1}$
- 4:  $a \leftarrow \sin^2(\frac{\Delta\phi}{2}) + \cos(\phi_{i-1}) \cdot \cos(\phi_i) \cdot \sin^2(\frac{\Delta\lambda}{2})$   
// Compute haversine distance
- 5:  $d_{i,i-1} \leftarrow 2R \cdot \arcsin(\sqrt{a})$  // Compute time difference
- 6:  $\Delta t \leftarrow t_i - t_{i-1}$
- 7: If  $d_i < \psi$  then
- 8:     If  $t_{cont} = 0$  then  $start \leftarrow i - 1$
- 9:     End if  $t_{cont} \leftarrow t_{cont} + \Delta t$
- 10: Else
- 11:     If  $t_{cont} \geq \epsilon$  then  $s \leftarrow s_a$
- 12:     Else  $s \leftarrow s_m$
- 13:     End if Label  $D_{start:i}$  with  $s, \tau$
- 14:     If  $s = s_m$  then
- 15:          $\tau \leftarrow \tau + 1$
- 16:     End if  $t_{cont} \leftarrow 0$
- 17:     End if
- 18: End for Label  $D_i$  with  $s_m, \tau$

---

Table 3. Summary of Notations for Algorithm 1

Symbol	Description	Symbol	Description
$i$	Index of dataset sequence	$T$	Total length of dataset
$t_i$	Timestamp at time step $i$	$\phi_i$	Latitude at time step $i$
$\lambda_i$	Longitude at time step $i$	$d_{i,j}$	Distance between $i$ and $j$
$a$	Intermediate value in haversine formula	$R$	Radius of earth
$\psi$	Distance threshold	$\epsilon$	Time threshold
$s_m$	Movement state	$s_a$	Anchorage state
$s$	Operation state at time step $i$	$\tau$	Trajectory ID
$t_{cont}$	Accumulated idle time	$start$	Start index of current segment

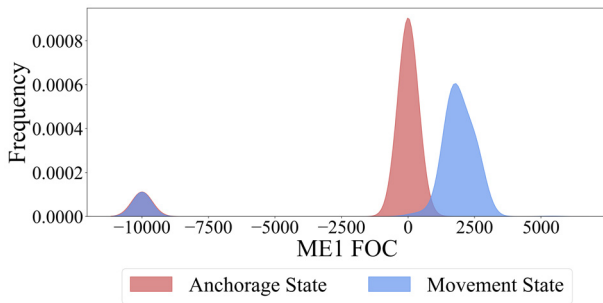


Figure 2. Distribution of Fuel Consumption by Operational Status

이후 <Figure 4>와 같이 정오 보고서 기반으로 분리된 항적 (파란색, (b))과 제안 방식으로 분리한 항적(빨간색, (a)+(b))을 비교하여 제안 방식의 정합성을 검증하였다. 두 항적을 시간 기준으로 매핑한 결과, 기존 정오 보고서에서는 항해 상태로 분류되지 않았던 구간이 제안 방식에서는 (a)와 같이 추가적으

로 항해 구간으로 식별되었고 연료 소모량 분포는 실제 항해 상태인 (b)와 유사한 경향을 보였다. 이는 제안된 방식이 정오 보고서에 비해 항해 구간을 정확하게 식별하여 항적을 구분할 수 있음을 보인다.

(2) 결측치 보간

센서 데이터는 선박의 실시간 운항 정보를 제공하지만 센서 오류, 통신 장애 등으로 인한 결측값이 빈번하게 발생한다. 특히 랜덤하게 발생하는 결측뿐만 아니라 연속적인 결측 구간의 발생 빈도도 높다. 이러한 데이터의 결측은 연속성이 중요한 시계열 데이터의 패턴을 왜곡할 가능성이 높으므로 적절한 보간 처리가 필수적이다. <Figure 5>는 센서 데이터에서 나타나는 다양한 결측 패턴에 대한 예시이며, 결측 유형에 따라 서로 다른 보간 방식이 필요함을 보인다. 이에 본 연구는 결측 구간의 특성에 따라 상이한 보간 방식을 적용하였다. 먼저 랜덤하게

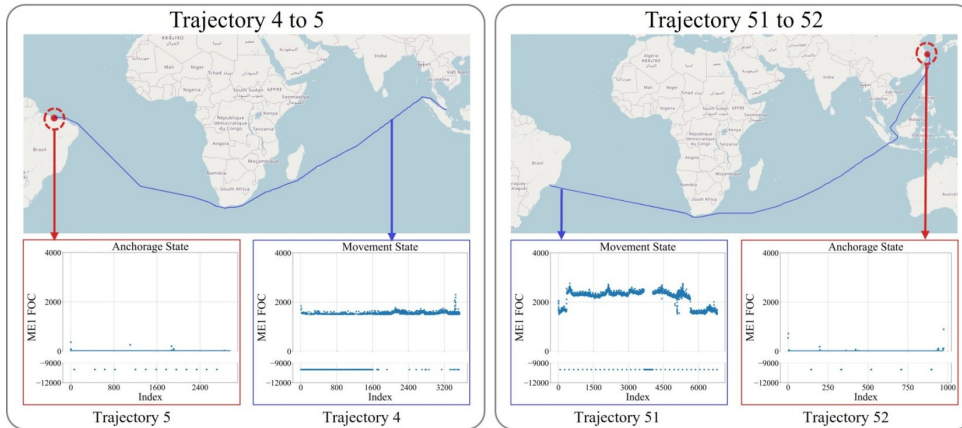


Figure 3. Fuel Consumption Comparison by each Trajectory Segmentation

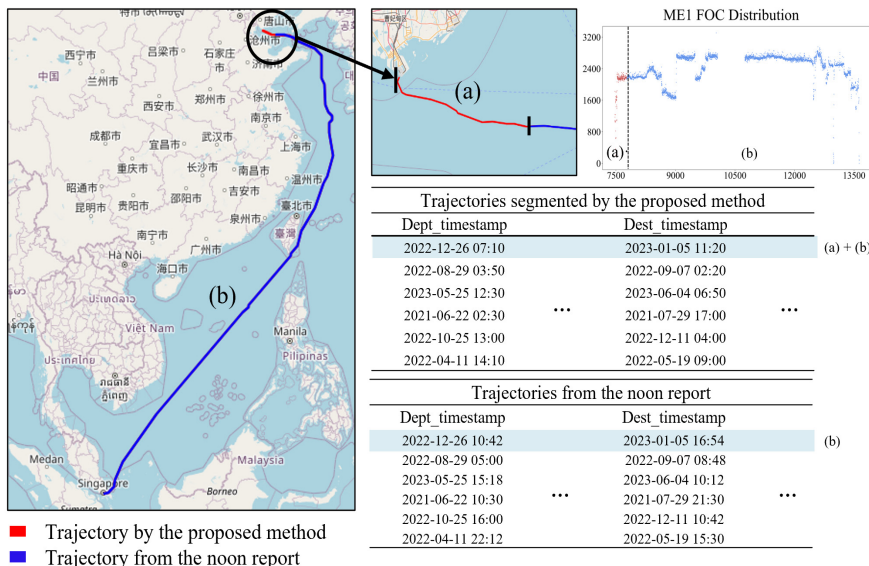


Figure 4. Comparison with Trajectories Segmented based on Noon Report

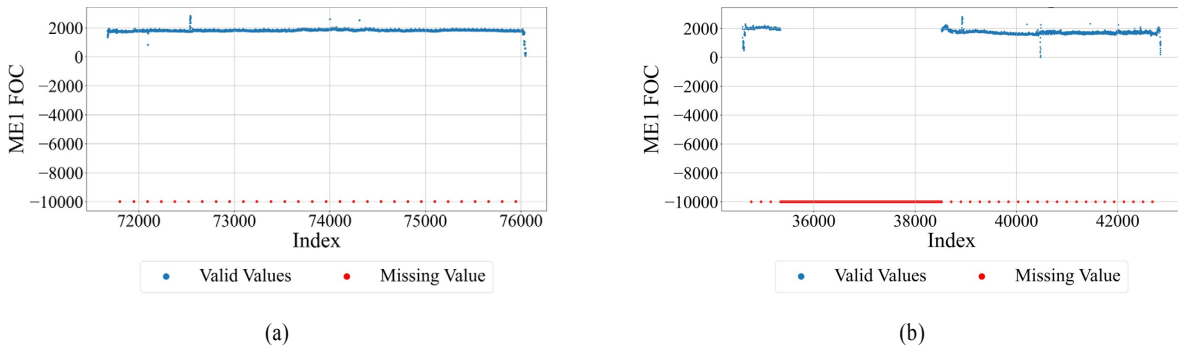


Figure 5. Different Missing Value Patterns (a) Random (b) Consecutive

발생한 결측의 경우 인접 데이터로부터 실제 패턴과 유사하게 보간이 가능하므로 곡선 보간으로 처리한다. 반면, 연속 결측 구간의 경우 단순 보간은 데이터의 시간적 연속성과 패턴을 왜곡할 가능성이 높기 때문에 Random Forest Imputation 및 유사 항적을 활용한 K-Nearest Neighbors (KNN) 기반 보간을 수행한다.

Table 4. Hyper-parameters of Random Forest

Parameter	Value
Number of Estimators	100
Maximum Depth	None
Minimum Samples Split	2
Minimum Samples Leaf	1
Random State	42

Random Forest Imputation은 결측이 있는 변수와 다른 변수 간의 관계를 비선형적으로 학습하여 결측값을 예측하는 방식으로, 변수 간의 상관성을 반영할 수 있는 이점을 갖는다(Tang et al., 2017). Random Forest 모델은 결측이 없는 구간에서 선정된 주요 입력 변수로 학습되었으며, 이를 활용하여 결측값을 보간하였다. 본 연구에서는 결측이 발생한 변수와 높은 상관성을 가지는 입력 변수를 선정하기 위하여 해당 변수만 무작위로 섞고 상관성이 높은 변수는 원래 값을 유지하여 각 변수의 독립적인 기여도를 평가하는 Conditional Permutation

Importance(CPI)(Debeer and Strobl, 2020)을 적용하였다. Random Forest 모델의 주요 파라미터는 <Table 4>에 제시된 값을 활용하였다. 이를 통해 결측이 발생한 변수와 강한 상관성을 가져 보간 성능에 실질적으로 기여하는 변수를 왜곡 없이 식별하여 모델의 입력 변수로 활용하였다.

다만, 주요 입력 변수 또한 동일 시점에 결측인 경우에는 모델 성능 저하로 인해 보간에 한계가 존재하므로 이를 보완하기 위하여 KNN 기반 보간을 추가로 적용하였다. KNN 보간은 특정 시점에서 결측이 발생했을 때 유사한 특성을 가진 과거 시점의 데이터를 기준으로 결측값을 추정하는 방식이다(Zhang et al., 2016). 본 연구에서는 KNN 보간을 단순히 전체 데이터에 적용하는 대신 유사한 운항 경로를 가지는 항적 그룹 내로 제한하여 적용하였다. 이는 활용한 데이터가 정기 항로를 반복 운항하는 컨테이너선의 센서 데이터로, 특정 지리적 위치에서 유사한 운항 패턴이 반복된다는 특성을 반영한 것이다. 그러나 유사한 구간이라도 실제 운항 중에는 선박의 적재 상태, 해양 기상, 계절 등 외부 요인에 따라 운항 패턴이 달라질 수 있다. 따라서 위도, 경도, 파고, 파랑 속도, 월, 홀수 등 외부 조건에 해당하는 변수들을 정규화하고 동일한 가중치를 부여한 뒤 해당 변수들을 기준으로 유사 항적 내에서 가장 유사한 K개의 시점을 탐색하여 결측값을 보간하였다. 본 연구에서는 K={3, 4, 5, 6}으로 실험 후 최종적으로 K를 4로 설정하였다. 이는 전체 항적을 대상으로 하는 기존 방식과 달리 운항 특성과 환경 조건이 유사한 데이터 내에서만 비교가 이루어지므로 실제 운항 환경과 보다 밀접한 값을 기반으로 결측값을 추정할 수 있다.

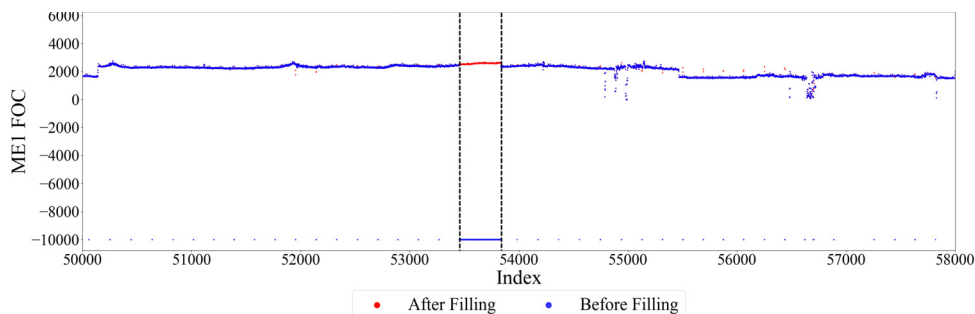


Figure 6. Imputation Results using the KNN-based Method

**Table 5.** Performance Comparison of Interpolation Method

Interpolation method	MSE(e+04)	MAE
Spline Interpolation	2.048	383.319
Linear Interpolation	5.367	97.498
Forward Fill	6.319	116.307
Mean Imputation	2.451	427.446
<b>Ours</b>	<b>2.012</b>	<b>90.392</b>

<Figure 6>은 제안한 KNN 기반 보간을 수행한 결과이며, 데이터의 패턴이 보존되는 형태로 보간이 이루어짐을 확인할 수 있다. <Table 5>에서는 무작위 결측을 발생시킨 후 교차 검증을 통해 제안한 세 단계를 거친 보간 방식과 기존의 단순 보간법을 비교하였다. 그 결과 제안 방식이 Mean Squared Error (MSE)와 Mean Absolute Error (MAE) 측면에서 가장 우수한 성능을 나타냈다. 특히 MAE를 기준으로 해석할 경우, 스케일 변화에 대한 민감도가 낮아 안정적인 보간 결과를 보인다.

### 3.3 변수 가공

기존 변수 정보를 보다 효과적으로 활용하기 위해 본 연구는 관련 연구 및 도메인 정보를 기반으로 기상 변수의 상대값 변환, 방향성 변수의 벡터화, 홀수 관련 파생 변수를 생성하였으며, 수식에 사용된 기호는 <Table 6>과 같다.

**Table 6.** Summary of Notations for Derived Variables

Symbol	Description
$V_w$	Wind speed
$V_s$	Ship speed (Speed over Ground, SOG)
$U$	East-west component of directional variable
$V$	North-south component of directional variable
$D_w$	Wind direction
$H_s$	Ship heading
$\theta$	Angular difference between wind direction and ship heading ( $D_w - H_s$ )

먼저 풍속, 조류 속도, 파고 등의 기상 데이터를 선박 기준 상대값으로 변환하여 항해 방향에 따른 연료 소모량의 영향을 고려한다. 이는 선박 진행 방향과 외부 기상 요소의 방향성 차이를 반영하여 생성된다. 예를 들어, 상대 풍속(Relative Wind Speed,  $V_{rw}$ )은 식 (1)과 같다.

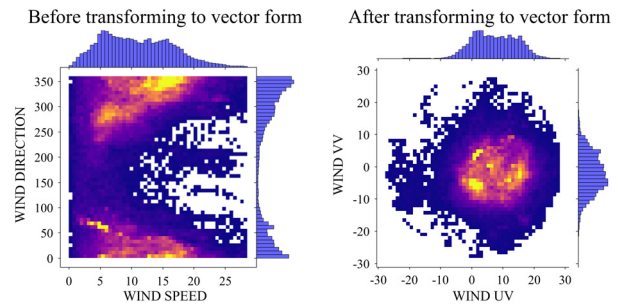
$$V_{rw} = \sqrt{(V_w \cos \theta - V_s)^2 + (V_w \sin \theta)^2} \quad (1)$$

또한, 풍향, 조류 방향, 파향과 같은 방향성 변수는  $0^\circ$ 와  $360^\circ$ 가 동일한 방향을 의미하는 원형 특성을 가지므로 이를 U-V 벡터 성분으로 변환하여 방향성과 크기를 동시에 반영하였다.

풍속과 풍향으로부터 변환된 벡터는 수식 (2), (3)과 같이 산출되며 <Figure 7>은 변환 전후의 데이터 분포를 나타낸다. 벡터 변환 전에는 풍속과 풍향이 독립 변수로 표현되어,  $0^\circ$ 와  $360^\circ$  사이 불연속 구간이 존재하고 데이터가 한쪽으로 치우쳐 분포한다. 반면 변환 후에는 V-벡터(WIND VV), U-벡터(WIND UV)가 방향성과 크기가 함께 반영된 벡터 형태로, 모델이 방향성 정보를 정확하게 반영할 수 있도록 한다.

$$U = V_w \cdot \cos(D_w) \quad (2)$$

$$V = V_w \cdot \sin(D_w) \quad (3)$$

**Figure 7.** Data Distribution Before and After Transformation of Directional Vector Variables

마지막으로, 선박의 적재 상태를 반영하기 위해 홀수 관련 파생 변수인 Draft Mean과 Draft Trim을 생성한다. 홀수는 선박의 중량에 따라 변하며 수면 저항 및 연료 소모에 직접적인 영향을 미친다(Shin *et al.*, 2021). Draft Mean은 선박의 평균 홀수를 통해 적재 중량 증가에 따른 수면 저항 변화를 반영하며 Draft Trim은 전후 홀수 차이를 통해 트림 상태에 따른 추진 효율 차이를 고려한다. 수식은 (4), (5)과 같다.

$$DraftMean = \frac{DraftFore + DraftAft}{2} \quad (4)$$

$$DraftTrim = DraftFore - DraftAft \quad (5)$$

모델에 활용할 입력 변수는 Recursive Feature Elimination (RFE) 방식(Liu *et al.*, 2021)을 활용하여 선정하였다. RFE는 특정 예측 모델을 기반으로 각 반복 단계마다 중요도가 낮은 변수를 제거하며 예측 성능이 가장 우수한 변수 조합을 선택한다. 이는 과적합을 방지하고 핵심 입력 변수 조합들을 효과적으로 도출하는 데 유용하다. 변수 가공 이후 RFE를 활용하여 최종적으로 선정된 변수는 Rel Wave Direction, Rel Wave Period, Wind UV, Ship Heading, Speed VG, Lat, Wind VV, Draft Meanf, Rel Wind Speed, Lon, Draft Trim로 총 24개의 변수 중 11개이다. 이때, LSTM 모델을 활용하여 각 변수의 RFE 점수를 산출했으며 교차 검증 및 통계적 검정(Wilcoxon Signed-Rank Test)을 통해 임계값인 0.7 이상의 변수들을 최종 입력 변수로 사용하였다.

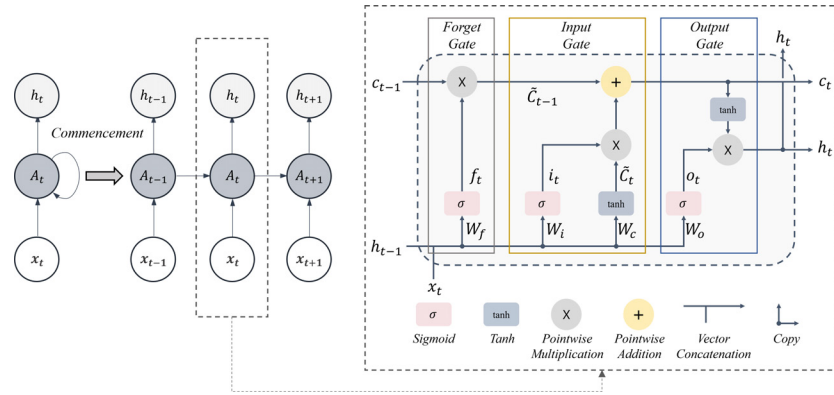


Figure 8. Structure of LSTM Cell

### 3.4 모델 학습 및 평가

본 연구에서는 선박 센서 데이터의 시간적 연속성을 보존하며 선박 연료 소모량을 예측하기 위하여 LSTM을 예측 모델로 활용하였다.

LSTM은 순환 신경망의 한 종류로 장기 의존성을 효과적으로 학습하기 위하여 <Figure 8>과 같은 구조를 갖는다. 내부에는 입력 게이트(Input Gate), 망각 게이트(Forget Gate), 출력 게이트(Output Gate)가 포함되어 있어 시점별로 어떤 정보를 기억하고 잊을지를 선택적으로 조절한다(Staudemeyer and Morris, 2019). 각 시점의 입력 데이터가 셀 상태(Cell State)를 따라 흐르며 과거 정보가 누적되는 구조는 시간 흐름에 따라 변화하는 연료 소모 패턴을 학습하는 데 적합하므로 LSTM을 예측 모델로 활용하였다. 본 연구에서 활용한 LSTM 모델의 하이퍼파라미터는 <Table 7>과 같다.

Table 7. Hyper-parameters of LSTM

Parameter	Value
Hidden Dimension	64
Number of Layers	3
Dropout	0.1
Learning Rate	0.001
Batch Size	32

## 4. 실험

### 4.1 실험 설계

본 연구는 기존 선행 연구에서 적용된 주요 방법론과의 비교 실험을 통해 항적 구분 방식 및 데이터의 시계열 특성과 외부 변수의 적절한 활용 여부가 예측 성능에 미치는 영향을 검증한다. 실험은 총 두 가지로 구성된다.

첫째, 실험 1에서는 입력 변수의 가공 없이 시계열 특성의 활용이 예측 성능에 기여하는 바를 검증한다. 이를 위해 비교

모델로는 머신러닝 기반 다변량 예측 모델(XGBoost, Random Forest)과 통계 기반의 전통적 시계열 예측 모델(AutoRegressive Intergrated Moving Average with eXogenous regressors(ARIMAX), Vector AutoRegression(VAR)), 딥러닝 기반 시계열 예측 모델(Gated Recurrent Unit(GRU), Temporal Convolutional Network (TCN), Transformer)을 적용한다. 특히, 동일한 모델 구조 하에서 항적 분리 방식의 차이가 예측 결과에 미치는 영향을 함께 분석한다. 항적 구성은 (1) 정오 보고서 기반(NR-NF), (2) 본 연구에서 제안한 운항 상태 기반(PM-NF)과 같다. 이를 통해 예측에서 항해 상태를 정확하게 반영한 항적 활용의 필요성을 검토한다.

둘째, 실험 2는 변수 가공의 필요성을 검증하기 위해 제안 방식으로 분리한 항적 및 실험 1과 동일한 모델 구성을 적용하되, 과생 변수 생성 및 RFE 기반 변수 선택이 선행된 입력 변수 조합을 사용한 예측을 수행한다. 해당 실험의 결과는 PM-FE으로 표기한다.

모든 실험은 학습 및 검증 데이터를 80:20 비율로 분할하여 수행하였으며 항적별로 현재 시점을 기준으로 남은 이후 구간에 대한 연료 소모량 예측 정확도를 평가한다. 시계열 예측 모델은 이전 60시점의 데이터를 입력으로 사용하고, 이후 12시점의 연료 소모량을 출력으로 설정하였다.

### 4.2 실험 평가 지표

실험 평가 지표는 MAE, RMSE, SMAPE를 활용하였다( $n$ : Total number of observations,  $i$ : Index of the observation ( $1 \leq i \leq n$ ),  $y_i$ : Actual value at time  $i$ ,  $\hat{y}_i$ : Predicted value at time  $i$ ).

MAE는 예측값과 실제값 간의 절대 오차 평균을 계산하며 절대 단위로 측정되는 오차를 직관적으로 해석할 수 있는 장점이 있다(Chicco *et al.*, 2021).

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (6)$$

Root Mean Squared Error(RMSE)는 오차 제곱의 평균에 제곱근을 취한 값으로, 큰 오차에 더 민감하게 반응하며 예측의 안정성을 확인하는 데 유용하다(Chicco *et al.*, 2021).

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (7)$$

Symmetric Mean Absolute Percentage Error (SMAPE)는 예측값과 실제값의 상대적 차이를 백분율로 나타내며 MAPE 대비 실제값이 0에 가까운 경우에도 안정적인 결과를 산출한다(Chicco *et al.*, 2021).

$$SMAPE = \frac{100}{n} \sum_{i=1}^n \frac{|y_i - \hat{y}_i|}{(|y_i| + |\hat{y}_i|)/2} \quad (8)$$

### 4.3 실험 결과

본 연구에서는 앞서 서술한 두 가지 실험 구성의 결과를 MAE, RMSE, SMAPE 세 가지 지표 기준으로 비교 및 분석하였다.

<Table 8>은 변수 가공 없이 수행된 예측 실험 결과로 항적 분리 방식 및 모델 구조에 따른 예측 결과를 비교한 것이다. 실험 결과, 트리 기반 머신러닝 모델(XGBoost, RandomForest)은 전체 연료 소모량 추세는 일정 부분 포착하나 시점 간 급격한 연료 소모 변화 구간에서는 오차가 크게 증가하는 경향을 보였다. 특히 RMSE 수치가 400 수준에 달한 것은 모델이 극단적인 오차 구간의 영향을 크게 받았음을 나타낸다. 이는 연료 소모량이 일부 구간에서 짧은 시간 내에 급증 및 급감이 반복되는 고변동 특성을 가지는데, 시계열 흐름을 반영하지 않고 각 시점을 독립적으로 예측하여 급변 구간을 고려하는 데에 한계가 존재한 것으로 해석된다.

또한, 전통적인 통계 기반 예측 기법(ARIMAX, VAR)은 학습 데이터에 내재된 복잡한 비선형성 패턴을 학습하는 데에 한계를 보인 반면, 비선형적 시계열 패턴을 효과적으로 반영할 수 있는 딥러닝 기반 시계열 모델은 단기적으로 변동성이 심한 구간에서도 상대적으로 높은 예측 성능을 보였다. 특히 LSTM은 MAE 기준 가장 오차가 낮으며 단기 고변동 구간에서도 안정적이다.

또한, 항적 분리 방식에 따라 모든 모델에서 일정 수준 이상의 예측 성능 차이가 나타났다. 정오 보고서 기반의 항적 분리

**Table 8.** Comparison Result of Experiment 1 (Performance Comparison by Trajectory Segmentation (NR-NF vs PM-NF) and Model Type)

Experimental Setting	Model Type		Metric		
			RMSE	MAE	SMAPE
NR-NF	Non Time-series model	XGBoost	472.01	251.02	13.21
		RandomForest	396.32	247.51	12.83
	Statistical Time-series model	ARIMAX	685.51	411.22	30.59
		VAR	652.18	406.14	29.98
	Deep Learning Time-series model	LSTM	203.47	81.59	5.16
		GRU	237.44	92.57	5.42
		TCN	212.52	88.72	5.27
	Transformer	216.60	85.84	5.12	
PM-NF	Non Time-series model	XGBoost	454.31 (↑ 3.76%)	240.14 (↑ 4.36%)	12.64 (↑ 4.32%)
		RandomForest	378.61 (↑ 4.48%)	236.34 (↑ 4.51%)	12.58 (↑ 1.95%)
	Statistical Time-series model	ARIMAX	651.25 (↑ 4.99%)	398.26 (↑ 3.15%)	29.02 (↑ 5.13%)
		VAR	639.21 (↑ 2.09%)	390.57 (↑ 3.83%)	29.29 (↑ 2.31%)
	Deep Learning Time-series model	<b>LSTM</b>	<b>186.84</b> <b>(↑ 8.17%)</b>	<b>72.14</b> <b>(↑ 11.58%)</b>	<b>4.59</b> <b>(↑ 11.03%)</b>
		GRU	220.33 (↑ 7.21%)	83.74 (↑ 9.54%)	5.09 (↑ 6.09%)
		TCN	193.15 (↑ 10.33%)	79.41 (↑ 10.47%)	4.91 (↑ 7.32%)
		Transformer	202.13 (↑ 7.09%)	77.42 (↑ 9.82%)	4.66 (↑ 9.02%)

Table 9. Comparison Result of Experiment 2 (Effect of Feature Engineering)

Experimental Setting	Model		Metric		
			RMSE	MAE	SMAPE
PM-FE	Non Time-series model	XGBoost	266.48 (↑41.36%)	138.07 (↑42.52%)	8.28 (↑34.50%)
		RandomForest	292.93 (↑22.64%)	162.58 (↑31.22%)	8.85 (↑29.66%)
	Statistical Time-series model	ARIMAX	640.62 (↑1.64%)	387.73 (↑2.64%)	29.81 (↑2.72%)
		VAR	521.58 (↑18.39%)	283.08 (↑27.52%)	21.04 (↑28.17%)
	Deep Learning Time-series model	<b>LSTM (Ours)</b>	<b>151.42</b> <b>(↑40.32%)</b>	<b>55.09</b> <b>(↑23.65%)</b>	<b>3.61</b> <b>(↑21.35%)</b>
		GRU	171.85 (↑18.94%)	62.72 (↑25.09%)	4.01 (↑21.41%)
		TCN	164.19 (↑15.01%)	62.35 (↑21.50%)	3.98 (↑18.92%)
		Transformer	168.73 (↑16.52%)	64.16 (↑17.10%)	4.12 (↑11.59%)

(NR-NF)에 비해 제안한 운항 상태 기반 분리(PM-NF)를 활용한 경우 전반적으로 높은 성능을 보였으며, 이러한 경향은 시계열 예측 모델에서 상대적으로 크게 나타났다. 개선 폭은 전체적으로 크지 않으나 제안 방식으로 분리된 항적을 활용할 경우 기존보다 더 많은 학습 데이터를 제공하며 해당 구간에서의 항해 패턴이 일정 부분 모델 학습에 반영되어 성능 개선에 기여할 수 있음을 보인다.

이러한 결과는 단순히 모델의 구조적 차이를 넘어 시간에 따른 누적 변화 및 반복 패턴과 같은 데이터의 시계열 특성 활용, 변수 간의 비선형적 상관관계 고려, 실제 운항 상태 기반의 항적 활용이 선박 연료 소모량과 같은 고변동성 데이터를 다루는 데에 효과적임을 보인다.

<Table 9>는 외부 해양 기상 정보 및 운항 조건 등 도메인 지식을 바탕으로 파생 변수를 생성한 후 변수 선택을 수행한 실험 결과를 제시한다. 본 실험(Experiment 2)의 목적은 동일한 모델 구조를 유지한 상태에서 입력 변수 구성 변화가 예측 성능에 미치는 영향을 검증하는 데에 있다.

실험 결과, 모든 모델에서 세 지표가 일관되게 개선되어 입력 변수의 적절한 조합이 모델 예측 정확도 향상에 기여함이 나타났다. 특히 시계열 특성이 고려되지 않은 머신러닝 모델에서도 성능 개선 폭이 컸다는 점은 입력 변수의 구성 변화만으로도 중심값과 비율 예측의 정확도를 크게 개선할 수 있음을 보인다. 반면, ARIMAX는 외생 변수와 시계열 구성 요소 간의 관계를 선형적으로 가정하기 때문에 변수 가공 이후에도 성능 개선이 미미하였다. 딥러닝 기반 시계열 모델 중에서는 LSTM이 세 지표 모두에서 가장 낮은 오차를 보였다. RMSE와 MAE가 동시에 개선된 것은 평균 수준의 오차뿐 아니라 급격히 감소하거나 증가하는 구간에 대해서도 보다 정확하게 예측

했음을 나타낸다. 또한 SMAPE가 유의미하게 감소하여 연료 소비량의 스케일이 크게 달라지는 상황에서도 상대 오차 기준으로 일관된 성능을 유지했다. 실제 연료 소모량은 운항 조건에 따라 100g/h 이하에서 수천 단위까지 변동하므로 변수 가공을 통해 스케일 편차를 줄이고 주요 영향력을 반영하여 예측에 중요한 변수만을 선별한 학습이 성능 향상의 주요 요인으로 작용하였다.

이처럼 변수 가공은 <Figure 9>에서 보여 지듯이 단순한 성능 개선을 넘어 모델 오차 분포 안정화, 이상값에 대한 민감도 완화 등 다양한 측면에서 예측 정확도를 제고하는 데에 기여하였다. 특히 LSTM은 시간 흐름에 따른 누적 변화와 변수 간 상호작용을 함께 학습함으로써 고변동성 데이터를 다루는 환경에서도 평균 오차와 상대 오차 모두에서 높은 성능을 나타냈다. 또한, TCN과 Transformer 모델은 단기 예측임에도 불구하고 장기 패턴 학습에 유리한 구조적 특성을 활용하여 일정 수준 이상의 예측 안정성을 보였다. 이는 본 연구에서 설정한 12시점 기준의 단기 예측을 넘어, 향후 장기 예측으로의 확장 가능성이 존재함을 나타낸다.

<Figure 10>은 MAE를 기준으로 모델 구조, 항적 분리 방식, 입력 변수 구성의 변화가 예측 성능에 미친 영향을 종합적으로 비교한 결과를 제시한다. 통계 기반 예측 기법 및 머신러닝 모델은 변수 가공을 적용했음에도 불구하고, 딥러닝 기반 시계열 모델에 비해 전반적으로 높은 MAE를 보였다. 딥러닝 기반 시계열 모델 중에서도 LSTM은 모든 실험 환경 하에서 가장 낮은 오차를 나타냈다. 이러한 결과는 연료 소모량과 같은 고변동성 데이터를 예측할 때 시계열적 특성과 운항 상태를 반영한 데이터 구성 및 변수 가공이 딥러닝 기반 모델과 결합되어 활용될 경우 예측 정확도와 안정성 향상에 기여함을 시사한다.

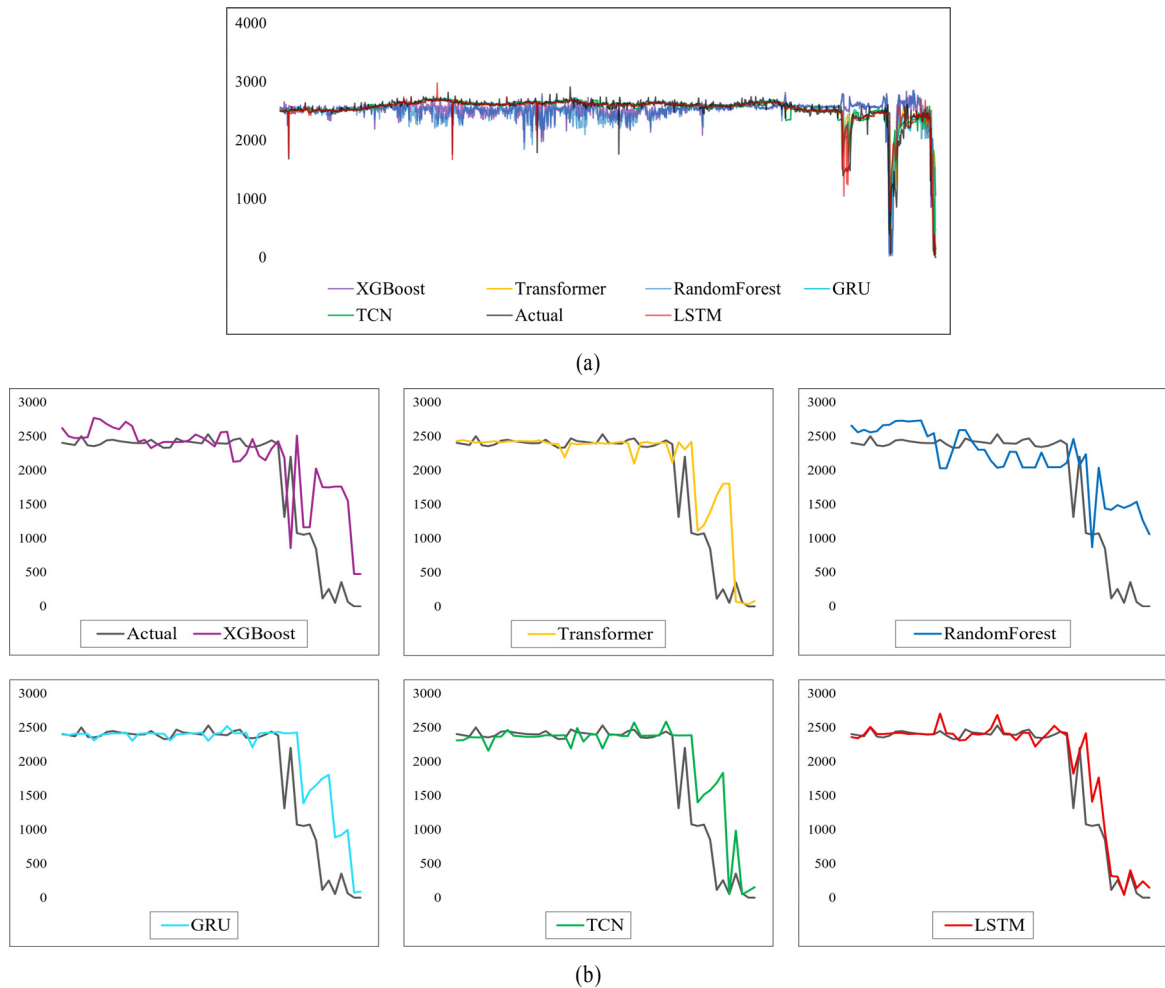


Figure 9. Experimental Results of Experiment 2. (a) Comparison Results between Ground Truth and Predicted Values (b) Last 48 Time Steps

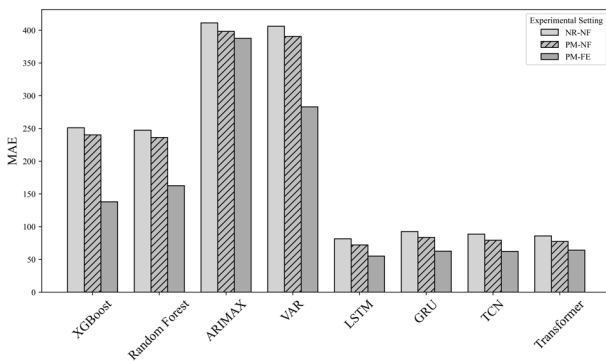


Figure 10. MAE Comparison Across Experimental Settings

### 5. 결론

본 연구에서는 선박 센서 데이터를 활용하여 항해 및 정박 상태를 구분하고, 이를 기반으로 항적을 구분하였다. 이후 항해 상태의 항적에 대해 결측값을 보간하고 외부 환경 변수를 반영한 항적별 연료 소모량 시계열 예측 프레임워크를 제안한다.

실험 결과, 딥러닝 기반 시계열 모델이 모든 실험 환경 및 지표에서 높은 성능을 보였다. 이는 센서 데이터의 시계열 특성과 다양한 변수들의 비선형 관계를 고려하는 것이 예측 성능 향상에 효과적임을 나타낸다. 또한, 운항 상태를 기반으로 분리된 항적을 활용한 경우 동일한 모델 구조에서도 전반적으로 예측 성능이 일부 개선되어 항적 구분의 정확성 또한 모델 성능에 영향을 미치는 것으로 나타났다. 한편, 외부 환경 변수 및 파생 변수를 포함하여 구성된 입력 변수 조합을 활용하여 시계열 예측을 수행하였을 때 가장 우수한 성능을 나타내어, 도메인 지식을 반영한 변수 가공이 예측 정확도 개선에 기여함을 확인하였다.

이러한 결과는 선사들이 최근 감축률이 상향되어 규제 대응 부담이 더욱 증가한 CII 기준을 충족하기 위한 선박 속도 산정, 최적 운항 계획 수립 등에 실질적으로 활용될 수 있다. 특히 본 연구의 제안 방식은 결측과 이상치가 빈번한 선박 센서 데이터를 시계열 데이터로 의미 있게 활용하여 연료 소모량 예측 성능을 높이는 프레임워크로서의 의의가 있다.

다만, 본 연구는 단기 예측에 초점을 두고 있으며 단일 선박 데이터에 기반한 모델 검증으로 장기 예측 및 모델 일반화에

는 한계가 있다. 따라서 향후 연구에서는 장기 예측을 위한 모델 고도화뿐만 아니라 이를 기반으로 한 항로 최적화에 초점을 두고 연구를 진행할 계획이다. 또한 다양한 선종 및 항로 데이터로 확장 적용하고, Cross-validation 등을 활용한 시계열 교차 검증을 수행하여 모델의 과적합을 방지하고 일반화 성능을 개선할 것이다.

## 참고문헌

- Barzi, F. (2004), Imputations of missing values in practice: Results from imputations of serum cholesterol in 28 cohort studies, *American Journal of Epidemiology*, **160**(1), 34-45.
- Bergström, M., Gosala, V., Depken, J., Fitz, A., Euskirchen, F., and Ehlers, S. (2023), A Simulation-Based Approach for Evaluating Merchant Fleet Decarbonization Strategies, *Proc. ASME Int. Conf. on Ocean, Offshore and Arctic Engineering*, 5.
- Cepowski, T. and Drozd, A. (2023), Measurement-based relationships between container ship operating parameters and fuel consumption, *Applied Energy*, **347**, 121315.
- Chen, Y., Sun, B., Xie, X., Li, X., Li, Y., and Zhao, Y. (2024), Short-term forecasting for ship fuel consumption based on deep learning, *Ocean Engineering*, **301**, 117398.
- Chicco, D., Warrens, M. J., and Jurman, G. (2021), The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation, *PeerJ Computer Science*, **7**, e623.
- Debeer, D. and Strobl, C. (2020), Conditional permutation importance revisited, *BMC Bioinformatics*, **21**.
- Dong, Z., Xia, C., Fang, K., and Zhang, W. (2022), Effect of the carbon emissions trading policy on the co-benefits of carbon emissions reduction and air pollution control, *Energy Policy*, **165**, 112998.
- Gkerekos, C., Lazakis, I., and Theotokatos, G. (2019), Machine learning models for predicting ship main engine Fuel Oil Consumption: A comparative study, *Ocean Engineering*, **188**, 106282.
- Grapsas, N. (2023), Contemporary assessment of International Maritime Organization's Carbon Intensity Indicator requirement on tanker vessels' chartering and employability prospects, *Proc. Global NEST Int. Conf. on Environmental Science & Technology*.
- Hu, Z., Zhou, T., Osman, M. T., Li, X., Jin, Y., and Zhen, R. (2021), A novel hybrid fuel consumption prediction model for ocean-going container ships based on sensor data, *Journal of Marine Science and Engineering*, **9**(4), 449.
- Karagiannidis, P. and Themelis, N. (2021), Data-driven modelling of ship propulsion and the effect of data pre-processing on the prediction of ship fuel consumption and speed loss, *Ocean Engineering*, **222**, 108616.
- Kim, H. J., Park, M. G., and Lee, J. H. (2023), A Study on the Prediction of Fuel Consumption of Bulk Ship Main Engine Using Explainable Artificial Intelligence, *Journal of Navigation and Port Research*, **47**(4), 182-190.
- Kim, M., Lee, J., An, S., and Hwang, D. (2024), Proposals on Effective Implementation of the Carbon Intensity Indication of Ships, *Journal of Marine Science and Engineering*, **12**(11), 1906.
- Lei, L., Wen, Z., and Peng, Z. (2021), Prediction of Main Engine Speed and Fuel Consumption of Inland Ships Based on Deep Learning, *Journal of Physics: Conference Series*, 2025.
- Liu, W. and Wang, J. (2021), Recursive elimination-selection algorithms for wrapper feature selection, *Applied Soft Computing*, **113**, 107956.
- Liu, Y., Wang, K., Lu, Y., Zhang, Y., Li, Z. R., and Huang, L. (2024), A Ship Energy Consumption Prediction Method Based on TGMA Model and Feature Selection, *Journal of Marine Science and Engineering*, **12**(7), 1098.
- Ma, Y., Zhao, Y., Yu, J., Zhou, J., and Kuang, H. (2023), An interpretable gray box model for ship fuel consumption prediction based on the SHAP framework, *Journal of Marine Science and Engineering*, **11**(5), 1059.
- Maria, E., Budiman, E., Haviluddin, H., and Taruk, M. (2020), Measure distance locating nearest public facilities using Haversine and Euclidean Methods, *Journal of Physics: Conference Series*, 1450.
- Melo, R., Figueiredo, N., Tobias, M., and Afonso, P. (2024), A Machine Learning Predictive Model for Ship Fuel Consumption, *Applied Sciences*.
- Perera, L. P. (2017), Handling big data in ship performance and navigation monitoring, *Smart Ship Technology*, 89-97.
- Rauca, L. and Batrinca, G. (2023), Impact of Carbon Intensity Indicator on the Vessels' Operation and Analysis of Onboard Operational Measures, *Sustainability*.
- Ren, F., Wang, S., Liu, Y., and Han, Y. (2022), Container ship carbon and fuel estimation in voyages utilizing meteorological data with data fusion and machine learning techniques, *Mathematical Problems in Engineering*, **2022**, 4773395.
- Rissman, J., Bataille, C., Masanet, E., Aden, N., Morrow, W. R. III, Zhou, N., Elliott, N., Dell, R., Heeren, N., Huckestein, B., Cresko, J., Miller, S. A., Roy, J., Fennell, P., Cremmins, B., Koch Blank, T., Hone, D., Williams, E. D., de la Rue du Can, S., Sisson, B., Williams, M., Howarth, N., Schmidt, T. S., Davis, S. J., Pales, A. F., Minx, J. C., Creutzig, F., Purohit, P., Mima, S., Miketa, A., Fekete, H., Levi, P., Prakash, S., and Helseth, J. (2020), Technologies and policies to decarbonize global industry: Review and assessment of mitigation drivers through 2070, *Applied Energy*, **266**, 114848.
- Shi, M., Yang, T., Liang, Y., and Wang, C. (2024), Ship fuel consumption prediction model based on LSTM, *Proc. Int. Conf. on Electronic Engineering and Informatics (EEI)*, 520-523.
- Shi, Q., Dai, W., Santerre, R., and Liu, N. (2020), A Modified Spatiotemporal Mixed-Effects Model for Interpolating Missing Values in Spatiotemporal Observation Data Series, *Mathematical Problems in Engineering*, **2020**, 1-15.
- Shin, D., Yang, C., and Jeon, H. (2021), Real Ship Experimental Analysis on the Influence of Fuel Oil Consumption Considering Ship Speed and Ocean Environment, *Marine Engineering*, **8**, 99-110.
- Sim, S., Park, J. H., and Bae, H. (2022), Deep collaborative learning model for port-air pollutants prediction using automatic identification system, *Transportation Research Part D: Transport and Environment*, **111**, 103431.
- Smith, T. W. P., Aldous, L., and Bucknall, R. (2013), Noon report data uncertainty, *Low Carbon Shipping Conference*.
- Staudemeyer, R. C. and Rothstein Morris, E. (2019), Understanding LSTM—A tutorial into long short-term memory recurrent neural networks, arXiv preprint, arXiv:1909.09586.
- Sun, W., Tang, S., Liu, X., Zhou, S., and Wei, J. (2022), An improved ship weather routing framework for CII reduction accounting for wind-assisted rotors, *Journal of Marine Science and Engineering*, **10**(12), 1979.
- Tang, F. and Ishwaran, H. (2017), Random forest missing data algorithms,

- Statistical Analysis and Data Mining: The ASA Data Science Journal*, **10**(6), 359-447.
- Trakakis, A., Daremas, N., Papatheodorou, M., Chatzinikolaou, S., Papadopoulos, C., and Kaiktsis, L. (2023), CII compliance: Effects of operational and technological measures on current levels and time development of calculated CII, Day 1, Tue, March 07, 2023.
- Venkatesan, I., Dharanish, M. B., Lakshmi, S. M., and Pavithra, S. (2025), Adaptive Weather Routing for Advanced Ship Navigation: A Comprehensive Review. In *2025 3rd International Conference on Intelligent Data Communication Technologies and Internet of Things (IDCIoT)*, IEEE, 1267-1273.
- Wang, H., Yan, R., Wang, S., and Zhen, L. (2023), Innovative approaches to addressing the tradeoff between interpretability and accuracy in ship fuel consumption prediction, *Transportation Research Part C: Emerging Technologies*, **157**, 104361.
- Xiao, Z. and Lam, J. S. L. (2023), Analysing exhaust emission of oil tanker vessels using big data in the port of Singapore, *International Journal of Shipping and Transport Logistics*, **16**(3-4).
- Yan, R., Wang, S., and Du, Y. (2020), Development of a two-stage ship fuel consumption prediction and reduction model for a dry bulk ship, *Transportation Research Part E: Logistics and Transportation Review*, **138**, 101930.
- Yang, H., Sun, Z., Han, P., and M. (2023), Data-driven prediction of ship fuel oil consumption based on machine learning models considering meteorological factors, *Journal of Engineering for the Maritime Environment*.
- Yuan, Z., Liu, J., Zhang, Q., Liu, Y., Yuan, Y., and Li, Z. (2021), Prediction and optimisation of fuel consumption for inland ships considering real-time status and environmental factors, *Ocean Engineering*, **221**, 108530.
- Zhang, Z. (2016), Introduction to machine learning: k-nearest neighbors, *Annals of Translational Medicine*, **4**(11), 218.
- Zhang, M., Tsoulakos, N., Kujala, P., and Hirdaris, S. (2024), A deep learning method for the prediction of ship fuel consumption in real operational conditions, *Engineering Applications of Artificial Intelligence*, **130**, 107425.
- Zhou, T., Hu, Q., Hu, Z., and Zhen, R. (2022), An adaptive hyper parameter tuning model for ship fuel consumption prediction under complex maritime environments, *J. Ocean Eng. Sci.*, **7**(3), 255-263.

Zhou, Y., Pazouki, K., Murphy, A. J., Uriondo, Z., Granado, I., Quincoces, I., and Fernandes-Salvador, J. A. (2023), Predicting ship fuel consumption using a combination of metocean and on-board data, *Ocean Engineering*, **285**(2), 115509.

## 저자소개

**이현하**: 부산대학교 산업공학과 학사과정에 재학 중이다. 연구 분야는 딥러닝, 빅데이터 분석, 인공지능이다.

**김도희**: 부산대학교 산업공학과에서 2019년에 학사, 2024년에 박사학위를 취득하였고, 2024년부터 현재까지 부산대학교 인간중심 - 탄소 중립 글로벌 공급망 연구센터에서 연수연구원에 이어 연구교수로 재직 중이다. 관심 분야는 시계열 분석, 인공지능, 딥러닝이다.

**추정호**: 대전대학교 경제학과에서 2024년 학사학위를 취득하고 부산대학교 산업공학과 석사과정에 재학 중이다. 연구분야는 딥러닝, 건전성 및 예측 관리, 운영 유지보수이다.

**조상민**: 부산대학교 산업공학과에서 2024년 학사학위를 취득하고 부산대학교 산업공학과 석사과정에 재학 중이다. 연구분야는 딥러닝, 시계열 분석, 프로세스 마이닝이다.

**배혜림**: 서울대학교에서 1996년에 학사, 1996년 석사, 2002년 박사학위를 취득하였고, 2002년부터 2003년까지 삼성카드에서 근무했으며, 2005년부터 부산대학교 산업공학과 교수를 역임하고 2024년부터 부산대학교 데이터사이언스전문대학원 교수로 재직 중이다. 관심 분야는 정보시스템 설계, 클라우드 컴퓨팅, 비즈니스 프로세스 마이닝, 항만 물류, 인공지능이다.