

Bi-LSTM을 이용한 LNG 이중연료 선박의 탄소배출 예측 및 변수 중요도 해석 연구

김현주¹, 김동현², 김경현^{3*}

¹국립부경대학교 산업및데이터공학과 박사수료, ²국립부경대학교 융합공학부(기계조선에너지시스템공학전공) 교수,

³국립부경대학교 실습선(백경호) 기관장

Predicting Carbon Emissions from LNG Dual-Fuel Vessels and Analyzing Feature Importance with Bi-LSTM

Hyun-Ju Kim¹, Dong-Hyun Kim², Kyong-Hyon Kim^{3*}

¹Ph.D. Candidate, Department of Industrial and Data Engineering, Pukyong National University

²Professor, Department of Smart Machine Mobility Engineering, Pukyong National University

³Chief Engineer, Pukyong National University Training ship BAEK KYUNG

요약 최근 국제해사기구(IMO)의 온실가스(GHG) 중기조치에 따라 2028년부터 총톤수 5,000톤 이상 선박은 연료집약도(GFI) 데이터를 수집해야 하며, 이에 따라 연료 효율 향상과 탄소배출 예측의 중요성이 커지고 있다. 본 연구는 이러한 규제 대응을 위해 시계열 분석에 강점을 가진 Bi-LSTM(Bidirectional Long Short-Term Memory) 모델을 적용하여 LNG 이중연료(Dual-Fuel, DF) 엔진 선박의 CO₂ 배출량을 정밀 예측하고, 모델의 해석 가능성을 확보하는 것을 목표로 하였다. 2023년 2월~9월 수집된 약 23만 건의 실운항 데이터 중 정상 항해 구간(6.75노트 이상) 약 1만 건을 활용하였으며, 타깃 변수는 연료 소비량 기반 시간당 CO₂ 배출량으로 설정하였다. 학습 결과, Bi-LSTM 모델은 R² 0.9761, RMSE 0.2113, MAPE 4.60%로 높은 예측 성능을 보였다. 또한 SHAP 분석을 통해 선박 속도(SPEED_VG)와 화물탱크 증기온도(CARGO_VAPOR_TEMP_MEAN)가 주요 영향 변수로 확인되었으며, 두 변수의 증가가 CO₂ 배출량 상승에 기여함을 정량적으로 입증하였다.

주제어 : LNG 이중연료 선박, CO₂ 배출 예측, Bi-LSTM, SHAP 분석, 시계열 데이터

Abstract In response to the International Maritime Organization's (IMO) mid-term greenhouse gas (GHG) strategy, all vessels of 5,000 gross tonnage or more will be required to collect fuel intensity (GFI) data from 2028, underscoring the growing importance of fuel efficiency enhancement and accurate carbon emission prediction. This study aims to address these regulatory demands by applying a Bi-LSTM (Bidirectional Long Short-Term Memory) model, which is well-suited for time-series analysis, to precisely predict CO₂ emissions from LNG dual-fuel (DF) engine vessels while ensuring model interpretability. 230,000 real voyage data points collected between February and September 2023 were processed, and about 10,000 valid normal navigation records (above 6.75 knots) were selected for model training. The target variable was defined as the hourly CO₂ emission rate based on LSFO and LNG fuel consumption. The Bi-LSTM model achieved superior predictive performance, recording an R² of 0.9761, RMSE of 0.2113, and MAPE of 4.60%. Furthermore, SHAP (SHapley Additive exPlanations) analysis identified vessel speed (SPEED_VG) and cargo vapor temperature (CARGO_VAPOR_TEMP_MEAN) as the most influential variables, quantitatively demonstrating that increases in these factors contribute to higher CO₂ emission predictions.

Key Words : LNG dual-fuel ships, CO₂ emission prediction, Bi-LSTM, SHAP analysis, Time-series data

*교신저자 : 김경현(hyunkk@pknu.ac.kr)

접수일 2025년 09월 18일

수정일 2025년 10월 07일

심사완료일 2025년 10월 16일

1. 서론

국제해사기구(International Maritime Organization, IMO)의 최신 온실가스(Greenhouse Gas, GHG) 증기 조치에 따르면, 2028년부터 총톤수 5,000톤 이상의 모든 선박은 연료집약도(Fuel Intensity, GFI) 데이터를 수집해야 하며, 2029년부터 이를 제출해야 한다. 특히, GHG 배출량이 높은 연료를 사용하는 선박은 Remedial Units(톤당 100 USD), Surplus Units, 또는 추가 비용(톤당 380 USD) 등을 통해 초과 배출량을 상쇄해야 하므로, 고탄소 연료 선박의 부담은 지속적으로 증가할 전망이다[1].

이에 따라, 고탄소 연료를 사용하는 기존 디젤 중심의 선박은 LNG, 암모니아 등의 저탄소 이중연료 시스템으로의 전환이 가속화되고 있으며, 동시에 운항 중 연료소모량 및 탄소배출량을 정확하게 예측하고, 이를 결정짓는 운항/환경 요인을 정량적으로 해석하려는 수요가 급증하고 있다. 이는 단순한 기술적 연구를 넘어, 규제 대응, 연료비 절감, 기업 수익성 확보와 직결되는 문제로 부상하였다 [2-3].

특히 LNG 이중연료 선박은 연료 전환 과정에서 발생하는 연소 특성 변화, 저온탱크 내 증발가스(BOG, Boil-Off Gas) 재활용, 연료 혼합비 조정 등 복합적인 요인으로 인해 탄소배출량의 변동성이 상대적으로 크며, 운항 조건에 따라 연료 사용 패턴이 비선형적으로 변화하는 경향이 있다. 이러한 열역학적·운항적 상호작용은 기존 단일 연료 선박보다 배출량 예측의 불확실성을 높이고, 주요 변수의 기여도를 명확히 구분하기 어렵게 만든다. 따라서 LNG 이중연료 선박의 CO₂ 배출량을 정확히 예측하기 위해서는, 시계열적 패턴과 변수 간 비선형 관계를 동시에 고려할 수 있는 고도화된 예측 모델과 함께, 결과를 정량적으로 해석할 수 있는 분석체계의 구축이 요구된다.

최근에는 이러한 복잡한 예측 문제에 대응하기 위해, 고성능 기계학습(Machine Learning) 모델과 딥러닝 모델이 연료소모량 예측에 활용되고 있다. 특히 SHAP(SHapley Additive exPlanations) 기법을 통해 예측 모델의 해석 가능성(Explainability)을 확보하고, 평균 흡수, 풍속, 속력 등 주요 변수의 영향도를 시각화하려는 시도도 늘고 있다 [4-6].

Zhang et al.은 Bi-LSTM(Bidirectional Long Short-Term Memory)에 Attention 메커니즘을 결합하여 벌크선 운항 데이터를 기반으로 연료 소모량을 예

측하였고[6], Safaei et al.은 VLCC 선박의 AIS 및 Noon Report 데이터를 활용해 다중회귀분석 모델을 구축하였다[7]. Yuksel et al.은 해상 디젤 엔진에 대해 다양한 ML 모델을 비교하고, 가장 성능이 뛰어난 모델을 도출하였으며[8], Zwart et al.은 Noon Report 데이터를 활용한 Grey-box 모델을 통해 트림 최적화 기반의 연료소모량 추정 모델을 제시하였다 [9].

그러나 기존 연구는 대부분 단일 연료 기반 선박 또는 정적 변수 중심의 분석에 국한되어 있으며, LNG 이중연료 선박의 시계열 운항 데이터를 기반으로, 예측 정확도와 해석력을 동시에 확보하려는 연구는 드물다. 이에 본 연구는 Bi-LSTM 기반의 시계열 예측 모델과 SHAP 해석 기법을 결합한 ‘해석 가능한 CO₂ 배출 예측 프레임워크’를 제안하며, 이를 통해 (1) CO₂ 배출량을 정밀하게 예측하고, (2) 주요 운항 변수의 영향 메커니즘을 규명하는 것을 연구의 두 축으로 설정하였다. 본 연구는 예측 정확도뿐만 아니라 규제 대응과 연료 전략 최적화를 위한 해석 가능한 지능형 예측 프레임워크로서의 가능성을 제시하고자 한다.

2. 데이터 구성 및 전처리 방법

2.1 데이터 개요

본 연구는 LNG 이중연료(Dual-Fuel, DF) 엔진을 탑재한 1척의 상선의 실제 운항 데이터를 기반으로 하였다. 해당 선박은 5,000톤 이상 국제해해 상선으로, 대표적인 연료 전환 사례에 해당하며, LNG 및 저유황중유(Low Sulfur Fuel Oil, LSFO)를 병행 사용하는 구조를 가진다. 또한 저압 가스분사 방식의 X-DF(eXpandable-Dual-Fuel) 타입 이중연료 엔진을 탑재하고 있으며, 이는 고압 분사 방식의 ME-GI(Methane-Ethane Gas Injection) 타입과 함께 현대 상선에서 가장 널리 사용되는 LNG 기반 이중연료 추진 기술 중 하나이다. 데이터는 2023년 2월부터 9월까지 약 8개월간 축적되었으며, 10분 간격으로 수집된 시계열 형태이다. 초기 수집된 원시 데이터는 총 약 23만 건의 샘플로 구성되어 있다. 데이터는 크게 두 가지 유형으로 구성된다. 첫째, 실시간 센서 기반 데이터는 항해 중 선박 자동화 시스템(AMS, VDR 등)을 통해 수집된 정보로, 기관 부하, 연료 유량, 기상 조건, 조타 상태, 카고 탱크 상태 등 다양한 운항 및 환경 요소를 포함한다. 둘째, Noon Report(일일 운항 보고서)는 항해사 또는 기관장이 입력한 보고 자료로, 하루 단위의 연

료 사용량, 속도, 화물 상태 등의 요약 정보가 포함되어 있다.

본 연구에서는 두 데이터셋을 시계열 기준으로 병합하여 총 956개의 변수(columns)를 포함하는 통합 데이터셋을 구축하였다. 이후 이상치 및 결측치 제거, 데이터 필터링, 변수 통합 등의 전처리 과정을 거쳐 최종적으로 약 1만 건의 유효 샘플을 확보하였으며, 이는 실제 운항 조건에서의 연료 소비 및 탄소 배출 예측을 위한 학습에 활용되었다.

2.2 이상치 및 결측치 처리

수집된 데이터에는 센서 오류나 통신 장애로 인해 이상치 및 결측치가 다수 포함되어 있었다. 먼저, -9999, 9997, 9999 등과 같은 명시적인 이상치는 모두 결측값으로 간주한 뒤 해당 행을 제거하였다. 또한, '-'와 같은 비수치형 결측값 역시 동일한 방식으로 처리하였다. 음수값과 같이 물리적으로 유효하지 않은 경우는 0으로 대체하였다.

2.3 데이터 필터링

이상치와 결측치를 제거한 데이터에는 정박 상태나 저속 항해 구간 등, 탄소배출량 분석과 직접적인 관련이 낮은 구간이 여전히 포함되어 있었다. 본 연구에서는 IMO 및 각국의 규제 대상이 선박 운항 중 발생하는 탄소배출량에 집중되고 있다는 점을 고려하여, 실질적인 항해 구간만을 분석 대상으로 한정하였다.

선박의 평균 항해 속도 분포와 정박 상태를 분석한 결과, 약 6.75노트(knots)를 기준으로 정상 운항 여부를 구분할 수 있었다. 이에 따라, 순간 속도가 6.75노트 미만인 구간은 정박 또는 비정상 저속 운항 상태로 간주하고 분석 대상에서 제외하였다. 이러한 필터링은 실제 운항 중 발생하는 탄소배출량에 집중하기 위한 조치로, IMO를 비롯한 국제적 규제가 정박이나 대기 중이 아닌 항해 중 탄소배출을 주로 규제 대상으로 삼고 있다는 점을 반영하였다. 실제 항해 이력 로그(Event Log) 분석 결과, 전체 분석 기간 중 약 3개월가량만이 정상 항해 상태로 확인되었으며, 대부분의 시간은 정박, 하역, 대기 등 비항해 상태로 기록되어 있었다. 이러한 점을 반영해 속도 기준에 따라 데이터를 필터링한 결과, 최종적으로 약 1만 건의 유효 데이터가 확보되었으며, 이는 정상 항해 조건 하에서의 탄소배출량 예측 모델 학습에 활용되었다.

2.4 변수 생성 및 통합 처리

본 연구에 활용된 선박은 2개의 주기관(Main Engine) 및 4개의 발전기(Generator Engine)를 병렬 운전하는 구조를 갖고 있으며, 각 엔진에는 별도의 센서가 설치되어 유사한 변수들이 중복 수집된다. 본 연구의 목적은 개별 엔진의 성능이 아닌, 선박 전체의 연료 사용 및 탄소 배출량을 예측하는 데 있으므로, 동일 개념의 변수들은 평균을 통해 통합 처리하였다. <Table 1>과 같이 각 엔진별 센서 데이터를 평균하여 대표값으로 재구성하였다. 연료 유량의 경우 LSFO와 LNG 각각에 대해 다중 유량 센서의 값을 합산하여 LSFO_FLOW_TOTAL과 LNG_FLOW_TOTAL을 정의하였다. 카고 탱크 관련 변수는 탱크별 온도 및 압력 데이터를 평균 처리하여 CARGO_VAPOR_TEMP_MEAN, CARGO_PRESS_MEAN 등의 파생 변수를 생성하였다. 이러한 변수 통합 과정은 불필요한 차원을 제거하고, 입력 변수의 의미적 명확성을 확보함으로써 모델 학습과 해석 과정의 효율성을 높이는 데 기여하였다.

<Table 1> Summary of Variable Aggregation and Generation

Variable Name	Description
ME_LSFO_FLOW	Average LSFO flow rate of main engines
ME_LNG_FLOW	Average LNG flow rate of main engines
GE_LSFO_FLOW	Average LSFO flow rate of generators
GE_LNG_FLOW	Average LNG flow rate of generators
DRAFT	Mean vessel draft (fore and aft)
CARGO_VAPOR_TEMP_MEAN	Average cargo tank vapor temperature
CARGO_PRESS_MEAN	Average cargo tank pressure

2.5 입력 변수 정의

모델 학습에 사용된 입력 변수는 총 27개이며, 선박의 운항 상태, 기상 해양 환경, 선체 및 카고 탱크 조건 등을 포괄한다. 변수는 실시간 센서 기반의 시계열 데이터로 구성되어 있으며, 연료 소비 및 탄소배출에 영향을 줄 수 있는 주요 항목들을 다변수 형태로 반영하였다. 입력 변수는 크게 운항 및 환경 변수, 선체 정보, 카고 탱크 상태, 화물/헤더 조건의 네 가지 범주로 구분되며, <Table 2>에 주요 변수를 요약하였다.

〈Table 2〉 Summary of Input Variables

Category	Variable Name
Navigation / Weather	SPEED_VG, COURSE_OVER_GROUND, SHIP_HEADING, RUDDER_ANGLE, ...
Hull Geometry	DRAFT
Cargo Tank Conditions	CTS_AVR_VAPOR_TEMP_SHIP, CTS_LIQUID_VOLUME_SHIP, ...
Cargo Header System	CARGO_VAPOR_TEMP_MEAN, VAPOR_HDR_AFT_TEMP, ...

2.6 타겟 변수 정의

본 연구의 예측 타겟은 시간당 이산화탄소(CO₂) 배출량으로 설정하였다. 탄소배출량은 LSFO 및 LNG 연료 소비량에 각 연료의 배출 계수를 적용하여 eq(1)~(3)과 같이 산정하였다.

$$LSFO_{flow} = ME_{LSFO,flow} + GE_{LSFO,flow} \quad eq(1)$$

$$LNG_{flow} = ME_{LNG,flow} + GE_{LNG,flow} + GCU_{LNG,flow} \quad eq(2)$$

$$CO_2 = \frac{LSFO_{flow} \times 3.114 + LNG_{flow} \times 2.75}{1000} \quad eq(3)$$

여기서 단위는 kg/h이며, 계수 3.114 및 2.750은 각각 LSFO 및 LNG 연료의 배출 계수 (tCO₂/t)이다. tCO₂/h로 환산하였다. LNG 소비량에는 BOG(Boil-Off Gas) 유량도 포함하여, 실질적인 연료 사용량에 기반한 배출량으로 추정하고자 하였다.

3. 예측 모델 구축 및 성능 평가

3.1 예측 모델 개요

본 연구에서는 선박 운항 중 발생하는 이산화탄소(CO₂) 배출량을 예측하기 위해 회귀 기반의 기계학습 및 딥러닝 모델을 활용하였다. 다양한 예측 모델 중에서 구조적 복잡도와 예측 정확도를 고려하여 Bi-LSTM(Bidirectional Long Short-Term Memory) 모델을 선정하였다. 특히, LNG 이중연료 선박의 운항 데이터가 다변량(multivariate) 시계열 형태로 구성되어 있으며, 시간에 따른 연속적 패턴과 변수 간 상호의존성이 높다는 점을 고려하였다. 기존의 단방향 LSTM은 과거 시점의 정보만을 활용하는 한계가 있으나, Bi-LSTM(Bidirectional Long Short-Term Memory) 모델은 순방향과 역방향의 두 개의 LSTM 층을 동시에 학습하여 시계열 데이터의 전후 관계를 모두 반영할 수 있다. 이를 통해 운항 구간 내에서의 엔진 부하, 선속, 연료전환 비율 등과 같은 변수가 CO₂ 배출량

에 미치는 시점별 비선형 영향 관계를 보다 정밀하게 학습할 수 있다. 이 모델은 각각의 특성에 따라 입력 변수와 타겟 변수 간의 비선형 관계를 효과적으로 학습할 수 있으며, 시계열적 변동성 및 다변량 환경에서도 강인한 예측 성능을 보인다[10].

본 연구에서 구축한 Bi-LSTM 네트워크는 두 개의 양방향 LSTM(hidden) 레이어와 하나의 완전연결(Dense) 출력 레이어로 구성되어 있다. 각 LSTM 레이어의 노드 수는 64개로 설정하였으며, 과적합 방지를 위해 Dropout(0.2)을 적용하였다. 입력 시퀀스는 10분 간격으로 수집된 시계열 데이터의 특성을 고려하여 48개 시점을 하나의 슬라이딩 윈도우로 구성하였다. 이는 약 8시간(48×10분)에 해당하는 연속 구간의 운항 데이터를 활용해 다음 시점의 CO₂ 배출량을 예측하도록 설계한 것이다. 활성화 함수로는 하이퍼볼릭 탄젠트(tanh)를 사용하였으며, 출력층에는 선형(linear) 활성화를 적용하였다. 이러한 구조적 설계는 모델이 단기 및 장기 패턴을 동시에 학습할 수 있도록 하여, 실제 LNG 이중연료 선박의 시계열적 CO₂ 배출 변동성을 효과적으로 반영하도록 설계되었다.

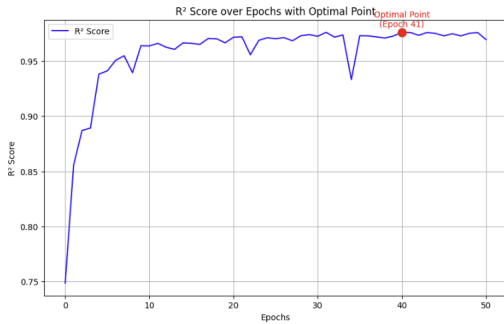
3.2 모델 학습

모델 학습은 전처리 및 필터링을 거친 약 1만 건의 정상 운항 데이터에 대해 수행되었다. 입력 변수는 선박의 항해 환경, 선체 정보, 카고 탱크 상태 및 헤더 조건 등 총 27개 항목으로 구성된다. 모델은 학습 데이터셋과 검증 데이터셋으로 분할하여 훈련되었으며, Bi-LSTM 모델의 경우 시계열 순서를 고려한 슬라이딩 윈도우 방식을 적용하였다. 학습은 PyTorch 기반으로 구현하였으며, 손실 함수는 평균제곱오차(MSE), 최적화 알고리즘은 Adam을 사용하였다. 조기 종료(Early Stopping) 조건은 검증 성능이 10 epoch 동안 개선되지 않을 경우로 설정하였다.

3.3 모델 성능

Bi-LSTM 기반 예측 모델의 성능은 검증 데이터셋을 기준으로 반복 학습(epoch)별 R² 점수의 변화와 주요 회귀 평가 지표를 통해 검토하였다. [Fig. 1]은 에폭 수에 따른 R² 점수의 추이를 나타내며, 최적 성능은 41번째 에폭(epoch)에서 확인되었다. 41번째 에폭에서 기록된 최고 R²는 0.9761이며, RMSE (Root Mean Squared Error) 0.2113, MAE (Mean Absolute Error) 0.1072, MAPE (Mean Absolute Percentage Error) 4.60%로 확인되었다. 모델 학습 초기에는 R² 점수가 급격히 상

승하며 예측력이 빠르게 향상되었고, 이후 10~15 에폭부터는 점진적인 개선 양상을 보였다. 과적합 방지를 위해 조기 종료(early stopping) 기준을 적용하였으며, R² 점수가 10 에폭 이상 개선되지 않은 시점에서 학습을 종료하였다.



[Fig. 1] R² Score over Epochs with Optimal Point

4. SHAP 기반 변수 중요도 해석

4.1 변수 중요도 해석

Bi-LSTM 모델의 예측 결과에 대한 해석 가능성을 확보하고, 주요 입력 변수들이 탄소 배출량 예측에 미치는 영향을 정량적으로 분석하기 위해 SHAP (SHapley Additive exPlanations) 값을 활용하였다.

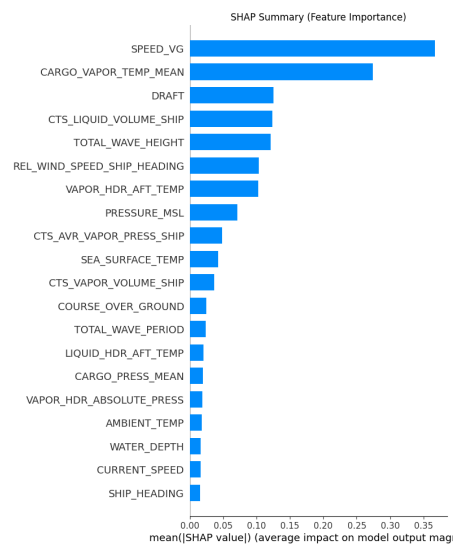
SHAP는 게임이론의 셰플리 값(Shapley Value)에 기반한 해석 기법으로, 각 입력 특성이 예측 결과에 기여한 정도를 공정하게 분배한다는 이론적 기반을 갖는다. SHAP는 모델 불문하고 적용 가능한 모델 비종속적 (Model-agnostic) 기법으로, 입력 변수의 조합을 변경하면서 예측값의 변화에 따른 각 변수의 기여도를 계산한다는 점에서 높은 해석력을 제공한다[11,12].

특히, SHAP는 전역적(Global) 해석뿐 아니라 개별 예측에 대한 지역적(Local) 설명을 제공할 수 있다는 점에서 복잡한 딥러닝 모델의 '블랙박스' 문제를 완화하는데 유용하다[13].

[Fig. 2]는 각 변수의 평균 SHAP 값을 기준으로 정렬된 절대적인 중요도 순위를 시각화한 결과이다. 분석 결과, 가장 높은 중요도를 보인 변수는 선박 속도(SPEED_VG)로 나타났다. 이는 연료 소비 및 배출량이 속도에 따라 비선형적으로 증가하는 선박 엔진의 특성을 반영하는 결과로 해석할 수 있다. 다음으로 높은 영향력을 가진 변수는 화물 탱크 증기 온도(CARGO_VAPOR_TEMP_MEAN)로, 이는 증발가스(BOG)의 양과 LNG 연료 사용량에 밀

접한 영향을 미친다. 이외에도 DRAFT(홀수), CTS_LIQUID_VOLUME_SHIP(화물 액체량), TOTAL_WAVE_HEIGHT(파고), 상대풍속(REL_WIND_SPEED_SHIP_HEADING) 등 외부 항해 조건 및 선체 하중과 관련된 변수들이 상위 중요 변수로 나타났다. 이는 선박의 저항력, 추진력, 증발가스량 등이 복합적으로 작용하여 연료 소모 및 배출량에 영향을 준다는 점을 시사한다.

반면, 일부 항해 변수(COURSE_OVER_GROUND, SHIP_HEADING 등)나 환경 변수(CURRENT_SPEED, AMBIENT_TEMP)는 상대적으로 낮은 SHAP 값을 보였다.



[Fig. 2] SHAP Summary Plot (Feature Importance)

[Fig. 3]은 Bi-LSTM 모델의 이산화탄소(CO₂) 배출량 예측 결과에 대한 전반적인 변수 중요도와 영향 방향을 전체 데이터셋 기준으로 시각화한 결과이다.

Y축은 평균 절대 SHAP 값에 따른 변수 중요도 순위를, X축은 예측값에 대한 기여도(양수: 배출량 증가, 음수: 배출량 감소)를, 색상은 각 변수의 실제 값(붉은색: 높음, 푸른색: 낮음)을 나타낸다.

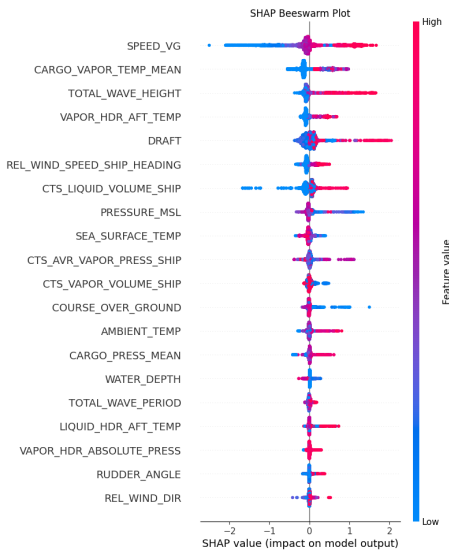
선박 속도(SPEED_VG)는 SHAP 값의 분포 폭이 가장 넓어, 모든 변수 중 가장 높은 중요도를 보인다. 높은 속도(붉은색) 조건은 CO₂ 배출량 예측을 현저히 증가시키는 주요 요인으로 작용함을 확인할 수 있다.

LNG 시스템 변수인 화물 탱크 증기 온도(CARGO_VAPOR_TEMP_MEAN)는 두 번째로 중요한 변수로 나타났다. 높은 온도(붉은색) 조건은 CO₂ 배출량 예측을 증가시키는 요인으로 작용하는데, 이는 LNG 이증연료

선박이 탱크 내 **증발가스(BOG, Boil-Off Gas)**를 연료로 활용하며, BOG 발생률이 온도 상승에 따라 증가하기 때문에 LNG 연료 연소량과 CO₂ 배출량이 함께 늘어나는 결과로 해석할 수 있다.

환경 관련 변수인 흘수(DRAFT)와 파고(TOTAL_WAVE_HEIGHT)는 각각의 값이 높을수록(붉은색) CO₂ 배출량 예측을 증가시키는 경향을 보였다. 이는 운항 중 파고 상승으로 인한 저항 증가를 극복하기 위해 더 큰 추진력이 요구되어 연료 소모가 증가함을 반영한 것으로 분석된다. 반대로, 낮은 흘수(푸른색) 조건에서는 선박의 저항이 감소하여 CO₂ 배출량이 줄어드는 경향을 확인할 수 있다.

또한, LNG 보조 시스템 변수인 증기 헤더 후방 온도(VAPOR_HDR_AFT_TEMP) 역시 온도가 높을수록(붉은색) CO₂ 배출량 증가에 기여하는 것으로 나타나, LNG 관리 시스템 전반이 CO₂ 배출량에 직접적인 영향을 미침을 보여준다.



[Fig. 3] SHAP Beeswarm Plot

5. 결론

본 연구는 LNG 이중연료 선박의 실제 운항 데이터를 기반으로 CO₂ 배출량을 정밀하게 예측하고, SHAP 기법을 활용하여 복잡한 운항 환경에서 각 변수의 기여도를 정량적으로 분석하였다. 분석 결과, Bi-LSTM 모델은 다변량 시계열 운항 데이터를 효과적으로 학습하여 R²

0.9761이라는 매우 높은 예측 정확도를 달성하였으며, 이는 LNG 선박의 비선형적이고 시계열적인 특성을 반영하는 데 있어 Bi-LSTM 기반 접근법이 가장 적합함을 보여주었다. 이러한 결과는 LNG 선박의 에너지 효율 관리와 배출량 예측 분야에서 Bi-LSTM 모델이 유용한 도구가 될 수 있음을 시사한다.

또한 SHAP 분석을 통해 CO₂ 배출량을 결정짓는 핵심 요인을 규명하였다. 특히 선박 속도(SPEED_VG)는 가장 중요한 변수로 나타났으며, 속도가 높아질수록 CO₂ 배출량이 크게 증가하는 경향을 보였다. 이와 함께 흘수(DRAFT)와 파고(TOTAL_WAVE_HEIGHT) 역시 배출량 증가에 기여하는 것으로 확인되었는데, 이는 선체 저항이 증가하면서 추진력이 더 많이 요구되는 물리적 메커니즘과 밀접히 연관된다. LNG 특성 변수인 화물 탱크 증기 온도(CARGO_VAPOR_TEMP_MEAN) 역시 두 번째로 중요한 변수로 도출되었으며, 이는 보일오프가스(BOG)의 발생 및 활용과 직결되어 CO₂ 배출량 증가를 유발하는 주요 요인으로 작용하였다. 특히 SHAP 의존성 분석을 통해, 화물 탱크 증기 온도의 영향이 고속 운항 조건에서 더욱 극대화된다는 상호작용 효과를 확인하였다. 이러한 결과는 단일 변수 제어만으로는 한계가 있으며, 운항 속도와 LNG 연료 시스템의 상호작용을 종합적으로 고려하는 것이 배출 저감을 위한 보다 효과적인 전략임을 시사한다.

본 연구는 Bi-LSTM 모델의 예측 성능과 SHAP 기법의 해석력을 통합하여, LNG 이중연료 선박의 CO₂ 배출 문제를 다각적으로 규명했다는 점에서 학술적 기여도가 크다. 더 나아가 실무적으로는 운항 관리자가 속도, 흘수, 탱크 온도와 같은 주요 변수를 제어하여 배출량을 줄일 수 있도록 지원하는 해석 가능한 AI 기반 운항 최적화 시스템 구축의 근거를 제공하였다.

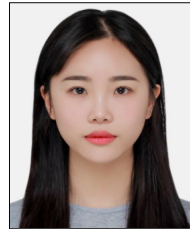
향후 연구에서는 본 연구에서 중요도가 높게 도출된 변수들, 특히 화물 탱크 증기 온도와 같은 LNG 연료 시스템 관련 요소를 보다 세부적으로 분석할 필요가 있다. 예를 들어 증기 온도 관리가 BOG 발생 및 연료 효율에 미치는 영향을 실험적·시뮬레이션 기반으로 검증하고, 다양한 운항 조건과 환경 요인을 반영한 확장 연구를 수행한다면, 보다 정밀하고 실용적인 CO₂ 배출 저감 전략을 제시할 수 있을 것이다. 이러한 추가 연구는 향후 친환경 연료 활용과 해운 산업의 탄소중립 달성에 실질적으로 기여할 것으로 기대된다.

REFERENCES

- [1] Korean Register, "IMO News Final: MEPC 83," KR Technical Publication, 2025. [Internet]. Available: [https://www.krs.co.kr/TECHNICAL_FILE/IMO%20News%20Final\(K\)%20MEPC%2083.pdf](https://www.krs.co.kr/TECHNICAL_FILE/IMO%20News%20Final(K)%20MEPC%2083.pdf)
- [2] J. Lee, J. Eom, Y. Jeong, H. Han and M. Kim, "The Development of a Machine Learning-Based Carbon Emission Prediction Method for a Multi-Fuel-Propelled Smart Ship by Using Onboard Measurement Data," Sustainability, Vol.16, No.6, Article 2381, 2024.
- [3] M. Handayani, Y. Kim, Y. Lee and H. Kim, "Navigating Energy Efficiency: A Multifaceted Interpretability of Fuel Oil Consumption Prediction in Cargo Container Vessel," Journal of Marine Science and Engineering, Vol.11, No.11, Article 2165, 2023.
- [4] V. N. Nguyen, V. Chung and H. P. Nguyen, "Internet of Things-Driven Approach Integrated with Explainable Machine Learning Models for Ship Fuel Consumption Prediction," Alexandria Engineering Journal, In press, 2025.
- [5] Z. Wang, Y. Lu and X. Zhao, "Improving Ship Fuel Consumption and Carbon Intensity Prediction Accuracy Based on a Long Short-Term Memory Model with Self-Attention Mechanism," Applied Sciences, Vol.14, No.6, Article 88526, 2024.
- [6] M. Zhang, A. Tsoulakos and P. Pistofidis, "A Deep Learning Method for the Prediction of Ship Fuel Consumption in Real Operational Conditions," Engineering Applications of Artificial Intelligence, Vol.130, Article 107425, 2024.
- [7] M. Safaei, M. Ghassemi, H. R. Ghaemi and S. A. Sadough Vanini, "VLCC's Fuel Consumption Prediction Modeling Based on Noon Reports," Journal of Marine Science and Application, Vol.18, No.3, pp.394-402, 2019.
- [8] S. Yuksel, H. Bayraktar, Y. Erdem and I. Karaboga, "A Comparative Study of Machine Learning Techniques to Predict the Fuel Consumption of Marine Diesel Engines," Energy Reports, Vol.9, pp.2363-2374, 2023.
- [9] B. Zwart, J. Bogaard and W. Bijl, "A Grey-Box Model Approach Using Noon Report Data for Trim Optimization and Fuel Consumption Prediction on a Chemical Tanker," Ocean Engineering, Vol.284, Article 116556, 2023.
- [10] H. J. Park, M. S. Lee, D. I. Park and S. W. Han, "Time-Aware and Feature Similarity Self-Attention in Vessel Fuel Consumption Prediction," Applied Sciences, Vol.11, No.23, Article 11514, 2021.
- [11] C.-H. Yang, C.-H. Wu, J.-C. Shao, Y.-C. Wang and C.-M. Hsieh, "AIS-Based Intelligent Vessel Trajectory Prediction using Bi-LSTM," IEEE Access, vol. PP, pp. 1-1, 2022.
- [12] Y. Zhu, Y. Zuo and T.-S. Li, "Predicting Ship Fuel Consumption based on LSTM Neural Network," in Proc. 2020 7th Int. Conf. on Information, Cybernetics, and Computational Social Systems (ICCSS), pp. 310-313, 2020.
- [13] Y. Nohara, K. Matsumoto, H. Soejima and N. Nakashima, "Explanation of Machine Learning Models Using Shapley Additive Explanation and Application for Real Data in Hospital," Computer Methods and Programs in Biomedicine, Vol.214, Article 106584, 2021.
- [14] T. Takanashi and M. Nishimura, "Shapley Additive Explanations for Knowledge Discovery in Aerodynamic Shape Optimization," Proceedings of AIAA SCITECH 2023 Forum, Article 0904, 2023.
- [15] T. A. Assegie, "Evaluation of the Shapley Additive Explanation Technique for Ensemble Learning Methods," Proceedings of Engineering and Technology Innovation, Article 9025, 2022.

김 현 주(Hyun-Ju Kim)

[정회원]



- 2017년 1월 : 한국해양대학교 기관공학(공학사)
- 2023년 8월 : 부경대학교 산업및 데이터공학과(공학석사)
- 2023년 9월 ~ 현재 : 국립부경대학교 산업및데이터공학과 (박사수료)

〈관심분야〉

선박 빅데이터, 인공지능, 머신러닝, 딥러닝

김 동 현(Dong-Hyun Kim)

[정회원]



- 2010년 2월 : 부경대학교 조선해양공학(공학사)
- 2019년 2월 : 국립부경대학교 조선해양공학(공학박사)
- 2023년 3월 ~ 현재 : 국립부경대학교 미래융합대학 스마트기계모빌리티전공 조교수

〈관심분야〉

해운 데이터 분석, 리스크 평가, 인공지능

김 경 현(Kyong-Hyon Kim)

[정회원]



- 1989년 2월 : 부산수산대학(부경대학교) 기관학과 (공학사)
- 2019년 8월 : 부경대학교 제어기계공학과(공학박사)
- 1993년 7월 ~ 현재 : 국립부경대학교 실습선 백경호 기관장

<관심분야>

자율운행 선박, 친환경 및 스마트 선박 건조