

## Prediction of drowning person's route using machine learning for meteorological information of maritime observation buoy

Jung-Wook Han\*, Ho-Seok Moon\*\*

\*Pilot, Republic of Korea Navy 6th air wing, Pohang, Korea

\*\*Professor, Dept. of Defense Science, Korea National Defense University, Nonsan, Korea

### [Abstract]

In the event of a maritime distress accident, rapid search and rescue operations using rescue assets are very important to ensure the safety and life of drowning person's at sea. In this paper, we analyzed the surface layer current in the northwest sea area of Ulleungdo by applying machine learning such as multiple linear regression, decision tree, support vector machine, vector autoregression, and LSTM to the meteorological information collected from the maritime observation buoy. And we predicted the drowning person's route at sea based on the predicted current direction and speed information by constructing each prediction model. Comparing the various machine learning models applied in this paper through the performance evaluation measures of MAE and RMSE, the LSTM model is the best. In addition, LSTM model showed superior performance compared to the other models in the view of the difference distance between the actual and predicted movement point of drowning person.

▶ **Key words:** Maritime distress accident, Maritime observation buoy, Machine learning, Prediction of drowning person's route, Surface layer current, LSTM

### [요 약]

해양조난사고 발생 시 해상 익수자의 안전과 생명 보장을 위해 구조자산을 활용한 신속한 탐색 및 구조작전은 매우 중요하다. 본 연구는 해양관측부이에서 수집되는 기상정보에 다중선형회귀분석, 의사결정나무, 서포트벡터머신, 벡터자기회귀, 순환신경망의 LSTM을 활용하여 울릉도 북서해역의 표층해류를 분석하고 유향과 유속에 대한 각각의 예측모형을 구축하여 예측된 유향과 유속 정보를 통해 해상 익수자의 이동경로를 예측하는 모형들을 제안한다. 본 연구에서 적용한 다양한 기계학습 모형을 MAE와 RMSE의 성능 평가척도로 비교해 볼 때 LSTM이 가장 우수한 성능을 보였다. 또한, 익수자 이동지점과 예측모형의 예측지점 간 거리 차이에 있어서도 LSTM이 다른 모형들에 비해 탁월한 성능을 나타내었다.

▶ **주제어:** 해상조난사고, 해양관측부이, 기계학습, 익수자 이동경로 예측, 표층해류, LSTM

- 
- First Author: Jung-Wook Han, Corresponding Author: Ho-Seok Moon
  - \*Jung-Wook Han (mysun2562@naver.com), Republic of Korea Navy 6th air wing
  - \*\*Ho-Seok Moon (bawooi@korea.ac.kr), Dept. of Defense Science, Korea National Defense University
  - Received: 2021. 11. 04, Revised: 2021. 11. 11, Accepted: 2022. 03. 14.
  - This paper is originally from the dissertation of Master's degree at KNDU(Korea National Defense University), and it has been updated and edited of the part of the original dissertation.

## I. Introduction

2014년 진도 여객선 침몰 사고, 2019년 표류선박 삼척항 식별 등 선박표류·충돌·화재로 인한 해상조난사고는 Fig. 1[1]과 같이 매년 꾸준히 증가하고 있는 추세이다.

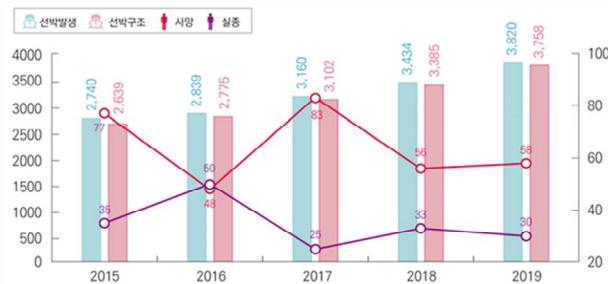


Fig. 1. Maritime Distress Accidents from 2015 to 2019

해상조난사고란 해상에서 선박·항공기 및 수상레저기구 등의 좌초, 충돌, 화재, 기관 고장, 추락 등으로 인하여 사람의 생명, 신체 및 선박·항공기·수상레저기구 등의 안전이 위험에 처한 상태[2]를 의미한다.

현재 우리나라 관할해역에서 해상조난사고가 발생하게 되면 가장 인접한 위치의 민·관·군 함정 및 항공 구조자산을 투입하여 해상 탐색 및 구조작전을 수행하고 있다. 이때 해상조난사고 발생지점 인근에 구조자산이 존재한다면 신속한 구조가 가능하겠지만 대부분의 경우 인접위치에 구조자산이 없으며 해상조난사고가 발생하고 관할기관에서 사고접수 후 구조자산이 사고 현장에 도착할 때까지 짧게는 1시간 내에서 길게는 수 시간의 시간 지연이 발생하게 된다. 그에 따라 사고해역에서의 표층해류 흐름에 의해 최초 발생위치와 구조시 익수자의 위치 간 차이가 발생하게 되며 이는 신속한 탐색 및 구조작전에 제한사항으로 작용하여 결론적으로 해상 익수자의 안전과 생명을 보장할 수 없게 된다. 그러므로 해상조난사고 발생 시 사고해역의 표층해류 분석 즉, 시간에 따른 표층해류의 유향과 유속 추정을 통해 최초 사고위치로부터 해상 익수자의 이동경로를 예측하여 신속한 탐색 및 구조작전을 수행하는 것은 익수자의 안전을 보장하기 위해 매우 중요한 사항이다.

현재까지 표층해류의 순환과 관련된 연구는 주로 표층 뜰개를 투하하고 그 궤적을 추적하여 표층해류의 순환을 연구하는 라그랑지 해류관측 방법에 대해 이루어져 왔으며 이 방법은 동해상에 투하된 표층 뜰개 자료의 분석결과로부터 동해 및 동중국해의 표층해류모식도 연구에 많이 사용되었다[3]. 이러한 라그랑지 해류관측은 투하된 뜰개부

만이 포함하게 된다. 반면 오일리식 정점 해류관측은 국립해양조사원에서 운용중인 해양관측부이 36개소에서 수집되는 기상정보로 짧게는 1년에서 길게는 수십년 동안 동·서·남해 및 연안해역의 기상정보를 축적하고 있다[4].

본 연구에서는 표층해류에 대한 보다 많은 정보를 갖고 있는 오일리식 정점 해류 관측인 해양관측부이에서 수집되는 기상정보를 이용하여 다양한 기계학습 기법을 적용해서 우리나라 주요항로 중 울릉도 북서해역의 표층해류 순환에 대해 연구하고자 한다. 이를 위해 기계학습 기법간의 성능 비교를 통해 최적의 기법을 선정하고, 선정된 기법을 활용해 표층해류에 따른 해상 익수자의 이동경로를 예측하고자 한다.

2장에서는 관련 이론과 기존 연구에 대해 설명하고, 3장에서는 주요항로인 울릉도 북서해역의 표층해류 순환 분석 및 해상 익수자 이동경로 예측을 위한 연구방법을 수립하고 기계학습 모형 설계 및 성능 검증을, 4장에서는 전체 내용에 대한 정리와 본 연구의 기여점, 제한사항 및 향후 연구방향에 대하여 제시한다.

## II. Preliminaries

### 1. Related works

해수면에 일정한 바람 및 기압골에 의해 수온약층에 해당하는 해수층의 이동인 표층해류와 관련된 연구들은 대부분 수치모형을 이용하여 연구되었다. 1998년 류청로 등 4명의 연구[5]에서는 우리나라 동남해역의 유류오염사고 자료와 해수 유동 모형인 2차원 수심 적분 수치모형과 유출유 거동 모형을 이용하여 계산된 유출유 확산예측 결과를 비교하였다. 또한 2015년 전연선 등 4명의 연구[6]에서는 미국 해양대기청(NOAA)의 해양 표층해류 시뮬레이터(OSCURS) 수치모형을 이용하여 계절별 동·서해로 유입되는 플라스틱 폐기물의 이동경로를 예측하고 효율적인 수거지점을 제안하였다. 표층해류와 관련된 연구들 중 수치모형 이외의 기계학습 기법을 활용한 연구도 있었다. 2017년 이찬재 등 3명의 연구[7]에서는 라그랑지 해류관측인 뜰개부이의 기상정보를 이용하여 여러 기계학습 기법간의 성능 비교를 통해 뜰개 이동 예측을 위한 최적의 기법을 선정하였다. 또한 2018년 이찬재 등 2명의 연구[8]에서는 뜰개부이의 기상정보를 이용하여 배킹과 부스팅의 앙상블 기법을 활용한 기계학습 기법을 통해 뜰개부이의 이동경로를 예측하였다. 표층해류와 관련된 기존 연구는 대부분 수치모형을 이용한 부유물의 예측과 관련된 연구

였으며, 기계학습 기법을 통한 표층해류 분석도 있었으나, 연구들이 모두 라그랑지 해류 관측인 뜰개부이 기상정보를 이용한 분석이 대부분이었다. 기존 연구 가운데 오일러식 정점 해류관측인 해양관측부이 기상정보를 이용한 표층해류 분석 연구는 확인하지 못하였으며, 또한 표층해류 분석과 연계하여 해상 익수자의 이동경로를 예측한 연구는 확인하지 못하였다. 본 연구에서는 Table 1과 같이 뜰개부이보다 상대적으로 많은 양의 데이터를 가지고 있는 해양관측부이 기상정보를 이용하여 표층해류를 분석하고 해상 익수자의 이동경로를 예측하는 연구를 제안한다.

Table 1. Merits and Demerits of Each Buoy

| Buoy     | Observation Buoy | Drifter Buoy |
|----------|------------------|--------------|
| Merits   | Big Data         | Movement     |
| Demerits | Fixed            | Small Data   |

## 2. Related theories

표층해류 분석 및 해상 익수자 이동경로 예측을 위해 본 연구에서는 다변량 자료의 예측기법 중 선형회귀분석 모형, 의사결정나무 모형, 서포트벡터머신 모형의 3가지 데이터마이닝 방법론과 벡터자기회귀 모형, 순환신경망 모형의 2가지 시계열 분석 방법론을 사용하였다.

### 2.1 Linear Regression, LR

선형회귀분석 모형은 설명변수 또는 독립변수라고 불리는 한 개 또는 두 개 이상의 변수들에 대하여 반응변수 또는 종속변수라고 불리는 다른 한 변수 사이의 관계를 선형적인 모형을 이용하여 규명하고, 이 규명된 함수식을 이용하여 설명변수들의 변화로부터 반응변수의 변화를 예측하고자 할 경우에 사용되는 통계적 분석기법[9]이다. 선형회귀분석 모형의 종류에는 설명변수가 한 개일 때 사용하는 단순회귀분석 모형과 설명변수가 두 개 이상일 때 사용하는 다중회귀분석 모형이 있으며 본 논문에서는 설명변수가 5가지 기상치로 다중회귀분석 모형을 사용한다. 다중회귀분석 모형의 수식은 식 (1)과 같다.

$$y_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip} + \varepsilon_i, \quad i = 1, 2, \dots, n \quad (1)$$

$y_i$  : response,  $x_{ip}$  : predictor,  $\beta_p$  : parameter,  $\varepsilon_i$  : error

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & \dots & x_{1p} \\ 1 & x_{21} & \dots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \dots & x_{np} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

### 2.2 Decision Tree, DT

의사결정나무 모형은 의사결정규칙을 도표화하여 관심 대상이 되는 집단을 몇 개의 소집단으로 분류, 예측, 세분

화를 수행하는 통계적 분석기법[10]이다. 이 방법은 Fig. 2와 같이 분석과정이 나무구조에 의해서 표현되기 때문에 분류와 예측을 목적으로 하는 다른 통계적 분석방법들에 비해 분석과정을 쉽게 이해하고 설명할 수 있다.

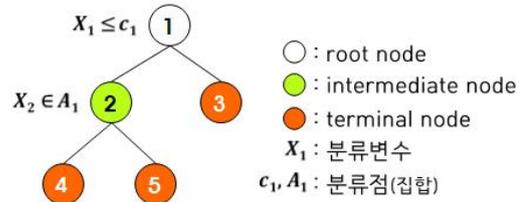


Fig. 2. Structure of Decision Tree

### 2.3 Support Vector Machine, SVM

서포트벡터머신 모형은 받침점의 역할을 하는 소수의 관측 개체들로 분류 및 회귀모형을 구축하는 통계적 분석기법[11]이다. 이 방법은 분류 경계면과 가장 가까이 있는 관측치들을 받침점으로 명명하여 분류를 위한 결정 경계면을 정의하는 역할로 사용한다. 따라서 서포트벡터머신 모형은 받침점만 잘 선택한다면 나머지 수많은 관측치들을 무시하고 결정 경계면을 정의할 수 있으므로 다른 기법들에 비해 속도가 빠르다.

### 2.4 Vector AutoRegression, VAR

벡터자기회귀 모형은 단변량 자기회귀모형을 다변량 자기회귀모형으로 확장시킨 모형으로, 예측할 변수의 과거값만 이용하는 자기회귀 모형에 추가하여 예측할 변수와 의존성 있는 변수들까지 고려하여 선형함수로 나타내는 확률적 과정의 통계적 분석기법[12]이다. 벡터자기회귀 모형의 수식은 식 (2)와 같다.

$$y_i(t) = a_i + \sum_{j=1}^n \sum_{k=1}^p w_{ij} y_j(t-k) + \varepsilon_i, \quad i = 1, 2, \dots, n \quad (2)$$

$y_i$  : variables,  $a_i$  : constant,  $w_{ij}$  : parameter,  $\varepsilon_i$  : error

$$\begin{bmatrix} y_1(t) \\ y_2(t) \\ \vdots \\ y_n(t) \end{bmatrix} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} + \begin{bmatrix} w_{11} \cdot w_{1j} \\ w_{21} \cdot w_{2j} \\ \vdots \\ w_{n1} \cdot w_{nj} \end{bmatrix} \begin{bmatrix} y_1(t-1) \\ y_2(t-1) \\ \vdots \\ y_n(t-1) \end{bmatrix} \dots \begin{bmatrix} w_{11} \cdot w_{1j} \\ w_{21} \cdot w_{2j} \\ \vdots \\ w_{n1} \cdot w_{nj} \end{bmatrix} \begin{bmatrix} y_1(t-p) \\ y_2(t-p) \\ \vdots \\ y_n(t-p) \end{bmatrix} + \begin{bmatrix} \varepsilon_1(t) \\ \varepsilon_2(t) \\ \vdots \\ \varepsilon_n(t) \end{bmatrix}$$

### 2.5 Recurrent Neural Network, RNN

순환신경망은 인공신경망의 한 종류로 내부 은닉층에 루프구조가 포함되어 있는 신경망이다. Fig. 3[13]과 같이 입력데이터가 은닉층인  $h_t$ 를 통과하면서 활성화 함수를 통해  $y_t$ 라는 출력데이터가 되며, 이때 순환구조에 의해 현 정보에 이전 은닉층인  $h_{t-1}$ 의 결과값이 쌓이게 되고 그에 따라 데이터가 순환되기 때문에 정보가 끊임없이 갱신된다. 순환



Table 3. The preprocessed Data

| Date             | W. Temp (°C) | Temp (°C) | Pres (hPa) | Wind (deg) | Wind Spd (Knots) | Current (deg) | Current Spd (Knots) |
|------------------|--------------|-----------|------------|------------|------------------|---------------|---------------------|
| 2020-09-30 23:00 | 22.85        | 20.05     | 1012.15    | 202.5      | 7.483784         | 202.5         | 0.5034546           |
| 2020-09-30 22:00 | 22.9         | 20        | 1012       | 202.5      | 8.552896         | 180           | 0.5802362           |
| 2020-09-30 21:00 | 23           | 20        | 1011.95    | 202.5      | 10.10796         | 180           | 0.5277526           |
| 2020-09-30 20:00 | 22.95        | 19.95     | 1011.75    | 202.5      | 10.01077         | 157.5         | 0.4976230           |
| 2020-09-30 19:00 | 22.9         | 20.35     | 1011.45    | 202.5      | 7.386592         | 180           | 0.5219210           |
| 2020-09-30 18:00 | 22.9         | 20.35     | 1011.15    | 202.5      | 7.386592         | 180           | 0.3576666           |
| 2020-09-30 17:00 | 22.9         | 20.35     | 1010.6     | 202.5      | 7.386592         | 135           | 0.1875806           |
| ⋮                | ⋮            | ⋮         | ⋮          | ⋮          | ⋮                | ⋮             | ⋮                   |

의 변화가 크지 않다는 특징을 이용하여 1시간 이전의 기상 자료값으로 대체하였다. 또한 원천 데이터의 단위 통합을 위하여 풍속, 유속 단위를 Knots 단위로 변환하였으며 16방위인 풍향을 360도로 변경하여 Table 3과 같이 전처리 과정을 거친 최종 데이터셋을 구성하였다.

## 2. Performance Evaluation Measures

다변량 자료의 예측기법 중 회귀분석 및 시계열 모형 등의 예측모형에서는 필연적으로 예측오차가 발생한다. 이러한 예측오차는 예측모형의 신뢰성 및 안정성을 평가하는 기준으로 적용된다. 본 연구에서는 이러한 예측오차를 이용하여 통계량을 구한 성능평가척도 중 MAE와 RMSE를 사용하여 표층해류 분석 및 해상 익수자 이동경로 예측 모형의 성능을 평가하였다. MAE는 오차의 크기만 고려하기 위해 식 (4)와 같이 실제값과 예측값의 차이에 절대값을 데이터의 크기로 나눈 값으로 모든 오차들에 대해 동일한 가중치가 주어진다.

$$MAE = \frac{1}{n} \sum_{i=1}^n |Y_i - \hat{Y}_i| \quad (4)$$

$$Y_i = \text{true}, \hat{Y}_i = \text{prediction}$$

RMSE는 MAE와 마찬가지로 오차의 크기만 고려하기 위해 식 (5)와 같이 실제값과 예측값의 차이를 제곱하여 데이터의 크기로 나눈 값으로 MAE와 달리 큰 오차에 대해서 패널티가 주어진다.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2} \quad (5)$$

$$Y_i = \text{true}, \hat{Y}_i = \text{prediction}$$

본 연구에서는 교차검증을 통해 모형의 성능을 평가하였다. 교차검증은 모형의 일반화 오차에 대해 신뢰할 만한 추정치를 얻기 위해 원천 데이터를 훈련 및 평가 데이터로 구분하여 모형을 평가하는 방법이다. 본 연구에서는 주어진 원천 데이터를 약 90%인 4,537개의 훈련 데이터와 약 10%인 477개의 평가 데이터로 분할하였으며, LSTM 모형의 경우 모형의 학습을 위한 검증용 데이터를 위해 약 10%인 477개의 검증 데이터를 훈련 데이터에서 추가로 분할하였다. 이렇게 분할된 데이터를 활용하여 5가지 예측모형을 학습 및 구축하였으며 구축된 모형에 대하여 평가 데이터를 통해 성능평가척도를 비교하였다.

또한, Fig. 6[16]은 해상 익수자의 수온별 최대 생존가능시간을 나타내는 그림으로, x축은 수온, y축은 해상 익수자의 최대 생존시간이다. 이 정보를 이용하여 모델 성능 평가 척도로 사고해역의 수온 범위에서의 최대 생존가능시간 내 실제 유향·유속에 따른 익수자의 이동지점과 예측모형에 의해 예측된 유향·유속에 따른 익수자의 이동 예측지점 간 시간대별 거리 차이의 평균치를 성능평가척도로 활용하였다.

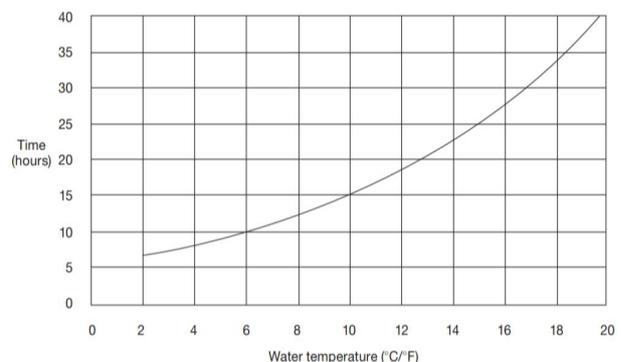


Fig. 6. Realistic Upper Limit of Survival Time for People in the Water Wearing Normal Clothing, from Time of Entry into Water

### 3. Structure of Various Models

본 연구에서 모형 구축을 위해 사용한 데이터의 독립변수는 수온·기온·기압·풍향·풍속 다섯 개이고, 종속변수는 유향·유속 두개이다. 이 데이터에 적용하는 모형의 종류는 다섯 가지이며, RStudio 3.6.3 및 Python 3.8.12을 통해 각 모형별로 유향과 유속을 예측하는 두 가지 예측 모형을 구축하고 시각화하였다.

#### 3.1 Linear Regression, LR

회귀 모형은 step-wise 변수선택 방법을 적용하여 유향과 유속의 반응변수에 대한 회귀분석 결과를 아래 Table 4, 5와 같이 도출하였다.

Table 4. LR Result of log(Current Speed)

| predictor   | estimate  | std.error | t-value | Pr(> t )     |
|---|-----------|-----------|---------|--------------|
| intercept   | 25.59248  | 1.968997  | 12.998  | <2e-16 ***   |
| temp  | 0.003048  | 0.001926  | 1.583   | 0.114        |
| press   | -0.026357 | 0.001905  | -13.835 | <2e-16 ***   |
| wind spd  | 0.013341  | 0.002015  | 6.622   | 3.96e-11 *** |
| F(3, 4533)=161.1, P-values : <2e-16***, R <sup>2</sup> : 0.10 |           |           |         |              |

\*\*\* P < 0.001, \*\* P < 0.01, \* P < 0.05, . P < 0.1

Table 5. LR Result of Current Direction

| predictor   | estimate | std.error | t-value | Pr(> t )    |
|---|----------|-----------|---------|-------------|
| intercept   | -787.978 | 217.333   | -3.626  | 0.000291*** |
| temp  | -5.52864 | 0.48426   | -11.417 | <2e-16 ***  |
| w. temp   | 3.26961  | 0.33764   | 9.684   | <2e-16 ***  |
| press   | 0.94357  | 0.21096   | 4.473   | 7.91e-06*** |
| wind  | 0.09165  | 0.01330   | 6.892   | 6.26e-12*** |
| wind spd  | 1.37246  | 0.22356   | 6.139   | 9.01e-10*** |
| F(5, 4531)=42.09, P-values : <2e-16***, R <sup>2</sup> : 0.05 |          |           |         |             |

\*\*\* P < 0.001, \*\* P < 0.01, \* P < 0.05, . P < 0.1

유속과 유향에 대한 회귀분석 결과 두 모형 모두 유의확률이 0.05 이하로 통계적으로 유의한 모형이나 설명력을 보면 두 모형 모두 10% 이하로 매우 낮아 사용하기 제한되는 수준이었다. 이런 이유는 유향·유속과 관련된 데이터가 주기성이 있고 비선형적이기에 선형 모형에 적합하지 않은 형상으로 평가된다.

#### 3.2 Decision Tree, DT

본 연구에서는 의사결정나무 모형 중에 CART 알고리즘을 사용하였다. 반응변수가 연속형으로 잔차제곱합의 감소가 최대가 되는 변수 방향으로 노드를 분리하였으며 complexity parameter를 1%로 지정하여 불순도의 상대

적 감소가 1% 미만일 경우 노드 분리를 종결하였다. 가지치기를 통해 구축한 log(유향)과 유속에 관한 모형 결과는 Fig. 7, 8과 같다.

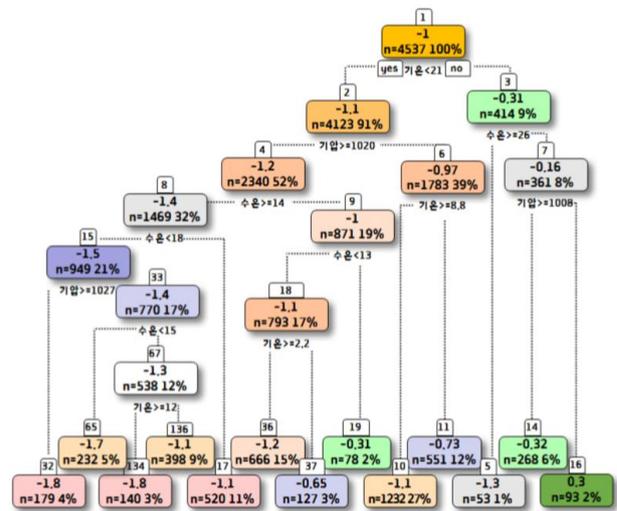


Fig. 7. Decision Tree Result for Current Speed

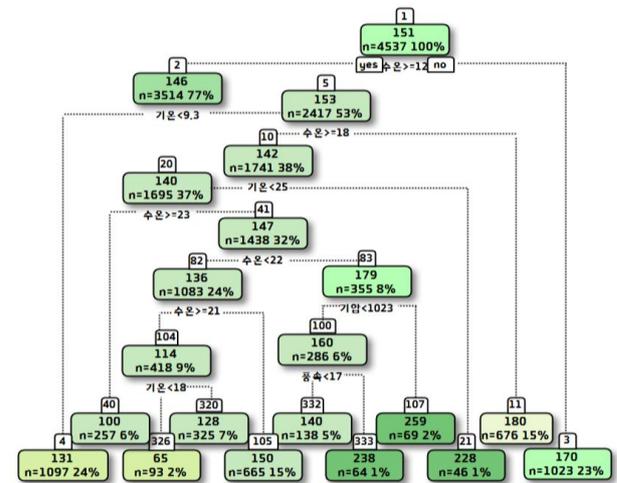


Fig. 8. Decision Tree Result for Current Direction

#### 3.3 Support Vector Machine, SVM

본 연구에서는 가우스 커널(radial kernel)을 사용하여 비선형 SVM 모형을 구축하였다. 모형의 최적 파라미터를 찾기 위해 10-fold 교차검증 방법을 사용하여 가우스 커널의 gamma값과 cost값의 조합 중 최적의 조합을 확인하였으며, Table 6, 7과 같이 각 파라미터 조합의 평균 총 오류율이 가장 낮은 조합을 선택하여 SVM 모형을 구축하였다.

Table 6. Tuning summary of Current Speed

| Parameter tuning of SVM |            |           |
|-------------------------|------------|-----------|
| Sampling method         | 10-fold CV |           |
| Best parameter          | gamma : 1  | cost : 10 |

Table 7. Tuning summary of Current Direction

| Parameter tuning of SVM |            |          |
|-------------------------|------------|----------|
| Sampling method         | 10-fold CV |          |
| Best parameter          | gamma : 10 | cost : 1 |

### 3.4 Vector AutoRegression, VAR

벡터자기회귀 모형은 반응변수인 유향과 유속도 설명변수로 함께 사용하여 모델을 구축하였다. 먼저 벡터자기회귀 모형을 적용하기 위해서는 예측될 변수의 정상성, 즉 평균과 분산, 공분산이 시간에 따라 불변하는 특징을 만족해야한다. 따라서 예측될 변수인 유속과 유향에 대해 정상성을 확인하기 위해 국소회귀를 통한 비선형 관계를 추정하는 분해기법인 STL(Seasonal and Trend decomposition)을 실시하였으며, STL 결과 두 예측변수인 유속과 유향 모두 계절성이 강한 것을 확인할 수 있다. 따라서 정상성을 만족하기 위해 Table 8과 같이 차분을 통해 계절성을 제거하였으며, 통계적 검정방법인 ADF(Augmented Dickey-Fuller) 검정 결과 두 변수 모두 1차 차분 시 유의확률이 0.05이하로 아주 작아 통계적으로 유의하며 귀무가설이 기각되므로 정상성을 만족한다고 볼 수 있었다.

다음은 벡터자기회귀 모형 구축을 위한 Lag Time인 P값 판단으로, 본 연구에서는 벡터자기회귀 모형의 P값 범위를 지정하여 격자탐색을 실시하였다.

탐색 결과 Fig. 9와 같이 24시간을 기준으로 AIC 값이 감소하다가 증가하는 경향이 있으므로 24시간이 최적의 AIC 값을 주는 sweet spot P값으로 확인되며 최종적으로 VAR(24) 모형을 구축하였다.

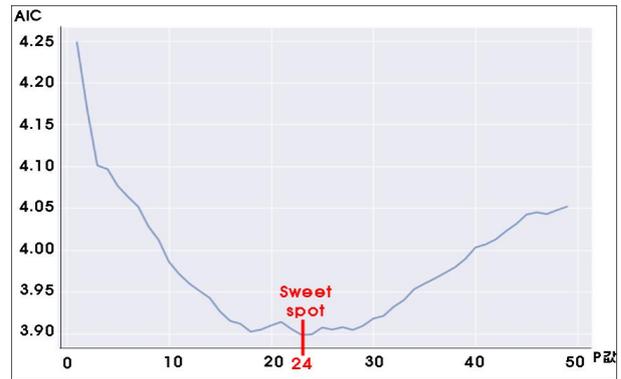


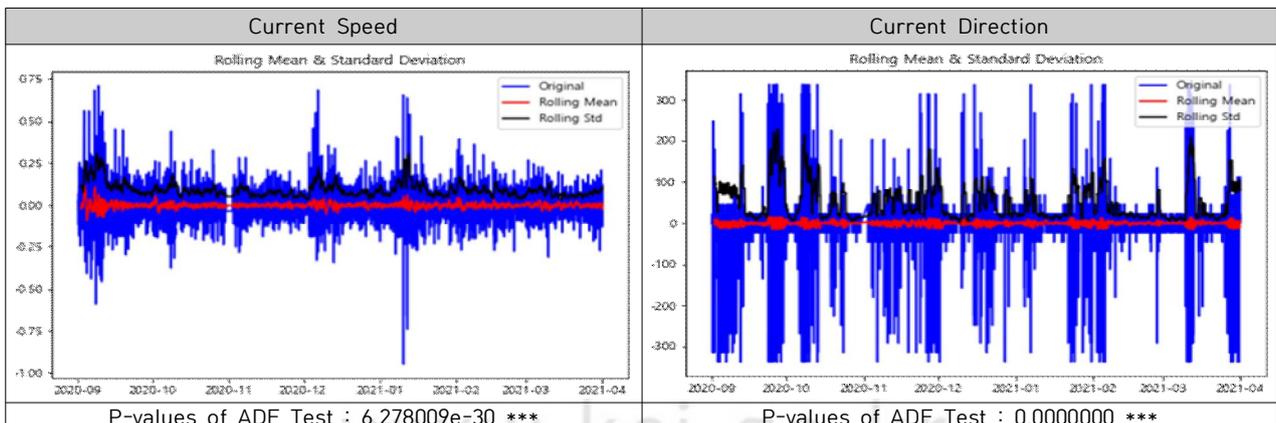
Fig. 9. Grid-search result to find the optimal lag time(P)

### 3.5 Recurrent Neural Network, RNN

본 연구에서는 순환신경망의 한 종류인 LSTM(Long Short-Term Memory) 모형을 구축하였으며, 이는 기존의 순환신경망 모형에서 발생할 수 있는 장기 의존성 문제를 해결할 수 있는 모형이다[17]. LSTM 모형을 구축하기 위해 수온·기온·기압·풍향·풍속·유향·유속의 7가지 기상치를 입력 데이터로 유향·유속의 2가지 기상치를 출력 데이터로 선정하였다.

또한 내부 구성 층의 개수와 위치, 층 내부의 매개변수 값에 대한 설정 기준이 없으므로 최적의 조합을 찾기 위해 층과 매개변수 값의 조합을 조정해가며 모형을 구축하고 학습 및 평가를 통해 가장 좋은 결과를 보여주는 모형을 구축하였다. Fig. 10이 본 연구에서 구성한 2-branch LSTM 모형의 구조이다. 구조를 살펴보면 유향, 유속, 나머지 기상치 데이터가 입력되는 입력층, 유향·유속 데이터에서 중요한 장기·단기 기억을 추출하는 LSTM 층, 유향·유속 데이터와 나머지 기상치 데이터를 합쳐주는 Concatenate 층, 입력과 출력을 완전 연결하는 Dense 층, 과적합을 방지하기 위해 지정된 비율로 데이터를 제외하는 Dropout 층으로 구성하였다.

Table 8. Differences and ADF TEST Results



\*\*\*  $P < 0.001$ , \*\*  $P < 0.01$ , \*  $P < 0.05$ , .  $P < 0.1$

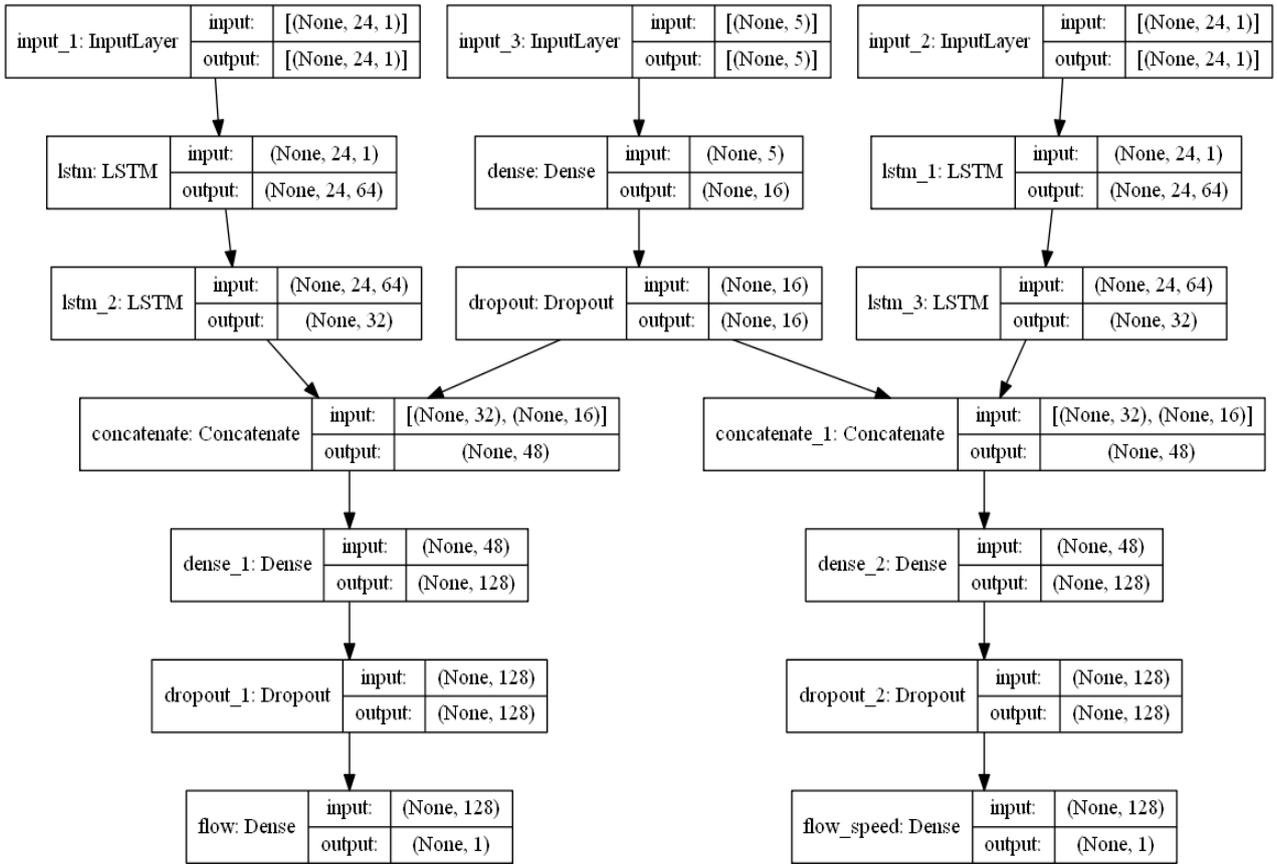


Fig. 10. 2-branch LSTM Model Structure

2-branch LSTM 모형의 학습 과정을 보면 Fig. 11과 같이 유향과 유속의 훈련 및 검증 데이터의 손실 함수값이 일정하게 수렴되는 것을 볼 수 있으며, 최종적으로 Table. 9와 같이 매개변수를 선정하여 2-branch LSTM 모형을 구축하였다.

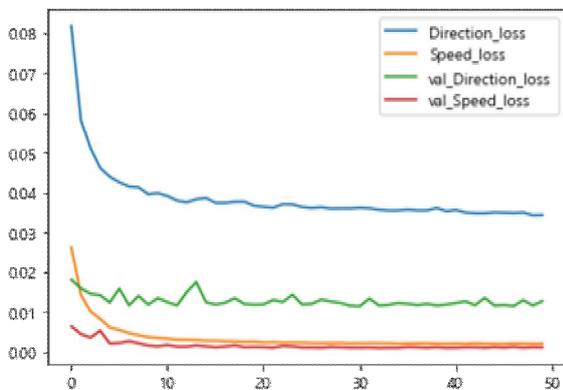


Fig. 11. Learning results of 2-branch LSTM Models

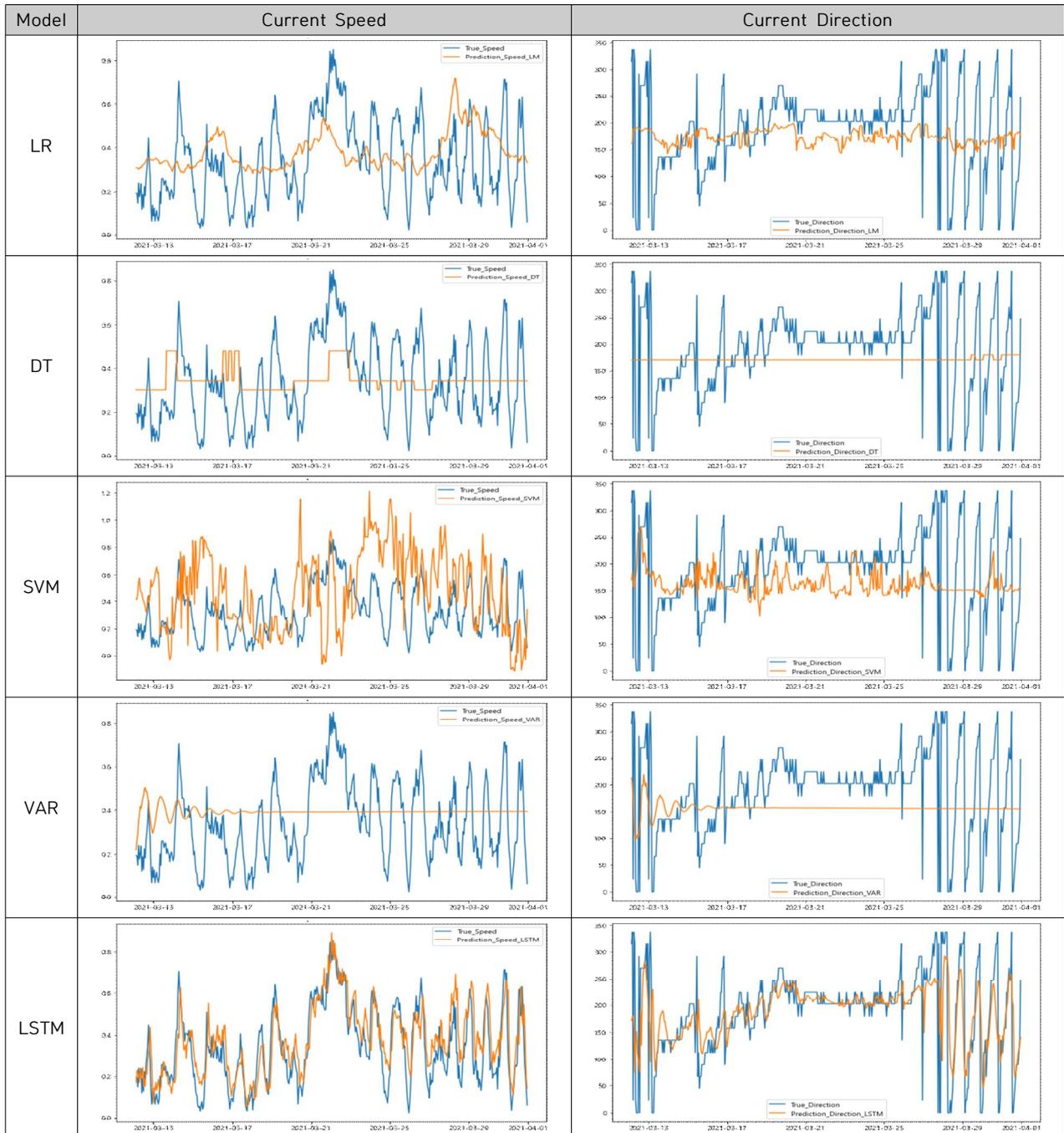
Table 9. Parameters of 2-branch LSTM Model

| Parameter           | Value in LSTM |
|---------------------|---------------|
| Epochs              | 50            |
| Batch_size          | 72            |
| Learning rate       | 0.01          |
| Activation Function | LeakyReLU     |
| Optimizer           | adam          |
| Loss Function       | MAE           |

#### 4. Analyzing the Results by Evaluation Measure

본 연구에서는 앞에서 구축한 LR, DT, SVM, VAR, LSTM을 통해 예측된 울릉도 북서해역의 유향·유속 데이터와 실제 유향·유속 정보 간의 오차를 성능평가 척도인 MAE와 RMSE로 분석하였다. 먼저 Table 10은 실제 유향·유속 정보와 각 예측 모형별 유향·유속 예측값의 그래프를 도식화한 것으로 DT와 VAR을 제외한 나머지 예측모형들은 대체로 실제 유향·유속 정보를 따라가는 경향이 있는 것을 볼 수 있다. DT의 경우 분류변수인 기상치 다섯 가지가 인접 시간대에 유사한 값을 가져 분류점에 따라 예측을 잘 하지 못하는 모습을 볼 수 있으며, VAR의 경우 예측 시간대가 길어질수록 유향·유속의 예측값이 평균으로 회귀하는 모습을 볼 수 있다. 또한 그래프에서 볼 수 있

Table 10. True and Predicted plot of Prediction Model



듯이 LSTM의 예측값이 실제 유향·유속 정보와 거의 유사하게 예측되는 것을 볼 수 있다. Table 11은 각 예측 모형별 성능평가 척도값을 나타내며 LSTM가 다른 예측 모형들에 비해 MAE와 RMSE가 모두 상대적으로 매우 작은 값을 가지는 것을 볼 수 있으며, 성능평가 척도에 따른 비교 시 LSTM가 유향·유속 예측에 가장 뛰어난 모형인 것을 볼 수 있다.

Table 11. Performance Evaluation Measure of Prediction Models

| Model       | Current Speed    |                  | Current Direction |               |
|-------------|------------------|------------------|-------------------|---------------|
|             | MAE              | RMSE             | MAE               | RMSE          |
| LR          | 0.170 kts        | 0.196 kts        | 61.90°            | 78.84°        |
| DT          | 0.152 kts        | 0.179 kts        | 64.98°            | 80.24°        |
| SVM         | 0.285 kts        | 0.356 kts        | 67.99°            | 84.09°        |
| VAR         | 0.164 kts        | 0.191 kts        | 68.49°            | 82.79°        |
| <b>LSTM</b> | <b>0.084 kts</b> | <b>0.106 kts</b> | <b>36.92°</b>     | <b>62.14°</b> |

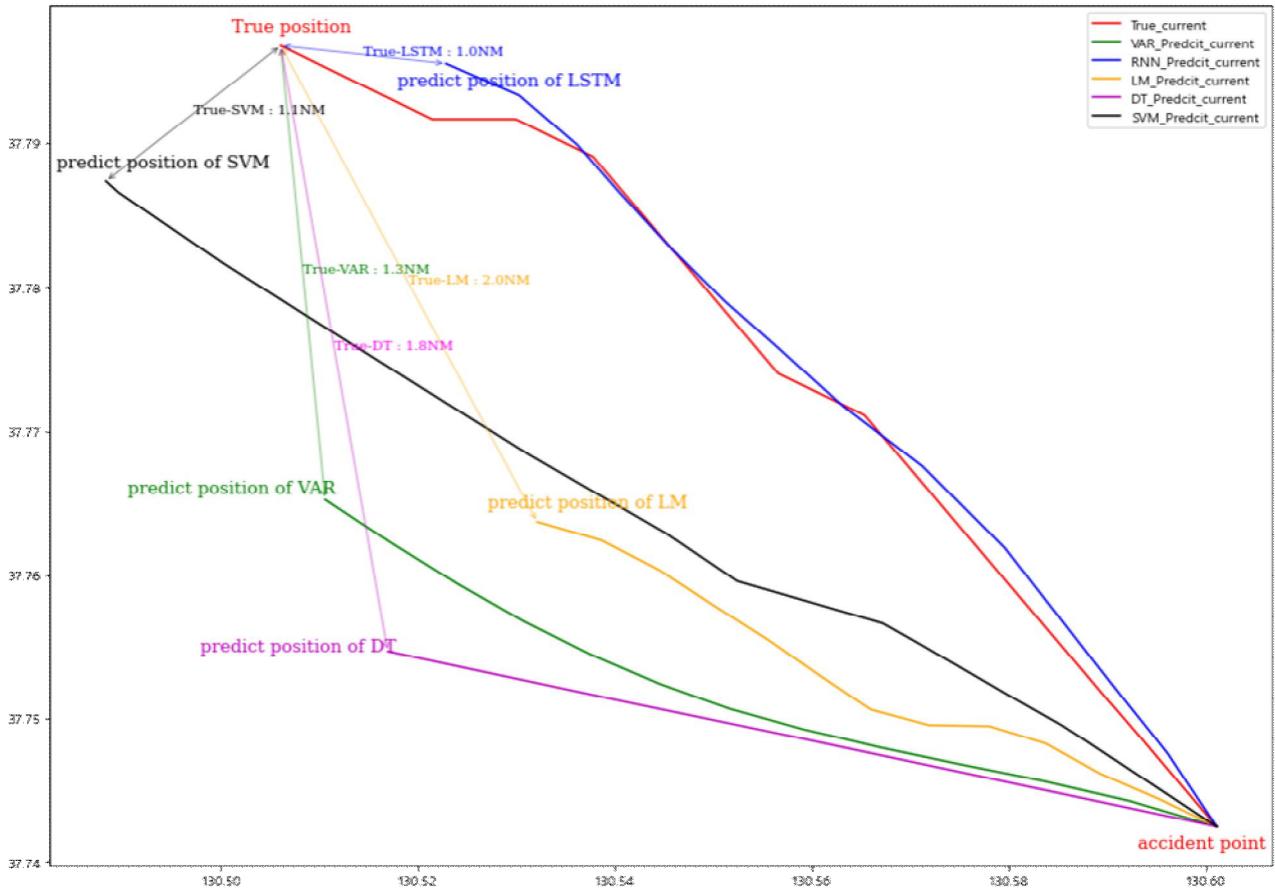


Fig. 12. Actual Route and Predicted Routes by Models of Drowning person at Sea

Table 12. Distance Difference Value Between the Actual and Models by Time Period

| Model | 1h    | 2h   | 3h   | 4h   | 5h   | 6h   | 7h   | 8h   | 9h   | 10h  | 11h  | 12h  | mean           |
|-------|-------|------|------|------|------|------|------|------|------|------|------|------|----------------|
| LR    | 0.36  | 0.65 | 0.90 | 1.13 | 1.32 | 1.48 | 1.63 | 1.69 | 1.78 | 1.84 | 1.94 | 2.02 | 1.40 NM        |
| DT    | 0.35  | 0.62 | 0.85 | 1.06 | 1.17 | 1.32 | 1.46 | 1.56 | 1.63 | 1.62 | 1.69 | 1.76 | 1.26 NM        |
| SVM   | 0.26  | 0.40 | 0.39 | 0.57 | 0.61 | 0.70 | 0.95 | 1.28 | 1.39 | 1.33 | 1.47 | 1.12 | 0.87 NM        |
| VAR   | 0.27  | 0.50 | 0.71 | 0.90 | 0.97 | 1.12 | 1.26 | 1.34 | 1.36 | 1.26 | 1.26 | 1.25 | 1.02 NM        |
| LSTM  | 0.351 | 0.43 | 0.40 | 0.37 | 0.41 | 0.43 | 0.52 | 0.54 | 0.66 | 0.88 | 0.99 | 1.00 | <b>0.58 NM</b> |

또한 해상조난사고 해역의 수온범위에 따른 해상 익수자의 최대 생존가능시간 내 익수자의 실제 이동위치와 예측 이동위치 간 거리차이 확인을 통해 각 예측모형의 성능을 평가하였다. Fig. 12는 실제 유향·유속에 따른 해상 익수자의 이동지점과 각 모형에 의해 예측된 유향·유속에 따른 해상 익수자의 이동 예측지점을 도식화한 그림이다. 이는 검증을 위한 평가 데이터가 21년 3월 14일의 울릉도 북서해역 기상자료이므로 당시 울릉도 해역의 최고 수온인 10.5°C 고려 해상 익수자의 최대 생존가능시간인 12시간 동안의 이동위치를 도식화하였으며 LSTM의 해상 익수자 예측지점이 다른 모형에 비해 실제 이동지점과 가장 근접한 것을 볼 수 있다. Table 12는 12시간 내 시간대별 해상 익수자 이동위치와 모형별 예측 이동위치 간 거리

차이값과 그 평균치로 이 또한 LSTM가 다른 모형에 비해 실제 위치로부터 이격된 평균거리가 0.58NM(0.9km)로 상당히 짧아 우수한 결과를 보인다는 것을 알 수 있었다.

#### IV. Conclusions

우리나라 해역에서 해상조난사고가 발생하면 해상 익수자의 안전과 생명을 보장하기 위해 신속한 탐색 및 구조작전은 매우 중요한 사항이다.

본 연구에서는 국립해양조사원에서 운용중인 울릉도 북서해역의 해양관측부이에서 수집되는 기상정보를 이용하여 다양한 기계학습 기법을 통해 표층해류 분석 모형을 구

축하였고, 나아가 기계학습 모형 간의 성능 비교를 통해 최적의 기법을 선정하였다. 또한 선정된 모형을 활용하여 해상조난사고 발생 시 표층해류의 순환에 따른 해상 익수자의 이동경로를 예측하는 방법을 제시하였다.

본 연구를 통해서 선정된 최적의 기법은 LSTM로, 성능 평가척도를 살펴보면 평가 데이터에 대해 유속 MAE 0.084kts, 유향 MAE 36.92°로 나타나며 해상 익수자의 수온별 최대 생존가능시간을 고려하여 시간대별 해상 익수자 이동위치와 LSTM의 예측 이동위치 간 거리 차이의 평균치가 0.58NM(0.9km)로 다른 모형들에 비해 탁월한 성능을 나타내는 것을 확인할 수 있었다.

본 연구의 기여점은 다음과 같다. 첫째, 해상 기상정보를 이용한 기계학습 기법을 통해 표층해류의 순환과 연계하여 해상 익수자의 이동경로에 대한 예측을 처음으로 시도한 연구이다. 둘째, 본 연구에서는 오일리식 정점 해류 관측인 해양관측부이에서 수집되는 수천시간 이상의 기상정보를 이용한 연구로 기존의 라그랑지 해류 관측인 뜰개 부이에서 수집되는 수백시간의 기상정보를 이용한 연구에 비해 데이터의 크기가 매우 커져 더욱 효과적인 모형을 구축한 연구이다. 셋째, 해상 기상정보를 이용한 표층해류의 순환 및 해상 익수자 이동경로 예측 연구에 딥러닝 기법인 LSTM을 활용한 모형 구축이 가장 좋은 방법론임을 보여 주고 있다. 넷째, 실제로 대부분의 해상조난사고에 대해 탐색 및 구조작전을 수행하고 있는 기관인 해군과 해경에서 본 연구결과를 직접적으로 활용할 수 있을 것이다. 즉 시간대별 익수자 예상위치를 기준으로 탐색 및 구조작전을 수행한다면 최초 해상조난사고 발생지점으로 이동하여 탐색 및 구조작전을 수행하는 현재의 방식보다 더욱 효과적으로 익수자의 안전과 생명을 보장할 수 있을 것이다.

본 연구의 제한사항 및 향후 연구방향은 다음과 같다. 첫째, 본 연구에 사용된 데이터가 울릉도 북서해역의 기상정보로 모형의 적용이 동해 해상에 한정되어 있다는 점이다. 현재 국립해양조사원에서 운용중인 해양관측부이 36개소의 기상정보를 모두 활용한 연구가 진행된다면 우리나라 동, 서, 남해 및 연안을 포함한 더욱 효과적인 모형 구축이 가능할 것이다. 둘째, 본 연구는 표층해류 분석을 통해 해상 익수자 이동경로를 예측한 것으로 실제 해상조난사고가 발생했을 때 구조자산이 예측지점으로 신속히 이동하더라도 탐색 시 시각에만 의존하게 된다면 신속한 탐색 및 구조는 여전히 제한된다고 볼 수 있다. 따라서 항공기 또는 함정에 장착된 영상 감시 시스템 등을 활용하여 해상 영상정보에 적합한 영상 인공지능 모형을 활용한 해

상 익수자 분류 모형 구축 등의 방법론에 대한 연구가 추후적으로 이루어진다면 더욱 효과적인 해상 탐색 및 구조작전능력 향상에 크게 기여할 수 있으리라 기대된다.

## REFERENCES

- [1] Korea Cost Guard, "2019 Maritime Distress Accident Statistical Yearbook", pp. 2, JUL. 2020.
- [2] Korea Cost Guard, "Act On The Search And Rescue, Etc. In Waters Article 2". AUG. 2019
- [3] Korea Hydrographic and Oceanographic Agency Ministry of Oceans and Fisheries Republic of Korea, "Research planning in the establishment of a system for predicting changes in ocean currents according to climate change, Result report", pp. 3-12, DEC. 2016
- [4] <http://www.khoa.go.kr/oceangrid/gis.category/refrence/distribution.do>
- [5] C. Ryu, J. Kim, D. Seol and D. Kang, "Oil Spill Behavior Forecasting Model in South-eastern Coastal Of Korea", Journal of the Korean Society for Maritime Environment & Energy 1(2), pp. 52-59, AUG. 2018
- [6] Y. Jeon, M. Hong, M. Park and Y. Choi, "Path Prediction and Suggestion of Efficient Collection Points for Maritime Plastic Debris Based on Betweenness Centrality Analysis", Journal of the Korean Society of Environmental Engineers 37(7), pp. 426-431, JUL. 2015. DOI: 10.4491/KSEE.2015. 37.7.426
- [7] C. Lee, G. Kim and Y. Kim, "Performance Comparision of Machine Learning Based on Neural Networks and Statistical Methods for Prediction of Drifter Movement", Journal of the Korea Convergence Society vol. 8, No. 10, pp. 45-52, OCT. 2017. DOI: 10.15207/JKCS.2017.8.10.045
- [8] C. Lee and Y. Kim, "Ensemble Design of Machine Learning Techniques: Experimental Verification by Prediction of Drifter Trajectory", Asia-pacific Journal of Multimedia Services Convergent with Art, Humanities, and Sociology, vol. 8, No. 3, pp. 57-67, MAR. 2018. DOI: 10.35873/ajmahs.2018.8.3.006
- [9] Faraway, J, "Linear Models with R 2<sup>nd</sup> Edition", A Chapman & Hall/CRC, pp.13, 2015
- [10] M. Heo, "Applied Data Analysis Using R", Freecaca, pp.239-248, OCT. 2014
- [11] M. Heo, "Applied Data Analysis Using R", Freecaca, pp.221-238, OCT. 2014
- [12] Hyndman. R.J and Athanasopoulos. G, "Forecasting: Principles and Practice, 2<sup>nd</sup> edition", OText, pp. 11.2, MAY. 2018
- [13] <http://ratsgo.github.io/natural%20language%20processing/2017/08/09/mnllstm/>

- [14] <http://www.khoa.go.kr/oceangrid/khoa/koofs.do>
- [15] Russom. P, "Big Data Analytics", TDWI Research, pp. 6, Fourth Quater. 2011
- [16] ICAO and IMO, "International Aeronautical And Maritime Search And Rescue Manual Volume II Mission Co-ordination", pp. N-19, 2016
- [17] Y. Noh and K. Cho, "A Text Content Classification Using LSTM For Objective Category Classification", Journal of the Korea Society of Computer and Information 26(5), pp. 39-46, MAY. 2021. DOI: 10.9708/jksci.2021.26.05.039

## Authors



Jung-Wook Han received the B.S. degree in oceanography from Korea Naval Academy, Korea, in 2013. And He received the M.S. degree in Defense Science (Military Operation Research) from Korea National

Defense University, Korea, in 2022. He is currently a Pilot in the Republic of Korea Navy 6th air wing. He is interested in Artificial Intelligence, Data Analysis and M&S.



Ho-Seok Moon received the B.S. degree in Chemistry from Korea Military Academy, Korea, in 1994. And He received the M.S. degree in Electronic Engineering from Korea University, Korea, in 2003. and Ph.D.

degrees in Industrial Engineering and Statistics from Korea University, Korea, in 2006 and 2010, respectively. Dr. Moon joined the faculty of the Department of Defense Science at Graduate School of Defense Management, Korea National Defense University, Nonsan, Korea, in 2015. He is currently a Professor in the Department of Defense Science, Korea National Defense University. He is interested in Big Data Analysis, M&S, and Artificial Intelligence.