Predicting Vessel Tracks in Waterways for Maritime Anomaly Detection

Finn-Matthis Minßen[1], Jannik Klemm[1], Matthias Steidel[2], Arto Niemi[1]

Many approaches to vessel track prediction and anomaly detection rely only on a vessel's positional data. This paper examines whether including tide and weather data into the track prediction model improves accuracy. We predict vessel tracks in waterways using a bi-directional Long Short-Term Memory (Bi-LSTM) approach and a transformer model. For this purpose, the boundaries of the Elbe and Weser river waterways are merged with vessel position data. Additionally, tide data, as well as weather information, will be used to train the model. To ascertain whether this additional data improves the accuracy, the models have been trained with and without tide and weather data and evaluated against each other. Furthermore, we have investigate whether the predictions can be used for detecting anomalous vessel behaviour. Our results show that the lowest average error and the best RMSE, MSE, and MAE values have been achieved with the Bi-LSTM, where no tide and weather data have been used for training. We have also found that the transformer model is more accurate than a linear prediction model, which is used as a baseline. In addition, we have shown that deviations between predicted and real tracks can be labelled as anomalous. The results have shown that including tide and weather data does not necessarily improve the predictions. Adding data with a low information content to train a machine learning model may introduce noise or bias into the model. We believe that this phenomenon explains our results. Thereby this paper shows that simply adding this data to train the track prediction model may not enhance the overall accuracy.

KEY WORDS

~ Vessel track prediction
~ Bi-directional LSTM
~ Transformer model
~ AIS data
~ Tide data
~ Weather data
~ Anomaly detection

[1] German Aerospace Centre, Institute for the Protection of Maritime Infrastructures, Bremerhaven, Germany
[2] German Aerospace Center, Institute of Systems Engineering for Future Mobility, Oldenburg, Germany

e-mail: arto.niemi@dlr.de

doi: 10.7225/toms.v13.n01.002
Received: 13 Dec 2023 / Revised: 22 Mar 2024 / Accepted: 25 Mar 2024 / Published: 20 Mar 2024

1. INTRODUCTION

The shipping industry is an important sector of the global economy, with around 80% of global trade by volume carried out by sea (UNCTAD, 2021). This high volume leads to dense traffic, especially in coastal regions, due to often limited and narrow waterways. This situation increases the likelihood of accidents, which pose major risks. For example, recently in the North Sea, the container ship *Mumbai Maersk* ran aground off Wangerooge (Szymanska & Murray, 2022), and a small cargo vessel (*Petra L*) collided with an offshore wind turbine (Voytenko, 2023). Predicting vessel tracks is becoming increasingly important to avoid these types of incidents, as these predictions can be used for detecting anomalous vessel behaviour (Nguyen & Fablet, 2021).

This paper uses two types of model to predict vessel tracks in waterways. The first is a bi-directional Long Short-Term Memory (Bi-LSTM) model that combines LSTM cells (Hochreiter & Schmidhuber, 1997), with bi-directional input data processing (Schuster & Paliwal, 1997). The second is a so-called transformer model (Vaswani et al. 2017). We use historical AIS data, tide information, weather data, and topological information as inputs for these models. The topological information is used for representing the waterways as grids, following the approach applied by Steidel et al. (2020). We then compare the results with a linear model consisting of a simple dense layer with a linear activation function. This model is a good baseline because it is simple and makes reasonable predictions.

These experiments are motivated by previous works. Steidel et al. (2020) use Kernel Density Estimation (KDE) for predicting vessel tracks. However, their approach predicts vessel behaviour independently of the behaviour in the previous corridor. This limitation can lead to great variations for each of the parameters considered. Our approach predicts tracks through multiple corridors. Furthermore, we compare the prediction accuracy for models with and without the weather characteristics: wind speed, wind direction, wave height, and tidal data. Steidel et al. (2020), for example, have already considered enhancing the prediction performance with weather information. A review paper by Zhang et al. (2022) has also argued that combining several data sources to predict tracks has been insufficiently studied. To the best of our knowledge, transformer models have so far only used AIS data for track prediction (Nguyen & Fablet, 2021).

This paper further investigates whether these predictions can be used to detect anomalous vessel behaviour. We have evaluated the viability of predictions with data that was not used for training. Track predictions have been made based on this data. If a vessel's behaviour diverges from the prediction, it has been flagged as anomalous. This approach is also motivated by the work of Steidel et al. (2020), as they mention that their approach can be used to detect when a vessel is moving outside a waterway or on the opposite lane of the waterway.

This paper is organised as follows: Section 2 describes recent works on vessel track prediction and anomaly detection. Our approaches are detailed in Section 3. Part 3.1 describes the input data sources, 3.2 the Bi-LSTM and transformer models, and 3.3 the anomaly detection approach. Section 4 provides the results. They show that the average prediction errors obtained with the Bi-LSTM and the transformer models are lower than those of the linear model. However, including tide and weather data does not improve accuracy. The approach can detect anomalous vessel behaviour. Yet in certain areas waterways are so distinct that vessels in them are often flagged as anomalous. This issue raises the question of whether these areas should be separately tackled – in other words, outside the prediction model as proposed here. A more thorough discussion of the results, as well as the conclusions, is presented in Section 5.

2. RELATED WORK

2.1. Vessel Track Prediction

The prediction of vessel tracks has been intensively researched in recent years. Zhang et al. (2022) summarise 57 different works covering this area. Their summary shows that machine learning has become increasingly important since 2020. State-of-the-art models often use variations of LSTM models to predict the expected vessel movement.

One approach, introduced by Mehri et al. (2021), trains separate LSTM models for each type of vessel. To train these models, AIS data were used from November 2017 to the end of December 2017 from the eastern coast of the United States. In addition, geographic information was used to simplify vessel tracks. The models were then evaluated using the Root Mean Square Error (RMSE) and the pointwise horizontal error. Their accuracy was compared with an ordinary LSTM model. The results show that the new method achieves a lower point-wise prediction error than an ordinary LSTM model.

In addition, variations of LSTM models were used to predict vessel tracks. Li et al. (2023) designed a Bi-LSTM to propose short-term vessel predictions. The model was trained with AIS data from 1,364 vessel tracks collected during one month in the Taiwan area. It was evaluated by comparing the accuracy with other methods, such as LSTM and Recurrent Neural Network (RNN) models. The results have shown that the Bi-LSTM has the lowest error rate in all categories compared to the other models. Based on this, Li et al. assert that the developed model can accurately predict short-term trajectories.

A Bi-LSTM model can also be found in the work of Liu et al. (2021). They developed a series of routing algorithms, where a Bi-LSTM was augmented with an 'attention mechanism' to predict the next position of a vessel along the trajectory. The attention mechanism should make the prediction more accurate by more effectively learning the dependencies upon the AIS data. This method was trained with AIS data from fishery vessels along the east coast of China from May 2015 to May 2018. The results showed that the methods predicted vessel positions to an error of less than 300 m after one hour, and an error of 2.73 km after nine hours when using an iterative process.

Zhang et al. (2021) combined two LSTM model variants to predict vessel tracks. First, an encoder-decoder LSTM model is used to extract features from AIS data. These extracted features are then combined back with the orginal AIS data. Then, in the final step, an attention-based Bi-LSTM is used to predict the subsequent vessel tracks. To train the model, AIS data from the US East Coast from January 2016 was used. Their results showed the MSE of the constructed model to be lower compared to an ordinary LSTM model. They concluded that a combination of a bi-directional LSTM and an attention mechanism improved the accuracy of real-time trajectory predictions.

A sequence-to-sequence LSTM for trajectory predicting was developed by Nguyen et al. (2018). They divided the Mediterranean Sea into grids of 1 x 1 nautical miles and trained a sequence-to-sequence LSTM model that predicts the trajectory by predicting the following grids that the vessel will approach. The model was then trained with AIS data from March to May 2015 and evaluated against a Gated Recurrent Unit (GRU) and other networks. The results indicated that the LSTM showed the lowest log perplexity and was therefore well suited for predicting vessel trajectories, as the authors note.

Another track prediction method based on an encoder-decoder LSTM architecture was proposed by Forti et al. (2020). In this method, the encoder consists of 64 LSTM cells and the decoder comprises 32 cells. The performance was evaluated so that the model was given five and 20 steps as input sequences and the next 20 output sequences were to be predicted, sampling the timestamps at two-minute intervals. The data was collected from June to September 2018 and include 534 different voyages from the port of Piombino to Portoferraio on the isle of Elba, Italy. The authors performed a qualitative comparison based on a stochastic Ornstein-Uhlenbeck (OU) model, developed by Millefiori et al. (2016) by calculating the RMSE. The authors state that the encoder-decoder LSTM method achieves a competitive performance compared to the OU method, especially near waypoints.

Sekhon & Fleming (2020) developed a method for short-term predictions of vessel tracks using LSTM with spatial and temporal attention. In addition to an LSTM model, their encoder uses spatial attention to extract information from neighbouring vessels that may affect the trajectory of the vessel under consideration. The decoder then uses both spatial and temporal attention to learn the important parts of the encoded vector. The final prediction is also made using LSTMs. The model was trained with AIS data from January 2017 around the port of San Diego, USA. The results show both the Average Displacement Error and the Final Displacement Error to be lower when using the LSTM in combination with spatial and temporal attention, compared to methods using only one of the two attention mechanisms, as well as an original LSTM architecture.

A different approach for trajectory predicting was developed by Nguyen & Fablet (2021). They argue that standard deterministic approaches, such as LSTMs, cannot capture the multi-model patterns involved in AIS data and are therefore ineffective for trajectory prediction. Nguyen & Fablet (2021) therefore propose a transformer structure to deal with the multimodal nature of AIS tracks. The transformer model contains eight layers, each equipped with eight attention heads. It was trained and tested with AIS data along the Danish coast during the first three months of 2019. The model was then evaluated, among other things, against the sequence-to-sequence LSTM model proposed by Forti et al. (2020). Thereby a significantly lower error in the forecast, measured with the haversine distance, is shown both in the first three hours and after ten hours. Nguyen & Fablet (2021) propose that the developed model is more suitable for capturing the multi-model nature of AIS data and extracting useful information from historical data than the compared models.

Recently, Zou et al. (2023) proposed combining the LSTM model with a transformer model. Their approach attempts to predict trajectories of multiple vessels at the same time in order to capture interactions between vessels. First, individual vessel trajectories are predicted with an LSTM model. Next, the so called vessel attention factor and motion gate parameters are calculated based on these predictions with a view to capturing interaction between vessels. This data is then fed into a transformer model to generate a more refined prediction.

Other concepts for vessel behaviour have been proposed. Löwenstrom et al. (2022) used a Markov Decision Process framework, where an agent observes the current state of the environment and takes action with a set interval. Their method is based on a neural network that was trained using imitation learning, similar to reinforcement learning. They considered the ship's type, wind conditions, and tidal data in their method.

Yet much of the existing research uses solely the vessel location information provided via the AIS to predict vessel tracks. Other factors, such as weather conditions, tides, and regional geographical characteristics also influence the track taken. As Zhang et al. (2021) pointed out, combining several data sources to predict vessel tracks has not yet been sufficiently researched.

In this paper, a data-driven model is trained to predict a vessel's next positions within a waterway. To create a dataset, AIS data, as well as data from waterways will be used. Furthermore, the paper examines the extent to which tide and weather information can influence predictions. Here an iterative approach is used to predict the next positions inside the waterway. The approach makes the prediction model slightly less complex and should therefore lead to more precise results.

2.2. Anomaly Detection

Stach et al. (2023) provide a thorough survey on anomaly detection in vessel traffic services. They state that most commonly used tools are rather simple, such as vessel-speed, a number of alerts, anchor watch or geofence-based monitoring. Because the task is safety-critical, the reliability and explainability of the technique play important roles. Therefore the most advanced techniques have not yet been used in practice.

The technique chosen for anomaly detection depends on the type of anomaly being studied. In literature, data-driven models for anomaly detection often consist of two parts. Firstly, the normal behaviour of a vessel is learned from historical data. The resulting predictions are then applied for detecting anomalies (Yan & Wang, 2019). An anomaly thereby indicates that a vessel's track is significantly different from the expected one, with a certain threshold. Therefore the methods can detect many forms of abnormal vessel behaviour. However, for detecting collision risks one must study encounters involving two or more ships (Guo et al., 2023). In these cases, a particular behaviour of an individual ship may appear normal. The hazardous situation becomes apparent only when the overall vessel traffic is considered.

Venskus et al. (2021) developed an autoencoder LSTM that predicts a region the vessel should be in after up to 2.4 hours. This model was trained and evaluated with AIS data from cargo vessels along the Danish waters from 2006 to 2020, as well as meteorological data from 2019 to 2020. With this approach, the model learns the normal vessel behaviour which can then be used to classify vessel behaviours as abnormal if the actual position of the vessel is outside the predicted region. The authors also developed a statistical wild bootstrap approach. Nevertheless, their results showed that the LSTM gave more practical predictions regarding vessels' behaviour. More recently Murray et al. (2023) tested autoencoders to detect anomalies in the Oslo fjord in Norway, where autonomous barges transport cargo between Moss and Horten.

Ristic (2014) developed an approach whereby anomalies can be detected based on specific positions, as well as on the speed of the vessel. To accomplish this, the area of interest is divided into 'cells' where vessels normally travel. Anomalies are then detected by comparing the position and speed with the distribution learnt. The method was tested with AIS data from January to May 2009 in the port of Jackson, Australia.

One approach that is frequently referred to when predicting anomalies is called the Density-Based Algorithm for Discovering Clusters (DBSCAN) (Ester et al., 1996). With this approach, clusters of arbitrary shapes can be efficiently discovered. To achieve this, a neighbourhood around each datapoint is defined from which dense regions of points are identified as clusters. Based on this approach and applied to the maritime sector, Pallotta et al. (2013) developed an unsupervised methodology to incrementally extract information from AIS data and detect low-likelihood vessel behaviour. This approach, called Route Extraction for Anomaly Detection (TREAD), was then further developed by the authors to detect whether a vessel is off-route, in reverse traffic on the route, or whether the speed is not compatible with the route followed (Pallotta & Jousselme, 2015). They used AIS data in the Ligurian Sea from January to February 2013, extracted this data into tracks, and then compared these tracks to the extracted routes. An anomaly is flagged when a threshold value of the route is exceeded. The results show that 87.3% of the routes are correctly classified as anomalous. The authors further state that the ability to detect anomalies is highly dependant on the regularity of the traffic patterns in the observed area.

Recently, Zhang et al. (2024) presented an anomaly detection method based on a graph attention network method. Their model has a time graph attention module and a feature graph attention module to capture temporal dependencies and correlations between ship features. The local and long-term ship feature correlations are then characterised with a joint detection strategy that employs reconstruction and prediction modules. The results from these characterisations are combined to calculate an anomaly score.

Using the developed model based on the approach of Steidel et al. (2020) to detect anomalies has the advantage that the information extracted from the waterways could also be used to detect whether a vessel is within these waterways. This information is relevant because, depending on the type of vessel, only within the waterways would there be a minimum depth at which the vessel could navigate. Also, instead of the KDE used by Steidel et al. (2020), a data-driven model could be used to predict the tracks more accurately within the waterways. Based on this approach, deviations from the 'true' or expected track could be measured, and if the deviations are above a certain threshold, the track would then be marked as anomalous.

3. MATERIAL AND METHODS

3.1. Maritime Data

3.1.1. Data Preparation

AIS data are generated in the automatic exchange of data between vessels, which gives information about their characteristics and positions. AIS data is collected from base stations that monitor traffic locally and from satellites collecting data on a global scale. Since the use of AIS is mandatory for specific vessels, this data gives a view of ship traffic worldwide and allow for the creation of vessel tracks (Stróżyna et al., 2022). We have extracted from the AIS data the Speed Over Ground (SOG), Course Over Ground (COG), vessel position, identification number (MMSI), and the time at which the vessel sent the message.

Based on the approach of Steidel et al. (2020), we assume that vessels in the area of interest have to sail within the waterways, because only there can a minimum depth be ensured. Waterways are bounded by starboard and port buoys that structure traffic within the waterways. These buoys are placed at indeterminate intervals and at crossings to help navigate within the waterways. Within these waterways, vessels are obliged to sail as close as possible to the starboard buoy (IALA, 2017). To extract the required information, the waterways are divided into a grid, where port side buoys p_i and starboard side buoys s_i form a cell c_i, as shown in Figure 1.

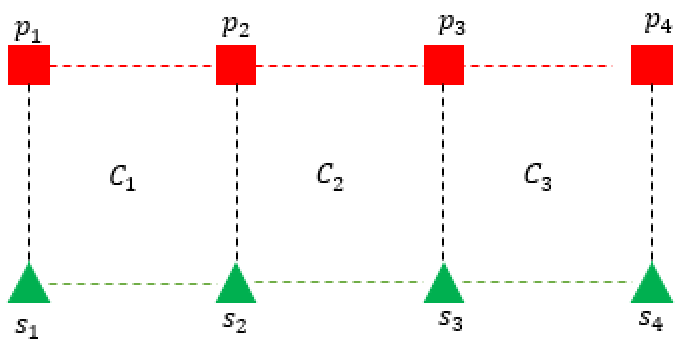

Figure 1. Waterways are extracted as grids

With this approach, a waterway forms a sequential series of cells. All AIS positions occurring within these cells can be filtered. This filtered AIS data is then used to create continuous AIS tracks along the waterways being considered. An AIS track is defined as a series of AIS messages for a particular MMSI received within one minute of the previous message. If the interval between received messages is longer than one minute, a new track is created. From these tracks we calculate transition points (TPs), where a vessel enters a cell, that is, when it passes between the starboard and port buoys of the respective cell. Positions before and after the ship enters the cell are linearly interpolated.

For each TP, we calculate the distance to the starboard buoy (d_s) at the crossing. This measurement simplifies the prediction, because the predicted position is indicated by means of just one value, instead of using the latitude and longitude. Figure 2 visualises the TP, as well as the d_s between the starboard and port buoys. We also determine the angle (β) at which the vessel moves from one cell to the next. As depicted in Figure 2, β is calculated using the TP, the position of the starboard buoy, and the first AIS position inside the new cell. We consider β to be important since, at cells where waterways cross, β may suggest which waterway a vessel will follow.

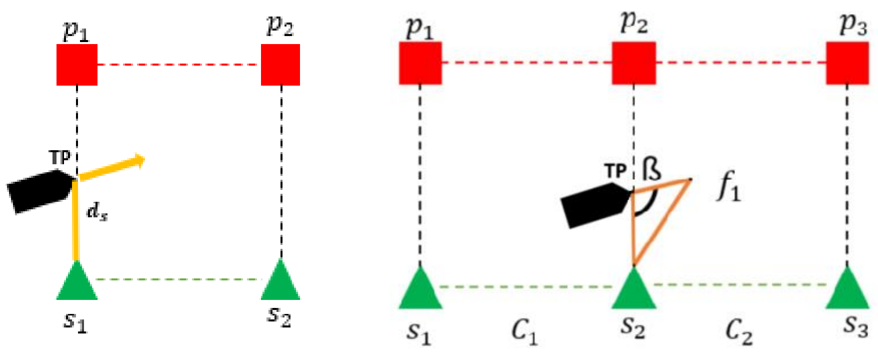

Figure 2. Visualisation of calculation of d_s and β

Furthermore, we have tested whether tide information improves the prediction accuracy using the tide information from the European Centre for Medium-Range Weather Forecasts (ECMWF) (Muis et al., 2022). This data comprises records of the water surface level from buoys located at specific positions along the coast and in rivers. The water surface levels from the two recording buoys at the beginning and end of the track are added to the track. These values are referred to as b_1 and b_2. The buoys used for this depend on the vessel's position, as will be shown in section 3.1.2. This data consists of records of the water level, recorded at ten-minute intervals. As the data collection frequency is different from AIS data, we have interpolated the water levels based on the time the vessel crosses the TP.

The influence of weather data on the prediction will also be investigated, the characteristics wind speed (w_s), wave height (w_h), and wind direction (w_d), extracted from the ERA5 dataset, also provided by ECMWF (Hersbach et al., 2023). The ERA5 dataset consists of latitude-longitude grids with a 0.25° x 0.25° resolution. We have used an algorithm to determine and select the data from the nearest grid point for each TP. The characteristics thereby obtained also had to be interpolated according to the time and position of the transition point.

Figure 3 summarises the process of generating TPs within waterways and adding tide data and weather information. In total, three different combinations of parameters have been tested. In the first approach, a dataset has been generated using only the information from the waterways in combination with AIS data. The second set also features the tide information, while the third set additionally includes the weather information.

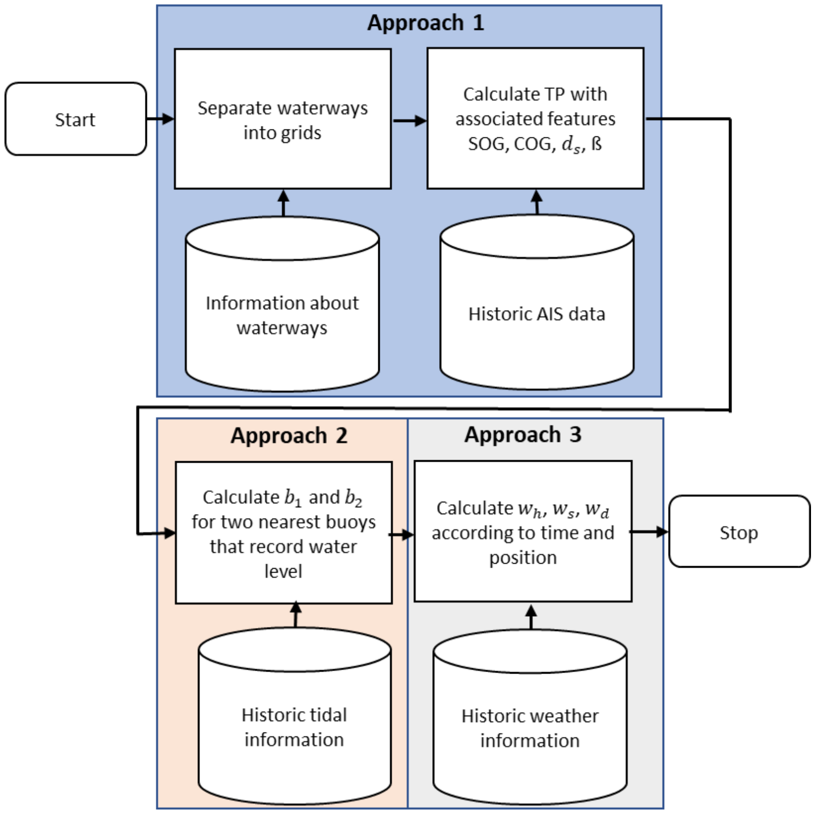

Figure 3. Summary of the process for generating TPs in waterways with tide and weather characteristics. The data generated from the preceding approaches is incorporated into the succeeding approaches.

3.1.2. Case Study Data from German North Sea Waterways

This paper uses commercially available AIS data, collected from terrestrial and satellite sources from January 1 2020 to April 16 2020 along the German Bight and the Elbe and Weser rivers. The region of interest is an area between 54.4° N, 10.7° E, 53.1° S, and 5.8° W. Figure 4 shows the waterways in this area and Figure 5 shows the AIS data. The AIS data has then been preprocessed by removing infeasible speed measures (2 kn < SOG < 30 kn), before generating the tracks.

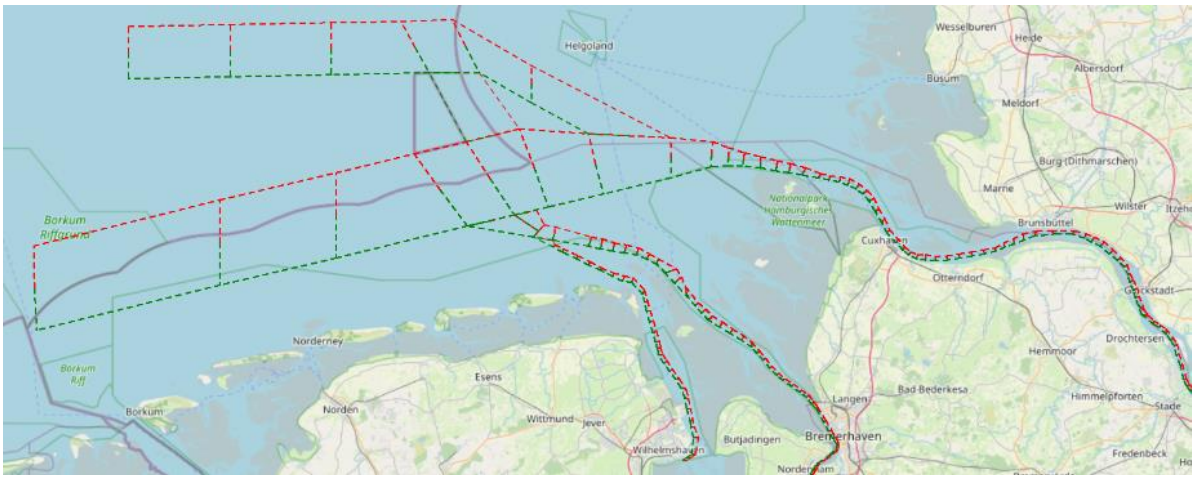

Figure 4. Waterways in the case study area

Figure 5. AIS data in the case study area from the year 2020

Regarding the tide information, Figure 6 displays the position of the buoys used in the area of interest. Depending on the position of the track, the measurements from the two nearest buoys are added to the AIS data. This means that the tracks along the Elbe get assigned the measuring levels of the buoys St. Pauli and Cuxhaven, and the tracks along the Weser the measuring levels of Bremen and the Alte Weser buoy. The tracks to or from Wilhelmshaven are mapped by the measuring levels of the buoys Wilhelmshaven and Wangerooge.

Figure 6. Buoys used for tide data (Bremen, Wilhelmshaven, Wangerooge, Alte Weser, Cuxhaven, St. Pauli)

In total, 11,167 tracks with at least 20 and up to 60 transition points are used. 2,877 tracks (26%) navigate along the Weser, 7,949 tracks (71%) along the Elbe, and the remaining 341 tracks (3%) navigate to or from Wilhelmshaven. This data is split into the training (72%), validation (8%), and the test (20%) datasets.

3.2. Machine Learning Models

LSTM models are widely used for predicting vessel tracks (Zhang et al., 2022). They are a type of recurrent neural network that uses gated units to selectively control the flow of information within the network. LSTMs consist of a memory cell and three types of gate: input gate, forget gate, and output gate (Hochreiter & Schmidhuber, 1997). The input gate regulates the flow of new information into the cell, the forget gate controls the flow of information out of the cell, and the output gate determines the information flow from the cell to the next hidden state. These gates allow selectively remembering or forgetting of information, as a result of which LSTMs are better equipped to model long-term dependencies in sequential data than the regular RNNs.

To extend the capabilities of LSTMs, Bi-LSTMs incorporate information from both past and future time steps of the input sequence. It is a type of bi-directional RNN (Schuster & Paliwal, 1997), and its architecture comprises two LSTM layers: one processing the sequence in a forward direction and the other in a backward direction. The outputs of these layers are concatenated to produce the final output, allowing access to information from both temporal directions and improving prediction accuracy compared to regular LSTM models.

LSTMs process input sequences sequentially, updating their internal state, one element at a time. This sequential nature limits parallelisation and can lead to increased computational complexity and training time, especially for long sequences. To address this limitation, Vaswani et al. (2017) propose the transformer model, which operates without any recurrence. The introduced transformer model employs a self-attention mechanism that enables efficient capturing of global dependencies by processing the entire input sequence in parallel. This attention mechanism calculates similarity scores for all pairs of positions in the sequence, allowing the model to learn deeper dependencies compared to LSTM models. The self-attention mechanism is combined with a feed-forward network, as well as normalisation layers, and is embedded in encoder and decoder layers. These components collectively enable the transformer model to effectively capture long-term dependencies, which makes it successful for time series predictions.

This work uses both Bi-LSTM and transformer models to predict vessel tracks inside waterways to compare these models. We use a Bi-LSTM model, where each LSTM has 128 units, followed by a dropout layer, and a dense layer that predicts possible future attributes. Overall, this model has at least 155,078 trainable parameters. This model will be compared with a transformer model that contains three transformer encoder layers, based on the architecture introduced by Vaswani et al. (2017). Using this architecture, the model contains at least 24,286 trainable parameters. This encoder layer is also followed by a dense layer predicting possible future attributes.

Ideally, the transformer model and the LSTM model should have a similar number of parameters. However, when we trained the models, a transformer model with a higher number of parameters was found to be overfitting. Therefore the numbers of parameters were optimised in order to obtain the highest possible prediction accuracy in the test dataset. The linear model, to which the Bi-LSTM and the transformer model are compared, consists of a simple dense layer with a linear activation function.

3.3. Anomaly Detection

In this paper we compare the predicted TP values against the real TP values to detect anomalous tracks. An anomaly in this context is a deviation from the prediction and the truth for one of the features SOG, COG, d_s, or β that exceeds a predefined threshold. The individual thresholds are determined for each feature based on a comparison between true and predicted values for the training data with which the model was trained. For this purpose, the standard deviation σ for the errors between the real and predicted values for the training data are calculated for each feature separately. The threshold for detecting anomalies can now be set, depending on the requirement.

To evaluate this concept, we have use a threshold of three standard deviations, since that includes 99.7% of all values as normal. This value is computed directly from the 99.7% quantile of the absolute error between the real and predicted values for the training data. One could also consider setting different thresholds for each variable. In a real-world application, setting the thresholds requires expert knowledge to ensure that these are meaningful (Kumpulainen, 2014). Additionally, the costs of false alarms and undetected anomalies have to be considered. It should be further emphasised that an anomaly in this paper represents only a deviation from the expected value. This does not mean that a detected deviation would represent a specific risk to vessel traffic within the predicted waterway.

4. RESULTS

4.1. Track Prediction

4.1.1. Comparison of Bi-LSTM and transformer models with different sets of input data

All models have been trained using the Adam optimiser and the Mean Square Error (MSE) to measure the prediction error. In this process, the models are trained to predict the next TP values based on the previous 10 TPs. These values are then used again with the previous nine TPs to predict the subsequent TP. Using this iterative approach, the five subsequent TPs have been predicted. The training was stopped whenever a model reached convergence, i.e. the point when additional training does not improve the model. This was determined using MSE calculated from the validation data. This metric is used in order to avoid a situation where the model is overfitting to the training data. However, local minima may exist. To be able to escape from a potential local minimum, the models are trained for ten additional epochs after a minimum has been reached.

To evaluate the influence of the tide and weather data, each model has been trained three times with different features based on the approaches shown in Figure 3. Considering these three approaches, transformer model training was stopped respectively after 30, 28, and 25 epochs, while Bi-LSTM models were stopped after 63, 38, and 81 epochs. Figure 7 shows MSE progression during training for Bi-LSTM and transformer models under approach 1. The MSE progressions with approaches 2 and 3 were not significantly different from the one shown in Figure 7. In all cases, transformer models converged faster than Bi-LSTM models. However, the accuracy of the Bi-LSTM models does not significantly improve over the training. Additionally, one must consider that our transformer models have a lower number of trainable parameters than our Bi-LSTM models.

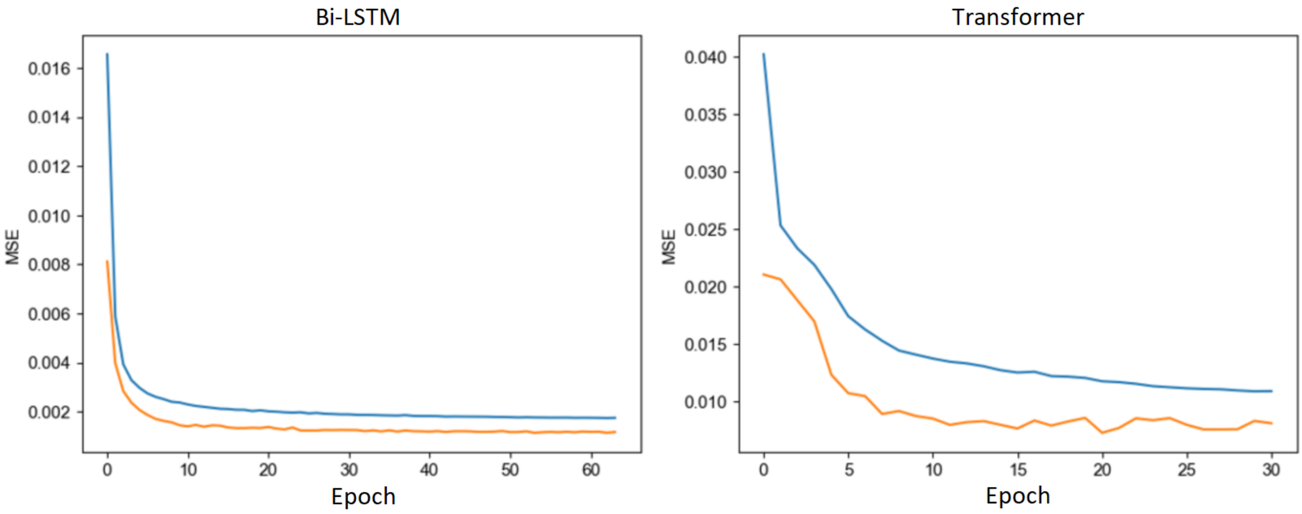

Figure 7. Training and validation errors for Bi-LSTM and transformer model with approach 1. Blue curves show the training error, and orange ones the validation error. Training is stopped ten epochs after the convergence.

Table 1 shows the average error for predicting the next five TPs compared to the ground truth for different sets of data, using both Bi-LSTM and transformer models. The errors are calculated using the test dataset.

Input Data	Model	d_s (m)	SOG (kn)	COG (°)	β (°)
	Linear	178.09	0.74	9.97	18.93
Approach 1 SOG, COG, d_s, β	Bi-LSTM	106.42	0.68	2.07	1.85
	Transformer	182.75	2.73	5.24	7.09
Approach 2 b_1, b_2	Bi-LSTM	197.46	1.02	2.16	2.03
	Transformer	171.21	2.75	5.06	5.95
Approach 3 w_h, w_s, w_d	Bi-LSTM	153.97	1.25	2.11	2.18
	Transformer	184.67	2.79	7.4	7.64

Table 1. Average prediction error of the five subsequent TPs compared to the ground truth. d_s is measured in meters (m), SOG is measured in knots (kn), while COG and β are measured in degrees (°). The linear model has been trained for each feature individually.

When considering d_s, the prediction of the Bi-LSTM model in the first approach achieves an average prediction error of 106.42 m. This result is the best one, and about 65 m more accurate than the best result achieved with a transformer model. This result has been achieved in approach 2, with tide data being added. The linear model achieves an average prediction error of 178.09 m.

When considering SOG, the best result is also obtained from the Bi-LSTM in the first approach. It is noticeable that the result does not differ much from that of the linear prediction, which reaches a deviation of 0.74 kn. The transformer model, on the other hand, delivers a higher prediction error at 2.72 kn. The error for the transformer model is similar, regardless of the data.

For COG and the β, the Bi-LSTM achieves 2.07 and 1.85° respectively in the first approach. These are by far the lowest prediction errors for these features. The Bi-LSTM model is followed by the transformer model, which achieves an average prediction error of 5.06° for COG and 5.95° for β, in the second approach. The largest deviation, with 9.97° for COG and 18.93° for β, is obtained with the linear prediction.

It is notable that the prediction results of the Bi-LSTM model become worse when more data features are added. However, the prediction results of the transformer model improved when tidal data was added. In both cases, when weather data was added, the results got worse.

Figure 8 shows the range of errors for each prediction step and for each feature, as well as the average, and the 99.7% quantile of the prediction errors from the Bi-LSTM with the first approach, the transformer model with the second approach, and the linear prediction. As can be seen, the average and the 99.7% quantile of the prediction errors increases consistently when predicting the later transition points for the features SOG and COG. However, this is not the case for the features d_s and β. This is of interest since the predicted errors for the first steps are included in the predictions of the last steps. Moreover, the outliers in the prediction of d_s, COG, and β for the Bi-LSTM model become lower in later predictions, as well as in the prediction of d_s and SOG with the transformer model. For the features d_s and SOG, the transformer model predicts the largest error range, with an error for d_s of up to 2,700 m relative to ground truth and up to 11 kn for SOG. Remarkably, the 99.7% quantile of the error for β under the transformer model increases significantly in the third step and then decreases again. When considering COG and β, the linear prediction performs the worst, apart from the third step, predicting large outliers of up to 120° for COG and 100° for β.

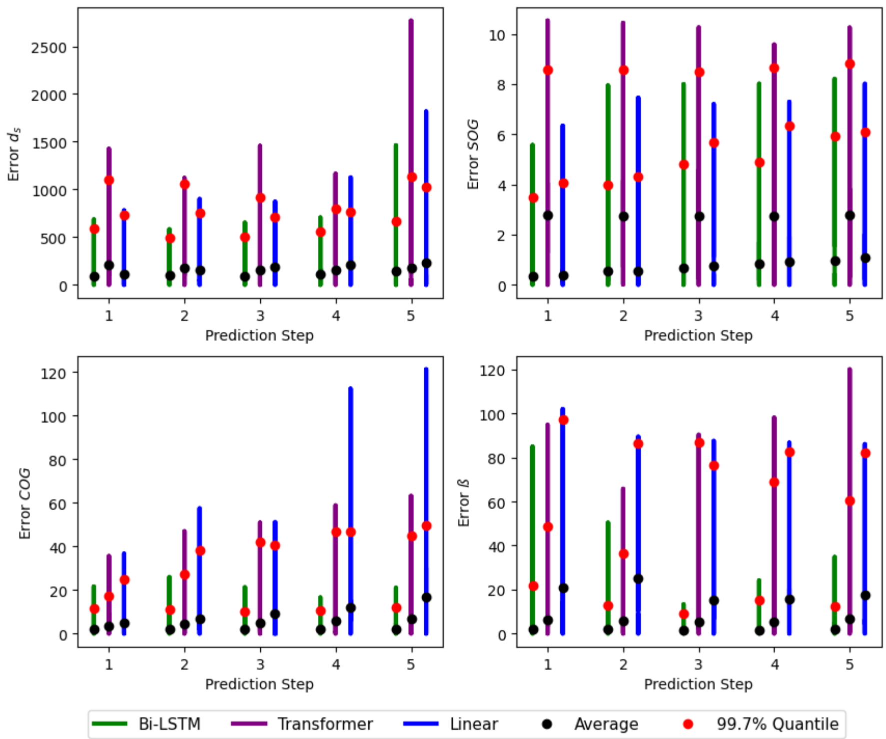

Figure 8. Range of prediction errors for the linear model, the Bi-LSTM model with the first approach, and the transformer model with the second approach

As in Figure 8, the prediction errors are shown in Figure 9 with the performance measures RMSE, Mean Square Error (MSE), and Mean Absolute Error (MAE) for the Bi-LSTM with the first approach, the transformer model with the second approach, and the linear prediction. The green lines show that the Bi-LSTM model has the lowest performance values for all four features d_s, SOG, COG, and β but the RMSE, MSE, and MAE for SOG do not differ greatly between the linear and Bi-LSTM model. The transformer model has lower performance values than the linear model for COG, and β. For d_s, the performance values from the transformer model are lower for the prediction steps from three to five compared to the linear model, but for the first two steps the values are higher for the transformer model. The transformer model has the largest prediction errors for SOG, compared to the Bi-LSTM and linear models.

To further evaluate the results, we divided the predictions from this model according to the waterways along the Elbe and Weser rivers and Wilhelmshaven. The prediction errors for these waterways can be seen in Table 2.

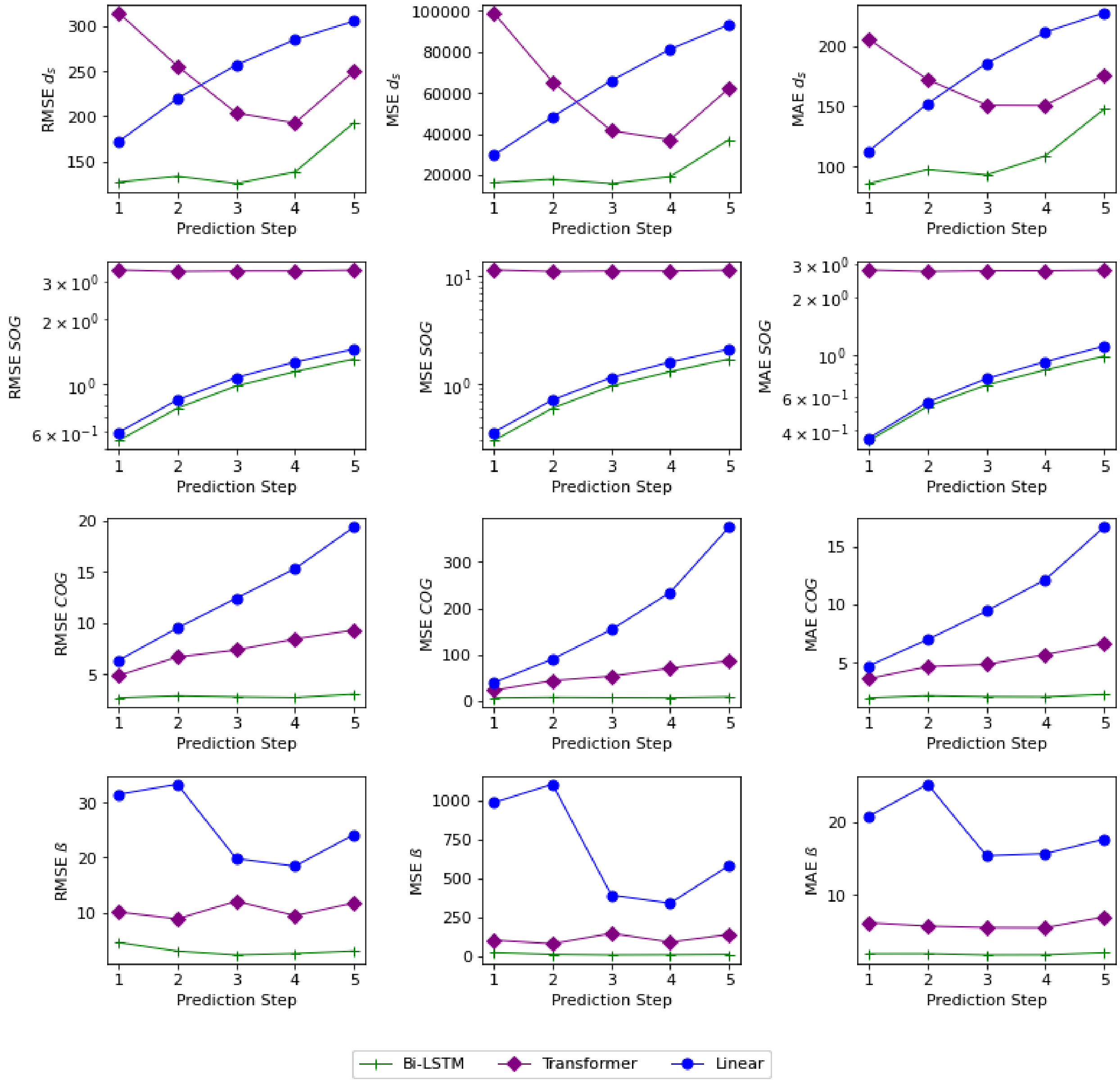

Figure 9. RMSE, MSE, and MAE of the linear model, the Bi-LSTM model from the first approach, and the transformer model from the second approach

Waterway	d_s (m)	SOG (kn)	COG (°)	β (°)
Elbe	100.57	0.65	2.17	1.87
Weser	115.15	0.71	1.90	1.76
Wilhelmshaven	105.15	0.82	2.22	2.64

Table 2. Average prediction errors in different waterways

The results show that the lowest average prediction error for d_s is achieved along the Elbe river, followed by Wilhelmshaven and the Weser waterways. For SOG, the lowest error is again obtained along the Elbe. However, the difference between the best and the worst results applicable to Wilhelmshaven is less than 0.2 kn. For the other features COG and β, the lowest errors are achieved along the Weser river. The prediction results for these two features are with less than a one-degree difference and their error values are close to each other. The fact that predictions to and from Wilhelmshaven are slightly worse than the others can be explained by the scarcity of data. Only 3% of our data came from this waterway. However, since the overall results of the predictions for the different waterways are close, it can be concluded that the concept presented here for predicting vessel tracks is generalisable for the trained waterways.

4.1.2. Assessment of prediction errors

When considering the range of prediction errors, as displayed in Figure 8 for d_s, the transformer model predicts a wider range of errors than the Bi-LSTM model. These outliers in the transformer model's prediction of the d_s occur for tracks where the waterway becomes significantly wider. This can be seen in Figure 10, where the d_s of the fifth prediction is 2,766 m away from the actual position. In this case, the Bi-LSTM makes a much more accurate prediction, although it also predicts an error of 263 m. In the example at hand, the vessel was sailing far to the left in the waterway, which is unusual, since it was assumed that vessels sail as far to the right as possible. This deviation from the assumption of the vessel's behaviour may also contribute to the large difference between the predicted distance and the actual distance. Accordingly, the assumption needs to be reconsidered, as it may not hold for all vessel types considered in the data.

The outliers in predicting β and COG occur mainly for one location in the waterway along the Elbe river, which is displayed in Figure 11. The waterway runs along Glückstadt and the buoys are arranged in such a way that the vessel seems not to be passing through the waterway from the north or south, but from the west or east. Since the arrangement of the buoys in the course of the waterway only gives that impression at this particular intersection, the models are unable to represent this transition well. However, had the model actually predicted this outlier, this would have indicated that the model had overfitted and made predictions too close to the training data.

In the case of the COG, it is particularly interesting that the linear predictions show significantly greater deviations in the later predictions than the other two models. This is also accompanied by a lower average prediction accuracy. However, the 99.7% quantile error is slightly higher for the transformer model. This illustrates that the linear prediction and the transformer model do not provide good results for this feature and that the Bi-LSTM model can predict the values significantly more accurately.

It is noticeable that all predictions can produce erroneous results. As shown in Figure 12, the Bi-LSTM model predicts values outside the waterway. This phenomenon occurs during the first approach in the Bi-LSTM model in 0.7% of all test predictions, and in the predictions made with the transformer model in 0.4% of all test predictions. The linear prediction only produces this error in 0.4% of all test predictions. Erroneous predictions mainly occur in curved passages, where the distance to the previous transition points becomes significantly smaller. With the introduction of tide information in the Bi-LSTM model, 36% of all tracks were predicted erroneously, which indicates that with the addition of these features, the model becomes distracted in the prediction and can no longer represent the dependencies between the features, as well as in the absence of any tide information. This can also be seen from the fact that the average error in the prediction of d_s increases by almost 90 m. However, the added features seem to help the transformer model, because all predicted features, except SOG, have become more precise.

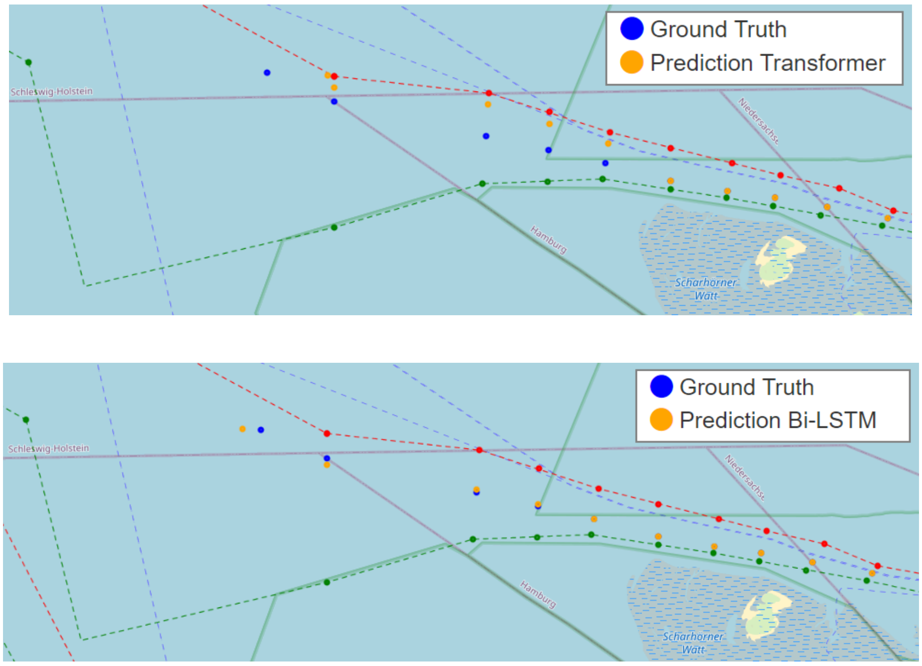

Figure 10. Examples of d_s outliers

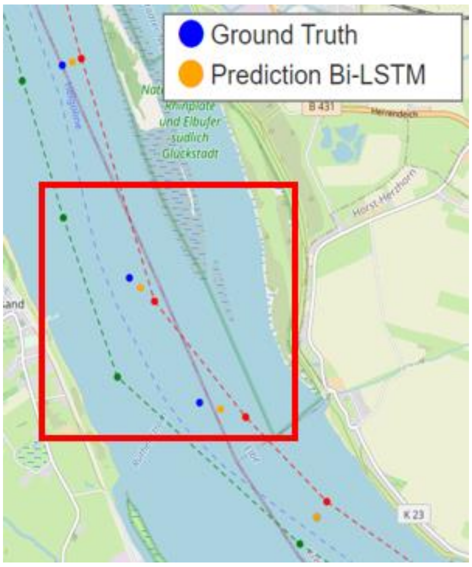

Figure 11. Section of the waterway where outliers of COG & β are predicted due to the arrangement of the starboard and port buoys. Although the waterway leads north and south, the arrangement of the buoys indicates a more westerly and easterly course, respectively

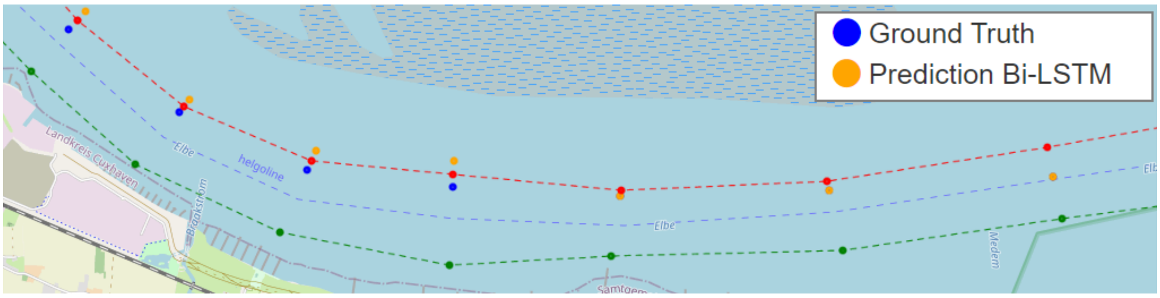

Figure 12. Erroneous track predictions outside waterways

4.2. Anomaly Detection

To evaluate the presented concept for anomaly detection, the 99.7% quantile of the prediction error is calculated for the all features from the training data between the real and predicted values. As previously mentioned, we use a 99.7% quantile, which corresponds to a threshold value of three σ. For individual parameters, this threshold represents the following deviations: d_s 576.42 m, SOG 4.64 kn, COG 11.29°, and β 13.16°.

To test the anomaly detection approach, we have used the best performing Bi-LSTM model to predict tracks using data from July 28 to August 6 2020. In total, predictions about the next five subsequent TPs are made for 932 tracks, of which 242 navigate along the Weser river, 643 along the Elbe river, and 47 navigate to and from Wilhelmshaven. The predicted features have been compared to the ground truths to ascertain whether the defined threshold for that feature has been exceeded. In such cases, the track has been flagged.

In certain cases, the model can predict a negative distance to the starboard buoy. Since this would mean that the vessel is sailing outside the fairway, these predictions are erroneous. For these cases, $d_s = 0$ is assumed instead of the negative distance.

The results of the marked transition points are shown in Table 3. 4.6% of the total tracks are marked by a deviation at β, 3.2% for COG, 2.8% for d_s, and 1.4% for SOG. This shows that the numbers of marked tracks are slightly different for each feature under consideration. The marked tracks are then once more divided into the individual waterways considered, whereby it is noticeable that, when considering d_s, none of the tracks along Wilhelmshaven are marked. For SOG, tracks are marked for Wilhelmshaven with 4.3%, 0.9% for the Elbe, and 2.1% along the Weser. When considering COG and β, tracks along the Weser are marked with 5.8% for COG and 10.7% for β, and for tracks along Wilhelmshaven 10.6% for COG and 12.8% for β. This is much higher than the tracks marked along the Elbe, which are only 1.7%.

	Overall	d_s > 576.42	%	SOG > 4.64	%	COG > 11.29	%	β > 13.16	%
Weser	242	5	2.1	5	2.1	14	5.8	26	10.7
Elbe	643	21	3.3	6	0.9	11	1.7	11	1.7
WHV	47	0	0	2	4.3	5	10.6	6	12.8
Overall	932	26	2.8	13	1.4	30	3.2	43	4.6

Table 3 Anomaly detection results

To illustrate some of the detected anomalies, Figure 13 shows two examples. In the anomaly shown on the left, the threshold value d_s is exceeded. The vessel changes its side in the waterway to the left, while it is predicted to remain on the right side. In the second anomaly, the threshold β was exceeded, as the vessel changes here first to the left and then back to the right side. The prediction was that the vessel would have remained on the right-hand side.

Overall, the marked tracks would need to be investigated further to determine whether the anomalies detected were actual deviations to the vessel's usual route or whether they represented unrealistic predictions made by the model. However, this is outside the scope of this paper. Nevertheless, the assumption can be made that these marked tracks are related to the outliers of the predictions mentioned in the previous section.

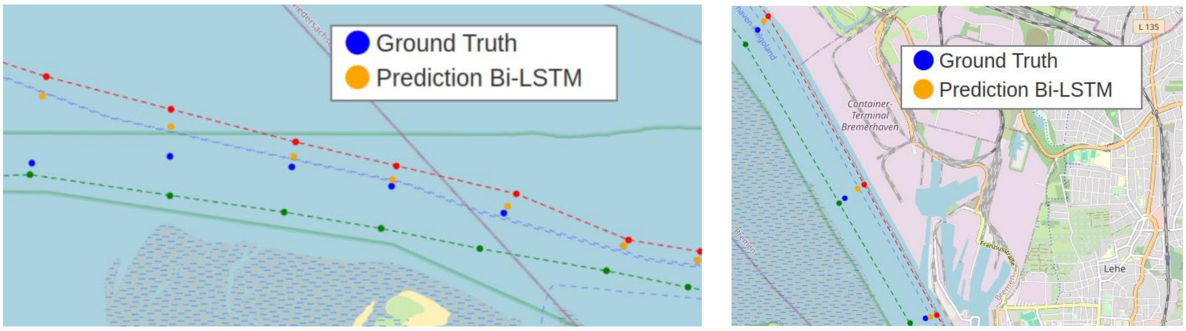

Figure 13. Anomalies for track prediction for parameter d_s (on the left side) and β (on the right side)

5. DISCUSSION AND CONCLUSIONS

In this paper models for predicting Transition Points (TP), representing vessel tracks in waterways, have been developed and tested. We have used historical vessel positions in the form of AIS data, weather information, and tide data. Positions of buoys have been extracted from sea chart information and have also been combined with AIS data to create TPs. The data resulting from this concept is then used to train a Bi-LSTM and a transformer model. These models have been used to predict the subsequent transition points, representing the track a vessel is expected to take.

The model that predicted the most accurate vessel tracks is a Bi-LSTM model trained without tide and weather information, focusing only on AIS data combined with positions of buoys that delimit waterways. Also, the transformer model has predicted a lower average error than the linear prediction for COG and β, as well as d_s for the prediction steps three to five. However, unlike the Bi-LSTM model, inclusion of tide data into the transformer model has improved the prediction, while the weather data has still decreased the accuracy.

Interestingly, Murray et al. 2023 attained somewhat similar results. Their anomaly detection approach worked best when only considering the position data. Models trained with the additional features SOG and COG could only detect parts of the anomalous trajectories. We believe that these results are caused by a phenomenon where adding data with a low information content to train a machine learning model may introduce noise or bias into that model (John et al., 1994), (Kuhn & Johnson, 2019).

Considering Bi-LSTM and transformer model, during the experiments, the transformer model overfitted as soon as more layers were added. Therefore, we have opted to use a transformer model which has fewer trainable parameters than the Bi-LSTM model. Future work should experiment with different architectures, which may prevent this phenomenon. Solving it would improve predictions. In addition, the concept developed here could also be used for specific vessel types, which would help in assessing the method for real-world use.

Furthermore, this paper has developed a concept for detecting anomalous vessel tracks. For this purpose, the prediction of the dataset used to train the model has been compared to the truth value and three standard deviations of the error have been set as the threshold for anomalous tracks, which corresponds to the 99.7% quantile of the error. In addition, another dataset has been processed in the same region, but at a different time. Having made the most accurate predictions, the Bi-LSTM model has been chosen to predict the tracks. As soon as the deviations were greater than the set threshold, they were flagged as anomalous tracks. In summary, this method can detect anomalous tracks by comparing predictions with actual measurements.

CONFLICT OF INTEREST

Authors declare no competing interests.

ACKNOWLEDGEMENTS

This paper uses data from Copernicus Climate Change Service information (accessed 2022). Neither the European Commission nor ECMWF is responsible for any use that may be made of the Copernicus information or data it contains.

REFERENCES

Ester, M., et al., 1996. A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. KDD'96: Proceedings of the Second International Conference on Knowledge Discovery and Data Mining, pp. 226 - 231.

Forti, N., et al., 2020. Prediction of Vessel Trajectories From AIS Data Via Sequence-To-Sequence Recurrent Neural Networks. IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP 2020, pp. 8936-8940. Available at: https://doi.org/10.1109/ICASSP40776.2020.9054421.

Guo, S. et al., 2023. Identification of Hazardous Encounter Scenarios Using AIS Data for Collision Avoidance System Testing. In: H. Le Sourne & C. Guedes Soares, eds. Advances in the Collision and Grounding of Ships and Offshore Structures. s.l.:CRC Press, pp. 43-50. Available at: https://doi.org/10.1201/9781003462170-8.

Hersbach, H. et al., 2023. ERA5 hourly data on single levels from 1940 to present, s.l.: Copernicus Climate Change Service (C3S) Climate Data Store (CDS). Available at: https://doi.org/10.24381/cds.adbb2d47.

Hochreiter, S. and Schmidhuber, J., 1997. Long Short-Term Memory. Neural Computation,9, pp. 1735–1780. Available at: https://doi.org/10.1162/neco.1997.9.8.1735.

IALA, 2017. The IALA maritime buoyage system, R1001, s.l.: International Association of Lighthouse Authorities.

John, G. H., Kohavi, R. and Pfleger, K., 1994. Irrelevant Features and the Subset Selection Problem. Machine Learning Proceedings 1994, pp. 121-129. Available at: https://doi.org/10.1016/B978-1-55860-335-6.50023-4.

Kuhn, M. and Johnson, K., 2019. Feature Engineering and Selection, A Practical Approach for Predictive Models. 1st ed. s.l.:CRC Press. Available at: https://doi.org/10.1201/9781315108230.

Kumpulainen, P., 2014. Anomaly detection for communication network monitoring. s.l.:Tampere University of Technology.

Li, H., Jiao, H. and Yang, Z., 2023. AIS data-driven ship trajectory prediction modelling and analysis based on machine learning and deep learning methods. Transportation Research Part E: Logistics and Transportation Review, 175, pp. 103152. Available at: https://doi.org/10.1016/j.tre.2023.103152.

Liu, C. et al., 2021. TPR-DTVN: A Routing Algorithm in Delay Tolerant Vessel Network Based on Long-Term Trajectory Prediction. Wireless Communications and Mobile Computing, 2021, pp. 15. Available at: https://doi.org/10.1155/2021/6630265.

Löwenstrom, J., Solano-Carrillo, E. and Stoppe, J., 2022. Learning Representative Vessel Trajectories Using Behavioral Cloning. Proceedings of the MARESEC 2022. Available at: https://doi.org/10.5281/zenodo.7143586.

Mehri, S., Alesheikh, A. A. and Basiri, A., 2021. A Contextual Hybrid Model for Vessel Movement Prediction. IEEE Access, 9, pp. 45600-45613. Available at: https://doi.org/10.1109/ACCESS.2021.3066463.

Millefiori, L. M. et al., 2016. Modeling vessel kinematics using a stochastic mean-reverting process for long-term prediction. IEEE Transactions on Aerospace and Electronic Systems, 52, pp. 2313–2330. Available at: https://doi.org/10.1109/TAES.2016.150596.

Muis, S. et al., 2022. Global sea level change time series from 1950 to 2050 derived from reanalysis and high resolution CMIP6 climate projections, s.l.: Copernicus Climate Change Service (C3S) Climate Data Store (CDS). Available at: https://doi.org/10.24381/cds.a6d42d60.

Murray, B., et al., 2023. Autoencoder-Based Anomaly Detection for Safe Autonomous Ship Operations. Proceedings of the 33rd European Safety and Reliability Conference (ESREL 2023), pp. 2885-2892. Available at: https://doi.org/10.3850/978-981-18-8071-1_P403-cd.

Nguyen, D. & Fablet, R., 2021. TrAISformer - A generative transformer for AIS trajectory prediction, s.l.: arXiv. Available at: https://doi.org/10.48550/arXiv.2109.03958.

Nguyen, D., Le Van, C. and Ali, M. I., 2018. Vessel Trajectory Prediction using Sequence-to-Sequence Models over Spatial Grid. DEBS '18: Proceedings of the 12th ACM International Conference on Distributed and Event-based Systems, pp. 258–261. Available at: https://doi.org/10.1145/3210284.3219775.

Pallotta, G. & Jousselme, A., 2015. Data-driven detection and context-based classification of maritime anomalies. 18th International Conference on Information Fusion, Fusion 2015, p. 1152–1159.

Pallotta, G., Vespe, M. and Bryan, K., 2013. Vessel Pattern Knowledge Discovery from AIS Data: A Framework for Anomaly Detection and Route Prediction. Entropy, 15, pp. 2218–2245. Available at: https://doi.org/10.3390/e15062218.

Ristic, B., 2014. Detecting Anomalies from a Multitarget Tracking Output. IEEE Transactions on Aerospace and Electronic Systems, 50, pp. 798–803. Available at: https://doi.org/10.1109/TAES.2013.130377.

Schuster, M. & Paliwal, K. K., 1997. Bidirectional recurrent neural networks. IEEE Transactions on Signal Processing, 45, pp. 2673 - 2681. Available at: https://doi.org/10.1109/78.650093.

Sekhon, J. & Fleming, C., 2020. A Spatially and Temporally Attentive Joint Trajectory Prediction Framework for Modeling Vessel Intent. Proceedings of Machine Learning Research, 120, pp. 318-327.

Stach, T., et al., 2023. Maritime Anomaly Detection for Vessel Traffic Services: A Survey. Journal of Marine Science and Engineering, 11, pp. 1174. Available at: https://doi.org/10.3390/jmse11061174.

Steidel, M., Mentjes, J. and Hahn, A., 2020. Context-Sensitive Prediction of Vessel Behavior. Journal of Marine Science and Engineering, 8, pp. 987. Available at: https://doi.org/10.3390/jmse8120987.

Stróżyna, M. et al., 2022. Data analysis in the maritime domain. s.l.:Poznań University of Economics and Business Press. Available at: https://doi.org/10.18559/978-83-8211-137-8.

Szymanska, Z. and Murray, M., 2022. Maersk container ship runs aground off German island. Reuters, 3 February.

UNCTAD, 2021. Review of Maritime Transport. s.l.:United Nations Publications.

Vaswani, A. et al., 2017. Attention is All you Need. 31st Conference on Neural Information Processing Systems (NIPS 2017), pp. 6000-6010. Available at: https://doi.org/10.5555/3295222.3295349.

Venskus, J., Treigys, P. and Markevičiūtė, J., 2021. Unsupervised marine vessel trajectory prediction using LSTM network and wild bootstrapping techniques. Nonlinear Analysis: Modelling and Control, 26, pp. 718-737. Available at: https://doi.org/10.15388/namc.2021.26.23056.

Voytenko, M., 2023. General cargo ship collided with wind turbine, damaged, North sea. Maritime Bulletin, 26 April.

Yan, R. and Wang, S., 2019. Study of Data-Driven Methods for Vessel Anomaly Detection Based on AIS Data. Smart Transportation Systems 2019, pp. 29–37. Available at: https://doi.org/10.1007/978-981-13-8683-1_4.

Zhang, S. et al., 2021. A Bi-directional LSTM Ship Trajectory Prediction Method based on Attention Mechanism. 5th IEEE Advanced Information Technology, Electronic and Automation Control Conference, IAEAC 2021, pp. 1987–1993. Available at: https://doi.org/10.1109/IAEAC50856.2021.9391059.

Zhang, X. et al., 2022. Vessel Trajectory Prediction in Maritime Transportation: Current Approaches and Beyond. IEEE Transactions on Intelligent Transportation Systems, 23, pp. 19980-19998. Available at: https://doi.org/10.1109/TITS.2022.3192574.

Zhang, Y. et al. 2024. Vessel Behavior Anomaly Detection Using Graph Attention Network, in Neural Information Processing, 30th International Conference, ICONIP 2023 Proceedings, Part V, pp. 291 – 304. Available at: https://doi.org/10.1007/978-981-99-8073-4_23.

Zou S. et al., 2023. A Dynamic Spatio-Temporal Refinement Network for Multi-Vessel Trajectory Prediction, in the Proceedings of the 2023 International Conference on Computer, Vision and Intelligent Technology (ICCVIT '23). Available at: https://doi.org/10.1145/3627341.3630416.