An investigation into computational methods for classifying fishing vessels to identify illegal, unreported and unregulated fishing activity

Benedict Grey, Lahcen Ouarbya and Tim Blackwell
Department of Computing
Goldsmiths, University of London
London, United Kingdom
bgrey001@gold.ac.uk; l.ouarbya@gold.ac.uk; t.blackwell@gold.ac.uk

***Abstract*—Illegal, unreported and unregulated (IUU) fishing undermines collective efforts to create a global model for sustainable fishing. Countering IUU fishing is an urgent priority given world population growth and increasing dependence on ocean-sourced food. This paper examines deep learning methods for the classification of fishing vessels with the intent to determine illicit fishing operations. This is achieved through supervised learning with highly irregular time series data in the form of signals from the automatic identification system (AIS). To deal with the intermittent frequency of AIS signals, two separate approaches have been followed: feature engineering with zero padding and linear interpolation. Fundamentally, this work suggests there exists a distinct relationship between vessel movement patterns and method of fishing. Two neural network architectures: stacked bidirectional GRUs and 1D CNNs with residual connection blocks, are leveraged on each data pipeline to produce four sets of results. The GRU with feature engineering achieves 95% accuracy despite severe class imbalance in the large datasets. The system can classify a vessel's fishing method over 24 hours in real-time to monitor behaviour in marine protected areas and detect gear discrepancies, safeguarding fish stocks in the process.**

***Keywords*—Deep learning, machine learning, data mining, supervised learning, automatic identification system, time series data, feature engineering, illegal fishing**

I. Introduction

Global demand for protein as a source of sustenance is increasing rapidly. The fishing industry is expanding to meet this growing demand. With this expansion, illegal, unreported and unregulated (IUU) fishing is catastrophically mitigating attempts to create a global model for sustainable fishing [1]. Consequently, global food security is at an ever increasing level of risk. Research suggests that IUU fishing accounts for between 20% to 35% of wild caught seafood [2]. The resultant economic loss and environmental damage [3], coupled with the threat to global food security make the prevention of IUU fishing a pressing environmental and economic matter.

This article is an examination of deep learning methods for the classification of fishing vessels through supervised learning with irregularly sampled, multivariate time series data, with the intent to determine illicit fishing operations. The stochastic nature of the data generated by the automatic identification system (AIS) leads to the first of two critical problems to be solved, processing irregularly sampled time series data. The second, is proving there exists a distinct relationship between the movement pattern of a vessel and it's method of fishing.

Thus, the core of the proposed system is identifying patterns in temporal geospatial AIS data. Ships that are engaged on international voyages and have a gross tonnage of 300 or more, as well as cargo ships not engaged on international voyages and passenger ships of any size, with a gross tonnage of 500 or more, are required to fit an AIS transponder [4]. The transponder broadcasts messages that contain the vessel's maritime mobile service identity (MMSI), GPS coordinates, course and speed. These messages are received terrestrially and by satellite. The data for this research project has been provided by the Global Fishing Watch (GFW), an international, independent non-profit organisation. AIS signals are transmitted randomly within varying time intervals depending on speed or course. The sampling frequency of the messages is numerous and irregular. Due to the large volume of data generated by the AIS, a clear path has emerged in the form of deep learning for the task at hand, deep learning models being most effective when trained on large quantities of data.

Fishing vessels are classified based on fishing method according to six classes: drifting longlines, fixed gear, pole and line, purse seines, trawlers and trollers. By categorising vessels into one of the proposed classes based solely on AIS data collected over a 24-hour period, the system can be used to identify fishing behaviour with a finite quantity of data. As a result, the technology can be leveraged in real-time by fishing authorities to locate suspicious fishing activity in maritime protected areas or locate discrepancies between registered fishing geartype and the model's observed geartype.

For the classification process to be successful, a deep learning model proven to be effective at learning from time series data must be selected. This proposal has therefore leveraged two neural network models against each other. The natural choice for modelling time series data would be a recurrent neural network (RNN) due to its ability to retain memory within sequences. However, a standard RNN has not been chosen due to the exploding/vanishing gradient problem. To combat this, new gated RNN models have been developed such as gated recurrent units (GRU) and long short-term memory (LSTM), the former of which has been selected as the first model. The second chosen model is a convolutional neural network (CNN) since recent research demonstrates they are effective at modelling long sequence data [5].

Before deep learning models are trained, an important issue must be addressed: how the nature of the irregularly sampled time series data affects the learning process. Established methods for handling temporal data tend to deal only with regularly sampled time series data. The proposed system using a CNN or GRU would assume the data is uniformly sampled, causing potentially erroneous results. The following techniques have been explored for handling this problem: zero padding with featurised time differences and linear interpolation.

Figure 1 illustrates the following experimental procedure: the first layer is the data preprocessing module, with different functionality depending on the branch i.e. padding with zeros or linear interpolation. For the branch of data with padded varying sequences, the time difference between observations is featurised. The second layer shows the comparison of the CNN against the GRU on both the interpolated and zero padded data. The final layer is evaluation, a comparison of results from each combination of diverging methods to identify the strongest overall model.

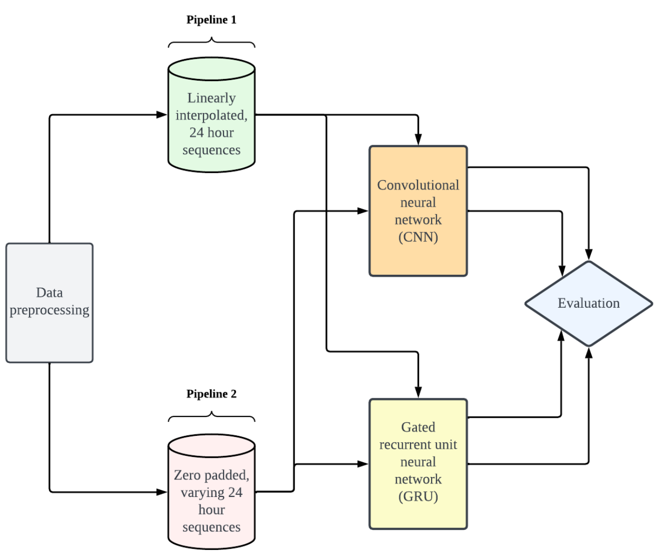


Fig. 1: The proposed system.

At the evaluation stage, analysis is conducted on the developed models using universally accepted metrics to determine effectiveness in a machine learning context. The construction of a confusion matrix allows accuracy, precision, recall and F1-score to be derived. By using these metrics, the most suitable model can be chosen from the branching system. Furthermore, the prevailing model can be compared with existing benchmark performances by competing architectures in the categories of fishing vessel classification and irregular time series modelling.

II. Related work

A number of research papers use trajectory data to classify the behaviour of vessels at sea. By identifying certain patterns in the data, inferences can be made about potentially illicit behaviour. Within the illegal fishing subset of research, several papers are examined. Kim and Lee [6] propose a method using AIS data collected rigorously over a year following 1380 fishing ships around Jeju island in South Korea. Their system uses a one dimensional convolutional neural network (1D CNN) based model incorporated with a fully connected network that utilises environmental data to aid the classification process. This technique is a novel, robust and seemingly accurate system for the classification of fishing geartypes. Kim and Lee [6] introduce two new evaluation metrics to measure their success in the form of a day-wise performance index (DPI) and trajectory window-wise performance metric (TPI). By comparing results obtained from their model in comparison with a support vector machine (SVM), a tried and testing machine learning algorithm supported by Lahmiri [7], using the TPI and DPI metrics, they produce average scores of 90.1% vs 76.9% and 96.3% vs 81.4% respectively. However, lacking from the research is the implementation of a universal metric such as the previously mentioned classification accuracy, precision, recall or F1-score. This would have provided more context for determining the efficacy of their model.

Kalaiselvi et al. [8] use a similar approach, introducing 1D CNNs into their system for classifying the data collected and labelled by the GFW. Their approach appears to be less rigorous, lacking environmental data compared to the model Kim and Lee use. However, considering Kalaiselvi et al. [8] use data from the GFW, which was not available at the time of Kim and Lee's [6] work, their system may represent a more global model of fishing vessel characteristics. In addition, Kalaiselvi et al. [8] use universal performance metrics: accuracy, precision, recall and F1-score. Producing impressive average scores of 95%, 89.75%, 94.5% and 88.5% respectively. Both Kalaiselvi et al. [8] and Kim and Lee's [6] papers are highly relevant.

Marzuki et al. [9] are referenced in both Kim and Lee [6] and Kalaiselvi et al.'s [8] work. Marzuki et al. [9] were early to adopt trajectory data for the task of classification in order to identify fishing geartype and consequently infer illicit behaviour from vessels. Marzuki et al. [9] use vessel monitoring system (VMS) data, which Kim and Lee [6] highlight has a problem: scarce signal frequency. This makes accurate inductions based on movement patterns difficult. The differences continue as Marzuki et al. [9] use SVMs as their model for classification, a traditional machine learning technique. Mean correct classification rate is their chosen metric for success, the same as previously mentioned classification accuracy, presenting their highest score of 97.6% accuracy.

Sánchez Pedroche et al. [10] also employ an SVM model as well as decision trees to determine fishing and non fishing behaviour, but with a key difference: they utilise AIS data. Sánchez Pedroche et al. [10] use a combination of accuracy and F1-score, exhibiting at their highest for decision trees 69.35% and 60.96% and for SVMs 75.3% and 72.62%. Although the data used is different to Marzuki et al. [9], the desired outcome of the paper is the same, to find patterns in IUU fishing to aid the protection of global food security and maritime biodiversity.

Kim and Lee [6] and Kalaiselvi et al. [8], as mentioned, have both used 1D CNNs for their classification tasks within the category of identifying IUU fishing. However Chen et al. [11], although not specifically aiming to identify illicit fishing behaviour, use 2D CNNs with AIS data for classification. This is an example of the increased usage of CNNs across fields, also supported by Villaruz [12] and Bunrit et al. [13]. Chen et al. [11] implement a novel approach that renders images of trajectory patterns displayed by all vessels using AIS transponders in order to classify static, normal navigation or manoeuvring. By creating images of the movement patterns through a process called ship movement image generation and labelling, Chen et al. [11] are claiming to harness the full ability of the CNN. This technique bypasses the issue of irregularly sampled time sequences due to the fact that each projection is taken from a set period of time, culminating in one image that serves as a single training example for their 2D CNN. Chen et al. [11] use familiar performance metrics. Accuracy is measured at 77.66%, precision at 92.55%, recall at 61.96% and F1-score at 76.38%. The application of CNNs for the task at hand is justified by promising results. This is supported by Li et al. [5] in their research paper on modelling long sequences with CNNs, proving they are competent and time efficient, with classification accuracy of 84.4% and robust performance across the board in the Long Range Arena [14].

A potential issue with Kim and Lee [6] and Kalaiselvi et al.'s [8] approach is the use of linear interpolation for regularising sample frequency in the AIS data. This creates inaccuracies, as noise is introduced to the dataset in the form of erroneous data points for the trajectory of the fishing vessels. The same problem applies to Marzuki et al. [9] due to their use of VMS trajectory data, the sparsity of the broadcast messages results in low spatial resolution, while the irregularity requires approximation, introducing noise. Therefore it is necessary to examine the existing architectures for handling irregularly sampled time series data. Weerakody et al. [15] conduct an in-depth survey and analysis of the established and cutting edge methods for handling irregular frequency time series data and point out the imbalance between the volume of techniques for dealing with regular vs. irregular time series data. Considering the rising quantity of data sampled at irregular intervals, there is growing incentive to develop the related technologies.

Weerakody et al. [15] continue to outline that RNNs are innately built for modelling time series data and that they have exclusive capacity for making sense of missing values in the data by utilising, instead of ignoring, the complicated temporal patterns that exist between the features and their sequence in time. Within the extensive review, Weerakody et al. [15] reinforce the standardised techniques for measuring classification algorithms within machine learning; accuracy, precision, recall and F1-score. This illustrates that despite the nature of the input data being irregular, the performance metrics are constant. In addition, the area under the receiver operator curve (ROC AUC) is used to evaluate various gated RNNs performance on irregularly sampled time series data, producing impressive results for the GRU-D, a modification of the mentioned GRU, with an AUC score of 0.8527.

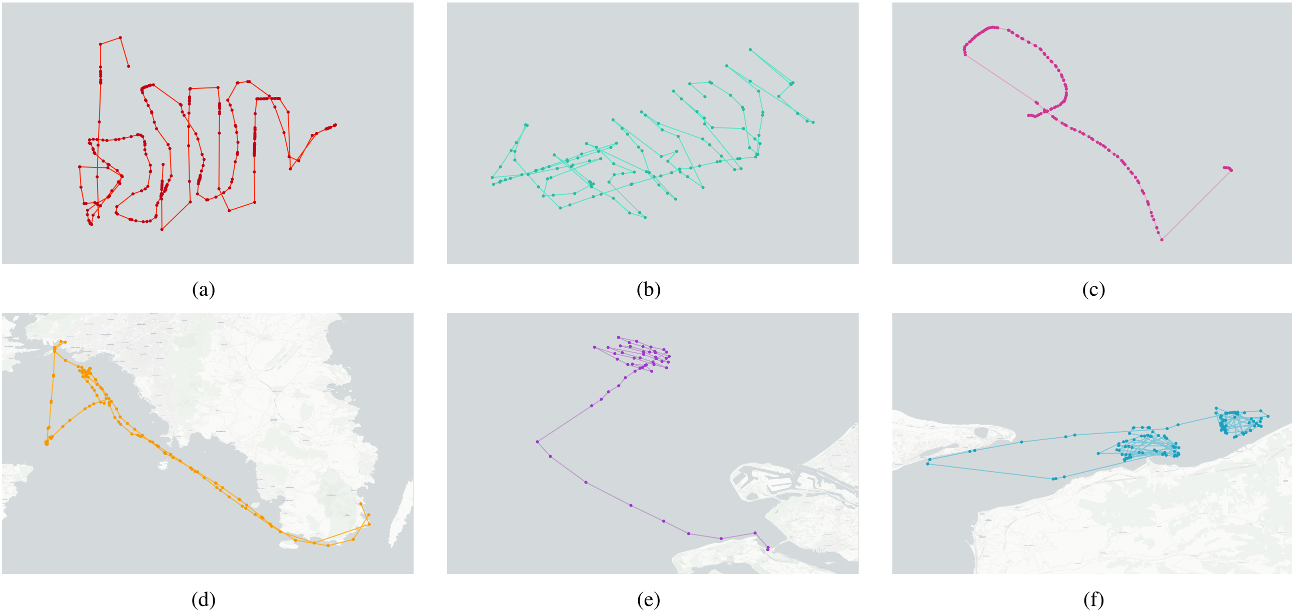
Shukla and Marlin [16] propose a novel deep learning approach for the task of supervised learning with irregular multivariate and univariate time series data. They introduce the multi-time attention network (mTAN), an encoder-decoder framework that leverages the attention mechanism to perform non-linear interpolation on the irregularly sampled observations for a number of machine learning applications. For classification, non-linearly interpolated, latent space embeddings are generated through the encoder module (a combination of multi-time attention blocks and a GRU), and are subsequently propagated through a fully-connected network to extract a decision. The proposal from Shukla and Marlin [16] is grounded in theory and supported with promising results, i.e. an accuracy of 94%. By interpolating the irregular time series data with a non-linear and proven method, and subsequently making predictions based on an architecture with an innate ability to model time series data, proven by Weerakody et al. [15] in the GRU, Shukla and Marlin [16] have presented a robust system that shows great potential in the field of irregular time series analysis.

The related research covers two broad topics: firstly, the attempt to use machine learning to determine IUU fishing and secondly the challenge of modelling irregular time series data. This proposal aims to incorporate the latter with the former in order to introduce a novel approach for classifying fishing vessels and subsequently identifying IUU fishing activity.

III. Data preprocessing

The automatic identification system (AIS) was designed for collision avoidance during voyages. For this reason, the signal transmissions are frequent, providing high-resolution multivariate data. Included in the signals is latitude and longitude, allowing vessel trajectories to be visualised, illustrated in figure 2. Although distinct patterns can be identified in the trajectories, the number of data points consisting of ambiguous and mixed behaviour make pattern recognition difficult for the human eye.

Due to the broadcast medium - very high frequency (VHF) channels, the signals are potentially noisy as messages can be delayed, lost or duplicated. This factor, coupled with the large quantity of data points, makes neural networks a suitable candidate for the task of pattern recognition. First however, the data must be processed into an optimal form for the models. The features of the data are MMSI, timestamp, course, speed, latitude and longitude. As this is a supervised learning classification task, the data must be labelled. This process has been done in advance by the GFW. The data is aggregated and cleaned, removing duplicates and features



(a) (b) (c) (d) (e) (f)

Fig. 2: Trajectory plots using Q-GIS: Drifting longline (a), fixed gear (b), pole and line (c), purse seines (d), trawlers (e), trollers (f).

that have no inference value. From here, the data processing pipeline diverges into two channels.

Linear interpolation is the backbone of the first pipeline, as shown in equation 1, where x_1 and y_1 and x_2 and y_2 are the first coordinates and second coordinates respectively, x is the point to perform the interpolation on and y is the interpolated value. Initially, the data points must be resampled into evenly spaced intervals. To proceed, a set of reference time points for every 24-hour sequence must be introduced. However, choosing an appropriate value for the data is not straightforward due to the high variance in lengths. On one hand, upsampling short sequences introduce noise in the form of fabricated data points and on the other, downsampling the longer sequences excludes valuable data. Using the calculated min-max mean, the chosen sampling frequency is one event per minute. When the reference time points are generated they are merged with the original data, linearly interpolated and resampled to the desired intervals.

$$y = y_1 + (x - x_1)\left(\frac{y_2 - y_1}{x_2 - x_1}\right) \tag{1}$$

$$\Delta t = t_{i+1} - t_i \tag{2}$$

The second pipeline segments the data into 24-hour sequences, regardless of length. The network will assume the sequences are spaced evenly, therefore the time difference between each element in the sequence is transformed into a feature, as shown in equation 2, where Δt is the time difference value, t_i and t_{i+1} are two sequential timestamp values. Finally, the sequences are zero padded to homogenise the data structure. This uniform size is set to the maximum sequence length in the dataset (2931). While this technique creates a large quantity of meaningless data, it preserves integrity, maintaining the AIS's high-resolution.

The MMSI must be removed from the data to mitigate any attempts at 'cheating' by the network. A simple marker such as ID that corresponds uniquely and frequently with a particular class, provides the network with shortcuts in the learning process instead of learning the underlying patterns in the causal features such as trajectory, speed and course. This conflict between causation and correlation, present universally in statistical analysis and machine learning, must be considered *a priori* during feature selection.

IV. Models

A. *Convolutional neural network*

By using artificial neural networks (ANNs) for the task at hand, it is acknowledged the objective is to find a function that approximates the underlying distribution of the data with accuracy [17]. Convolutional neural networks (CNNs) are a subclass of feed forward ANNs with alternating convolution and downsampling layers, trained by the backpropagation learning algorithm. Although convolutional neural networks were invented primarily for computer vision tasks, there has been recent research into their ability to forecast time series data [5]. Their structure is well suited to modelling temporal data as the kernels convolve over the sequences, producing feature maps that are subsequently downsampled to distil significant patterns in the data. When leveraging CNNs on time series data, a one-dimensional variation of the original two-dimensional CNN is implemented. Instead of convolving in multiple directions, the kernel passes over the sequence in one direction only.

Each feature of the multivariate time series data is transformed into an input channel. The one-dimensional kernels convolve over all the input channels, combining the information from all the independent variables into the feature maps, forward-propagating it through the network, while a loss function determines the error based on the desired outcome (the dependent variable) and the network's output. The gradient of the error with respect to the parameters of the network is calculated using partial derivatives and the chain rule in order to adjust the parameters in the direction of steepest descent on the multimodal error landscape: this is gradient descent via the backpropagation algorithm [18].

As the depth of the CNN grows to learn progressively more complex features, a familiar problem begins to emerge: the vanishing gradient. The gradient of the error with respect to the parameters being backpropagated through the network becomes increasingly small in the layers closest to the input, preventing learning from happening effectively as well as leading to numerical instability. Residual blocks were developed to mitigate this problem for CNNs in 2015 [He]. The novel idea was to include identity connections (also referred to as skip connections) which reinject the input features into the final layer of the residual block, before the ReLU activation. 1D CNNs with residual blocks have been implemented in this system to achieve stable results with a network deep enough to learn intricate patterns in the AIS data. Figure 3 illustrates the architecture of a 1D CNN with residual blocks implemented in this research. Equation 3 depicts the computations inside a 1D CNN for input dimension (N, C_{in}, L) and output dimension (N, C_{out}, L_{out}) where $\star$ is the sliding inner product operator, N is the batch size, C is the number of channels and L denotes the length of the input sequence.

$$out(N_i, C_{out_j}) = bias(C_{out_j}) + \sum_{k=0}^{C_{in}-1} weight(C_{out_j}, k) \star input(N_i, k) \tag{3}$$

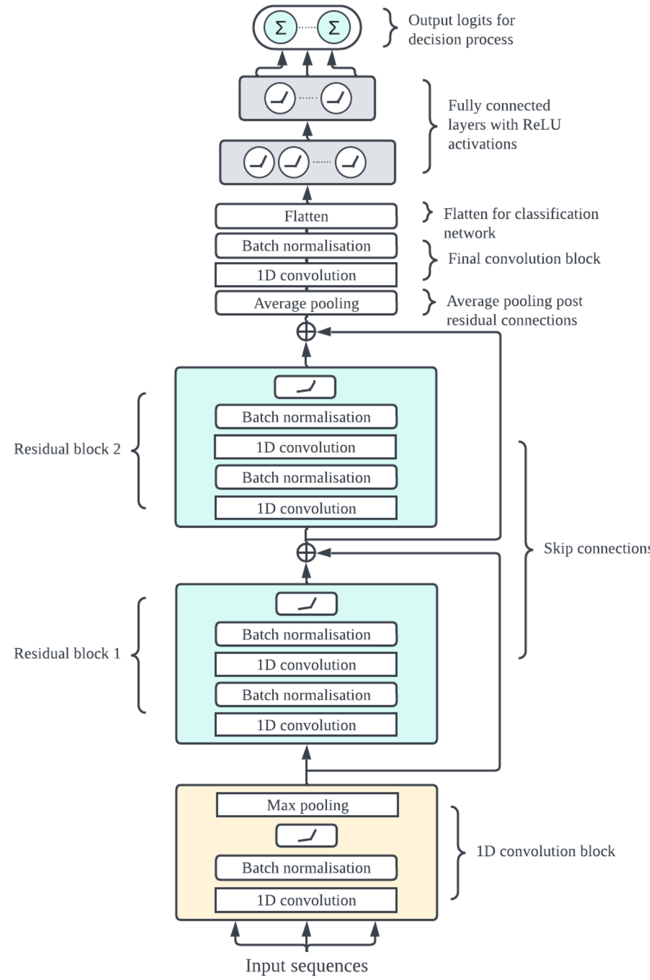


Fig. 3: 1D CNN architecture depicted showing residual blocks.

B. *Gated recurrent unit network*

As Weerakody et al. outlined, RNNs have an innate ability to model sequential data because they can incorporate information from previous steps in the sequence to make inferences [15]. However, they have been marred by the vanishing gradient problem. RNNs are susceptible to this issue because of the way they are trained using backpropagation through time (BPTT). The network is unrolled so that each time step is represented as a separate layer sharing the same parameters. The depth of the network grows linearly with time steps, which also leads to vanishing or exploding gradients.

$$\begin{aligned} r_t &= \sigma(W_{ir} \cdot x_t + b_{ir} + W_{hr} \cdot h_{(t-1)} + b_{hr}) \\ z_t &= \sigma(W_{iz} \cdot x_t + b_{iz} + W_{hz} \cdot h_{(t-1)} + b_{hz}) \\ n_t &= tanh(W_{in} \cdot x_t + b_{in} + r_t * (W_{hn} \cdot h_{(t-1)} + b_{hn})) \\ h_t &= (1 - z_t) * n_t + z_t * h_{(t-1)} \end{aligned} \tag{4}$$

Gated recurrent units address this problem by introducing update and reset gates. The update gate regulates the information passed through the network, while the reset gate determines how much should be excluded. By learning the parameters of these gates, the network is able to carry information from the inputs through the deep layers of the network, acting analogously to the skip connection in residual blocks mentioned previously. Figure 4 displays the structure of the stacked, bidirectional GRU implemented in this study. Equation 4 illustrates the calculations at the core of the GRU. Each layer computes the function for every element in the input sequence where x_i is the input at time t, h_t denotes the hidden state at time t, $h_{(t-1)}$ is the hidden state of the previous time step $(t-1)$, r_t is the reset gate, z_t denotes the update gate and n_t is the candidate hidden state. σ denotes the sigmoid function and $*$ denotes the element-wise product.

V. Evaluation Methodology

Overall accuracy on it's own is not sufficient for evaluation. Firstly, it assumes that the class distribution present in the data is representative of reality. Secondly, it fails to signify performance on under represented classes effectively. By using metrics such as precision, recall and F1-score (the weighted average of precision and recall as outlined by Grandini et al [19]) for individual classes, in conjunction with overall accuracy, a more comprehensive evaluation can be conducted.

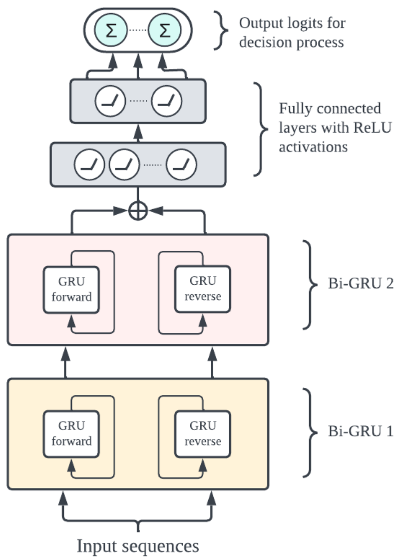


Fig. 4: Stacked, bidirectional GRU with a classification network appended.

VI. Results

The system shown in figure 1, is a comparison of the two data pipelines and classification modules, i.e. the linearly interpolated (pipeline 1) or zero padded (pipeline 2) data serving as input for both the 1D CNN with residual blocks and the stacked, bidirectional GRU. To prove statistical power, the common-sense prediction (CSP) model accuracy must be surpassed. The CSP amounts to predicting the most frequently observed class (drifting longlines) for every example, resulting in an overall accuracy of 56.63%. Early in the development process a minimal model outperformed the CSP: as shown in table I. The computational resources for all experiments conducted constitute a single machine with an AMD Ryzen 5950X CPU paired with an NVIDIA RTX 3090 GPU enabled for CUDA parallel computation.

	Accuracy	Precision	Recall	F1-score
All	57.93%	None	0.17898	0.14518
Drifting longline	–	0.58572	0.99422	0.73716
Fixed gear	–	0.0	0.0	0.0
Pole and line	–	0.0	0.0	0.0
Purse seines	–	0.0	0.0	0.0
Trawlers	–	0.42009	0.07965	0.13392
Trollers	–	0.0	0.0	0.0

TABLE I: Results produced by the minimal model demonstrate a lack of sufficient predictive power to learn the under-represented classes effectively.

Following the iterative deep learning workflow of training and validation, a trial and error exploration of the hyper-parameter space for each model was undertaken to produce optimal results. After exhaustive search of the hyperparameters, progress stalled. However, by implementing a technique known as non-random initialisation, introduced by Reed and Marks in Neural Smithing [20], the performance for both GRU and CNN models (structures in tables II and III, hyperparameters in tables IV and V) improved. Finally, using L1 norms (Manhattan distance), low magnitude weights were pruned to improve generalisation, thus enhancing performance on out-of-sample data to produce the results in tables VI, VII, VIII and IX. The non-random initialisation has a novel adjustment proposed in this paper to utilise validation accuracy as the defining metric as opposed to training loss in order to identify the starting weights with high potential for further training.

Layer	Name	Parameters	Dimensions
1	Input	–	128, 5, 2931
2	1D convolution Batch normalisation Leaky ReLU Max pooling	64, 5, 3 – – 2, 2	128, 64, 2929 – – 128, 64, 1464
3	Residual block(1D convolution Batch normalisation 1D convolution Batch normalisation Leaky ReLU)	– 64, 64, 3 – 64, 64, 3 – –	– 128, 64, 1464 – 128, 64, 1464 – –
4	Residual block(1D convolution Batch normalisation 1D convolution Batch normalisation Leaky ReLU)	– 64, 64, 3 – 64, 64, 3 – –	– 128, 64, 1464 – 128, 64, 1464 – –
5	Average pooling	64, 128	128, 64, 732
6	1D convolution Batch normalisation	128, 64, 3 –	128, 128, 730 –
7	Flatten	–	128, 46720
8	Classification block(Fully connected 1 (ReLU) Fully connected 2 (ReLU) Fully connected 3 (ReLU))	– 128, 46720 64, 128 6, 64	– 128, 128 128, 64 128, 6
	Total	6,064,966	

TABLE II: 1D CNN with residual blocks structure (for zero padded pipeline), the total number of trainable parameters significantly outnumber the GRU model.

As illustrated by the results for zero padded data in tables VI and VIII, the stacked, bidirectional GRU outperforms the 1D CNN with residual blocks in terms of overall accuracy and individual class metrics. Given the data serving as input for both models is indeed irregular, Weerakody et al.'s [15] research on the suitability of gated recurrent neural networks for modelling irregular time series data holds firm. Results from the linearly interpolated data demonstrate again the superiority of the GRU network as depicted in tables VII and IX, but also show adverse affects on the performance of the CNN, with accuracy down from 93.3% to 90.6%. The GRU responded well to both data pipelines, although accuracy is lower with 94% compared to 95.379%, the macro F1-score is marginally higher for the linearly interpolated data at 0.92868 compared to 0.92457, show in tables VIII and IX.

As demonstrated by the experiments, the best model for the task at hand is the stacked, bidirectional GRU, while the answer for the prevailing data pipeline is not so straightforward. There is a case to be made for utilising linear interpolation when the class imbalance is severe due to the improved per-

Layer	Name	Parameters	Dimensions
1	Input	–	64, 2931, 5
2	GRU stack 1 *forward*(input - hidden hidden - hidden)	– 192, 5 192, 64	64, 2931, 64 – –
3	GRU stack 1 *reverse*(input - hidden hidden - hidden)	– 192, 5 192, 64	64, 2931, 64 – –
4	GRU stack 2 *forward*(input - hidden hidden - hidden)	– 192, 128 192, 64	64, 2931, 64 – –
5	GRU stack 1 *reverse*(input - hidden hidden - hidden)	– 192, 5 192, 64	64, 2931, 64 – –
6	Classification block(Fully connected 1 (ReLU) Fully connected 2 (ReLU) Fully connected 3 (ReLU))	– 128, 128 64, 128 6, 64	– 64, 128 64, 64 64, 6
	Total	126,918	

TABLE III: Stacked, bidirectional GRU structure (for zero padded pipeline), although the total number of trainable parameters number less than the competing CNN model, bidirectionality introduces additional computation to training.

Hyperparameter	Value
Learning rate	3e-04
Optimiser	AdamW
Loss	Cross entropy loss
Output channels convolution	64
Kernel size	3, 1
Pool size	2
Number of convolution layers	6
Fully connected layers	3
Dropout	False
Batch size	128
Shuffled	True
Sequence length	2931
Non-random weight initialisation	True
L1-norm global unstructured weight pruning	False
Epochs	35

TABLE IV: CNN hyperparameters, depicted is only one of many combinations explored.

Hyperparameter	Value
Learning rate	3e-04
Optimiser	AdamW
Loss	Cross entropy loss
Number of stacked recurrent units	2
Bidirectional	True
Fully connected layers	3
Dropout	0.1
Batch size	64
Shuffled	True
Sequence length	2931
Non-random weight initialisation	True
L1-norm global unstructured weight pruning	True
Epochs	68

TABLE V: GRU hyperparameters (zero padded data).

formance on under represented classes. However, the overall accuracy is still superior for the zero padded data when paired with the GRU. The decision to adopt linear interpolation or zero padding with a time difference feature, depends on the data distribution and the importance of classifying accurately the under represented classes. In the case where smaller classes are highly important, these experiments would suggest

	Accuracy	Precision	Recall	F1-score
All	93.302%	0.87731	0.87549	0.87608
Drifting longline	–	0.97026	0.97094	0.9706
Fixed gear	–	0.81932	0.85871	0.83855
Pole and line	–	0.77558	0.80205	0.78859
Purse seines	–	0.89934	0.8386	0.86791
Trawlers	–	0.91952	0.92489	0.9222
Trollers	–	0.87983	0.85774	0.86864

TABLE VI: CNN results on zero padded test data.

	Accuracy	Precision	Recall	F1-score
All	90.619%	0.8785	0.86836	0.87219
Drifting longline	–	0.94921	0.96996	0.95947
Fixed gear	–	0.81932	0.79946	0.80638
Pole and line	–	0.91631	0.83571	0.87415
Purse seines	–	0.876	0.77802	0.82411
Trawlers	–	0.88632	0.9051	0.89561
Trollers	–	0.82973	0.92192	0.8734

TABLE VII: CNN results on linearly interpolated test data.

	Accuracy	Precision	Recall	F1-score
All	95.379%	0.92317	0.92653	0.92457
Drifting longline	–	0.97895	0.97796	0.97845
Fixed gear	–	0.86385	0.92324	0.89256
Pole and line	–	0.88621	0.87713	0.88165
Purse seines	–	0.92177	0.87851	0.89962
Trawlers	–	0.95384	0.94434	0.94906
Trollers	–	0.93443	0.95798	0.94606

TABLE VIII: GRU with zero padded test data: highest overall accuracy.

	Accuracy	Precision	Recall	F1-score
All	94.01%	0.92945	0.9282	0.92868
Drifting longline	–	0.97006	0.97204	0.97105
Fixed gear	–	0.90009	0.86639	0.88292
Pole and line	–	0.92829	0.90427	0.91612
Purse seines	–	0.8652	0.89939	0.88196
Trawlers	–	0.93673	0.93611	0.93642
Trollers	–	0.97633	0.99099	0.98361

TABLE IX: GRU linearly interpolated test data results.

utilising linear interpolation. In cases where the more prevalent classes carry equal or greater significance, the high-resolution and time difference feature of the zero padded pipeline takes on greater importance.

VII. Conclusion

Use of the vast quantity of high-resolution data generated by the automatic identification system (AIS) can be a powerful tool in the prevention of illegal, unreported and unregulated (IUU) fishing. This paper has proposed a system for processing the highly irregular and noisy time series data to extract meaningful information in the form of fishing method identification. When leveraged in real-time, this system reveals key information about potentially suspicious behaviour in marine protected areas or falsely registered equipment.

By procuring results with strong statistical power, i.e. outperforming the common-sense prediction model using a robust and proven set of methodologies, the proposed system has been successful. In addition, the contemporary results obtained by Kalaiselvi et al. [8] of 0.885 macro F1-score have been surpassed by the GRU with scores of 0.92868 and 0.92457 for linearly interpolated and padded data respectively. This is despite the proposed system including two smaller classes: pole and line and trollers, introducing added class imbalance, further complicating the task. The high degree of accuracy of the experiments conducted in this study has provided substantial evidence supporting the existence of a clear relationship between vessel movement patterns and method of fishing.

Moving forward, the introduction of non-linear interpolation will be the main focus through the implementation of the multi-time attention network introduced by Shukla and Marlin [16]. Additionally, traditional machine learning classification algorithms such as random forests and support vector machines will be introduced for further comparison of results. This future work remains novel and ambitious, with a strong foundation of results to build upon.

REFERENCES

[1] FAO, *The State of World Fisheries and Aquaculture - 2006 (SOFIA)* (The State of World Fisheries and Aquaculture (SOFIA) 2006). Rome, Italy: FAO, 2007, 170 pp., ISBN: 978-92-5-105568-7.

[2] J. Vince, B. D. Hardesty, and C. Wilcox, "Progress and challenges in eliminating illegal fishing," *Fish and Fisheries*, vol. 22, no. 3, pp. 518–531, 2021, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/faf.12532, ISSN: 1467-2979. DOI: 10.1111/faf.12532.

[3] D. J. Agnew, J. Pearce, G. Pramod, *et al.*, "Estimating the worldwide extent of illegal fishing," *PLOS ONE*, vol. 4, no. 2, e4570, Feb. 25, 2009, Publisher: Public Library of Science, ISSN: 1932-6203. DOI: 10.1371/journal.pone.0004570.

[4] I. M. O. (IMO). "AIS transponders." (), [Online]. Available: https://www.imo.org/en/OurWork/Safety/Pages/AIS.aspx (visited on 11/07/2022).

[5] Y. Li, T. Cai, Y. Zhang, D. Chen, and D. Dey, *What makes convolutional models great on long sequence modeling?* Oct. 17, 2022. DOI: 10.48550/arXiv.2210.09298. arXiv: 2210.09298[cs,stat].

[6] K.-i. Kim and K. M. Lee, "Convolutional neural network-based gear type identification from automatic identification system trajectory data," *Applied Sciences*, vol. 10, no. 11, p. 4010, Jan. 2020, Number: 11 Publisher: Multidisciplinary Digital Publishing Institute, ISSN: 2076-3417. DOI: 10.3390/app10114010.

[7] S. Lahmiri, "Features extraction from high frequency domain for retina digital images classification," *Journal of Advances in Information Technology*, vol. 4, Nov. 1, 2013. DOI: 10.4304/jait.4.4.194-198.

[8] V. Kalaiselvi, P. R. D., J. Ranjani, V. K. S. M., and M. P. S., "Illegal fishing detection using neural network," in *2022 International Conference on Communication, Computing and Internet of Things (IC3IoT)*, Mar. 2022, pp. 1–4. DOI: 10.1109/IC3IOT53935.2022.9767876.

[9] M. I. Marzuki, P. Gaspar, R. Garello, V. Kerbaol, and R. Fablet, "Fishing gear identification from vessel-monitoring-system-based fishing vessel trajectories," *IEEE Journal of Oceanic Engineering*, vol. 43, no. 3, pp. 689–699, Jul. 2018, Conference Name: IEEE Journal of Oceanic Engineering, ISSN: 1558-1691. DOI: 10.1109/JOE.2017.2723278.

[10] D. Sánchez Pedroche, D. Amigo, J. García, and J. M. Molina, "Architecture for trajectory-based fishing ship classification with AIS data," *Sensors*, vol. 20, no. 13, p. 3782, Jan. 2020, Number: 13 Publisher: Multidisciplinary Digital Publishing Institute, ISSN: 1424-8220. DOI: 10.3390/s20133782.

[11] X. Chen, Y. Liu, K. Achuthan, and X. Zhang, "A ship movement classification based on automatic identification system (AIS) data using convolutional neural network," *Ocean Engineering*, vol. 218, p. 108 182, Dec. 15, 2020, ISSN: 0029-8018. DOI: 10.1016/j.oceaneng.2020.108182.

[12] J. Villaruz, "Deep convolutional neural network feature extraction for berry trees classification," *Journal of Advances in Information Technology*, vol. 12, Jan. 1, 2021. DOI: 10.12720/jait.12.3.226-233.

[13] S. Bunrit, N. Kerdprasop, and K. Kerdprasop, "Improving the representation of CNN based features by autoencoder for a task of construction material image classification," *Journal of Advances in Information Technology*, vol. 11, pp. 192–199, Jan. 1, 2020. DOI: 10.12720/jait.11.4.192-199.

[14] Y. Tay, M. Dehghani, S. Abnar, *et al.*, *Long range arena: A benchmark for efficient transformers*, Nov. 8, 2020. DOI: 10.48550/arXiv.2011.04006. arXiv: 2011.04006[cs].

[15] P. B. Weerakody, K. W. Wong, G. Wang, and W. Ela, "A review of irregular time series data handling with gated recurrent neural networks," *Neurocomputing*, vol. 441, pp. 161–178, Jun. 21, 2021, ISSN: 0925-2312. DOI: 10.1016/j.neucom.2021.02.046.

[16] S. N. Shukla and B. Marlin, "Multi-time attention networks for irregularly sampled time series," presented at the International Conference on Learning Representations, Mar. 18, 2021.

[17] C. M. Bishop, G. Hinton, C. M. Bishop, and G. Hinton, *Neural Networks for Pattern Recognition*. Oxford, New York: Oxford University Press, Nov. 23, 1995, 504 pp., ISBN: 978-0-19-853864-6.

[18] S. S. Haykin, *Neural networks: a comprehensive foundation*, 2nd ed.. Upper Saddle River, N.J.: Prentice Hall, 1999, xxi+842, ISBN: 978-0-13-273350-2.

[19] M. Grandini, E. Bagli, and G. Visani, *Metrics for multi-class classification: An overview*, Aug. 13, 2020. arXiv: 2008.05756[cs,stat].

[20] R. D. Reed and R. J. Marks, *Neural Smithing: Supervised Learning in Feedforward Artificial Neural Networks*. Cambridge, MA, USA: MIT Press, 1998, 346 pp., ISBN: 978-0-262-18190-7.