



**HAL**  
open science

## Apprentissage auto-supervisé pour la détection d'actions illégales lors de la surveillance du trafic maritime

P Bernabé, A Gotlieb, B Legeard, F Olaf Sem-Jacobsen, H Spieker

### ► To cite this version:

P Bernabé, A Gotlieb, B Legeard, F Olaf Sem-Jacobsen, H Spieker. Apprentissage auto-supervisé pour la détection d'actions illégales lors de la surveillance du trafic maritime. Conférence Nationale sur les Applications Pratiques de l'Intelligence Artificielle (APIA 2021), Jul 2021, Bordeaux, France. hal-03874246

**HAL Id: hal-03874246**

**<https://hal.science/hal-03874246v1>**

Submitted on 27 Nov 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Apprentissage auto-supervisé pour la détection d'actions illégales lors de la surveillance du trafic maritime

P. Bernabé<sup>1,2</sup>, A. Gotlieb<sup>1</sup>, B. Legeard<sup>2,4</sup>, F. Olaf Sem-Jacobsen<sup>3</sup>, H. Spieker<sup>1</sup>

<sup>1</sup> Simula Research Laboratory, VIAS - Oslo, Norvège

<sup>2</sup> Université de Bourgogne Franche-Comté, Institut FEMTO-ST - Besançon, France

<sup>3</sup> Statsat AS - Oslo, Norvège

<sup>4</sup> Smartesting - Besançon, France

{pierbernabe, arnaud, helge}@simula.no, frank.Sem-Jacobsen@statsat.no,  
bruno.legeard@univ-fcomte.fr

## Résumé

*La surveillance du trafic maritime est confrontée à des difficultés très importantes dans la détection des activités illégales en mer. Dans cet article, nous présentons les premiers résultats d'une méthode d'apprentissage auto-supervisé qui vise à déceler les déconnexions volontaires du système d'identification des navires. En traitant les données provenant de quatre satellites de surveillance norvégiens, notre modèle d'apprentissage vise l'identification de navires soupçonnés d'activités illégales telles que la pêche dans des zones protégées ou bien le franchissement de zones d'exclusion économique en temps réel. Dans cet article, nous présentons une approche fondée sur des techniques d'apprentissage auto-supervisé, et expérimentée à partir de données réelles.*

## Mots-clés

*Apprentissage automatique, Apprentissage auto-supervisé, Système d'identification automatique, Surveillance maritime.*

## Abstract

*The surveillance of maritime traffic is confronted with very important difficulties in detecting illegal activities at sea. In this article, we present the first results of a self-supervised learning method which aims to detect voluntary disconnections of the identification' system of vessels. By processing data from four Norwegian surveillance satellites, our learning model aims to identify vessels suspected of illegal activities such as fishing in protected areas or crossing economic exclusion zones in real time. In this article, we present an approach based on self-supervised learning techniques, and experienced from real data.*

## Keywords

*Machine Learning, Self-Supervised Learning, Automatic Identification System (AIS), Maritime Surveillance*

## 1 Introduction

### 1.1 Surveillance du trafic maritime

La surveillance du trafic maritime est une tâche complexe qui vise à identifier et contrôler l'activité des navires présents dans une zone maritime [12]. La surveillance a plusieurs objectifs, comme par exemple, identifier et guider les navires, aider à la prévention des collisions, lancer des missions de sauvetage en mer ou plus généralement, réguler le trafic maritime. Les systèmes de surveillance modernes visent aussi à détecter des activités illégales telles que les actions de piratage, de pêche illégale, d'intrusion dans les zones d'exclusion économique (EEZ), de transbordement de produits stupéfiants, de dégazage en mer, etc. La plupart du temps, la détection de ces activités illégales repose uniquement sur l'observation des navires, l'analyse visuelle des données et l'expertise des garde-côtes. Pourtant, l'exploitation automatisée des différentes sources de données pourrait s'avérer très utile pour la détection d'actions illégales [1]. En effet, les systèmes de surveillance maritime collectent des données provenant des communication-radio spécifiques émises par les navires, et celles-ci sont captées par différents moyens tels que des balises en mer, les sémaphores côtiers et des satellites dédiés à l'observation du trafic maritime.

L'émission de messages par des transpondeurs VHF<sup>1</sup> embarqués sur les navires d'un certain tonnage (i.e.,  $\geq 300$  tonnes) constitue la source d'information principale de la surveillance maritime. Ces messages s'appuient sur le protocole AIS (*Automatic Identification System*) qui transmet toutes les 2 à 10 secondes des informations telles que l'identifiant unique du navire (MMSI<sup>2</sup>), son statut (amarré, échoué, au mouillage, etc.), son type (cargo, pêche, plaisance, etc.), ses dimensions, sa route, sa vitesse, sa position (latitude, longitude), son cap, etc. Ces messages sont captés non seulement par les autres navires ainsi que les balises et les sémaphores lorsque les récepteurs sont à portée, mais

1. Very High Frequency

2. Maritime Mobile Service Identity

également par des satellites dédiés à la surveillance du trafic comme indiqué plus haut. Dans ce dernier cas, les données sont nommées S-AIS. Il est important de noter que le capitaine d'un navire a la possibilité de couper volontairement l'émission des messages AIS et ceci est parfois fait dans le but de réaliser des actions illégales en mer. Par exemple, la déconnexion AIS est fréquemment utilisée par certains capitaines de navire de pêche afin de pêcher en toute impunité dans une zone interdite ou fortement réglementée. De part le volume de messages AIS reçus, une analyse manuelle de ces données reste très hasardeuse et la plupart du temps inopérante. Par exemple, en une seule journée (19/5/2019), les satellites de surveillance du trafic maritime opérés par la Norvège ont réceptionné 3 518 649 messages AIS, qui sont à ajouter aux 4 192 381 messages captés par les stations terrestres des côtes norvégiennes.

## 1.2 Méthodes pour la détection automatique d'actions illégales

L'utilisation de moyens automatisés pour la détection d'actions illégales est donc devenu un enjeu pour la surveillance du trafic maritime [4]. Parmi ces moyens, la recherche de déconnexions AIS volontaires à partir des messages AIS transmis par les navires est apparu comme une première application cruciale [8, 7, 14, 11]. Avec le lancement de satellites de surveillance, la disponibilité de données S-AIS pour des zones habituellement hors de portée a permis de systématiser cette surveillance y compris en pleine mer. Une approche directe pour ce problème consiste à définir des règles qui caractérisent la déconnexion volontaire des transpondeurs VHF mais cette approche se heurte à plusieurs difficultés liées essentiellement à la qualité des données. En effet,

- les messages S-AIS captés par une flotte de satellites de surveillance sont habituellement très irréguliers à cause des conditions météo difficiles, des collisions entre messages, de la position des satellites qui orbitent autour de la terre et qui n'offrent pas une couverture permanente des mers ;
- certains messages S-AIS sont bruités car les données non mises à jour, ou bien émises par des transpondeurs de piètre qualité ;
- certains navires embarquent plusieurs transpondeurs et brouillent volontairement les pistes en changeant d'émission VHF ;
- certaines zones sont saturées en bateaux (e.g., à l'approche des ports ou bien dans des couloirs très empruntés) et conduit à la congestion des modules de réception des satellites ;

Devant l'impossibilité d'identifier un système de règles caractérisant la déconnexion AIS légale et illégale, un courant de recherche a émergé depuis quelques années, qui consiste à entraîner et déployer des modèles d'apprentissage automatique supervisé pour la détection d'actions illégales en mer. La plupart des travaux initiaux se sont concentrés sur la prédiction de trajectoires et la détection d'anomalies pour un navire donné, sans se préoccuper du reste du trafic maritime [12, 4]. De même, des modèles probabilistes du com-

portement individuel des navires à partir de données historiques AIS ont émergés tels que des modèles de Markov [1, 5] ou bien des réseaux de neurones hiérarchiques [6]. Cependant, ces modèles sont d'une part mal adaptés au traitement de données bruitées telles que les données AIS issues des satellites et d'autre part, en ne prenant en compte que les comportements individuels des navires, ils manquent souvent de pertinence pour détecter des actions illégales dans la multitude des navires traversant une zone maritime donnée. Récemment, l'utilisation de l'apprentissage multi-tâches pour l'entraînement d'un modèle probabiliste de trajectoires typiques de navires à partir de données AIS a donné des résultats très encourageants pour la surveillance du trafic maritime [9]. En particulier, en utilisant une représentation qui régularise la fréquence des messages et complète le jeu de données, cette approche a permis un bien meilleur traitement des données bruitées [10] même si les jeux de données utilisés proviennent, non pas des satellites de surveillance, mais des balises en mer.

## 1.3 Nos contributions

Les résultats présentés dans cet article s'articulent autour de trois contributions distinctes :

1. À partir de données S-AIS provenant du captage d'une flotte de satellites opéré par notre partenaire Norvégien StatSat AS, nous constatons que la présence d'évènements de déconnexion AIS est très fréquente et seule une très faible proportion correspond à des déconnexions volontaires. Notre approche, basée sur l'apprentissage profond, permet de détecter ces déconnexions de manière fiable. C'est, à notre connaissance, la première fois que les déconnexions volontaires sont recherchées et détectées dans des données S-AIS avec un modèle à base de réseau de neurones multi-couches. Ceci ouvre des perspectives intéressantes pour couvrir des zones maritimes situées en pleine mer, c'est-à-dire situées loin des côtes et des balises, et pour la détection temps-réel des déconnexions permettant une intervention plus rapide des garde-côtes ;
2. Nous utilisons des données non-étiquetées pour entraîner nos modèles à base de réseaux de neurones dans une approche d'apprentissage auto-supervisée. Ces données n'ont pas été annotées par des opérateurs capables de discriminer les déconnexions AIS volontaires des autres, ce qui facilite l'automatisation et la généralisation de notre approche ;
3. À l'inverse d'autres méthodes existantes, notre approche ne s'appuie pas sur la reconstruction de messages AIS manquants. Nous créons une représentation qui s'appuie sur les réseaux d'attention et bénéficions ainsi d'une plus grande précision dans l'analyse des données. C'est, à notre connaissance, la première fois que les réseaux d'attention sont utilisés pour traiter des données AIS. Les résultats expérimentaux que nous avons obtenus démontrent les bénéfices liés à cette représentation.

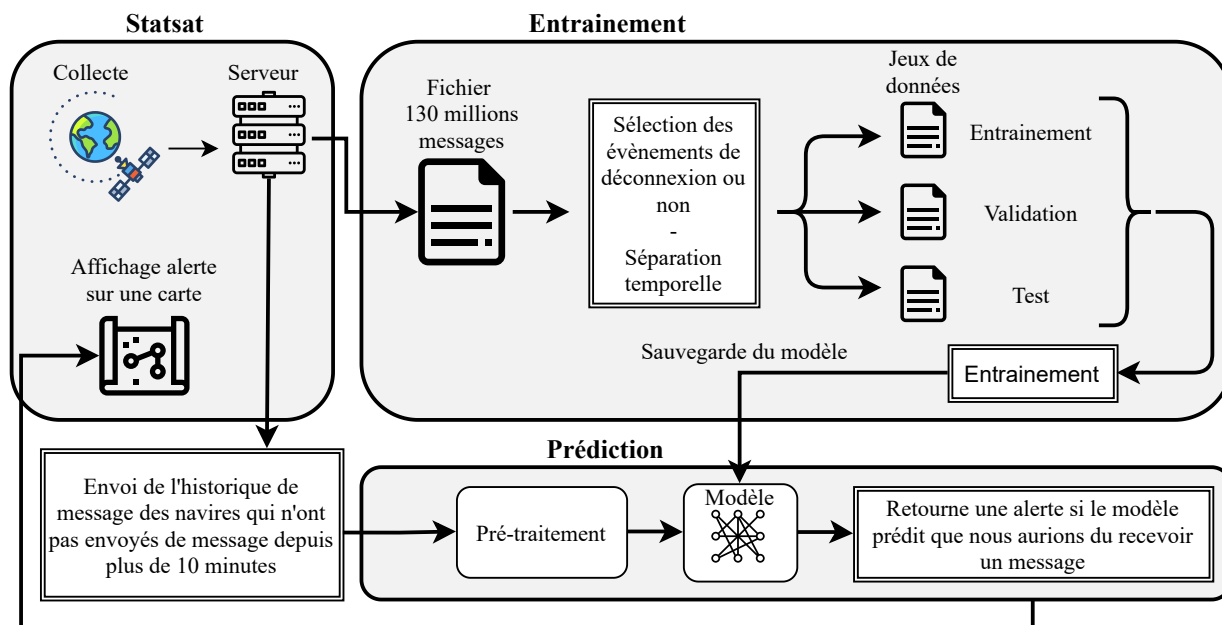


FIGURE 1 – Schéma de fonctionnement général de notre méthode d'apprentissage auto-supervisé pour la détection de déconnexions AIS volontaires.

## 1.4 Plan du papier

Le reste de cet article est organisé de la manière suivante : la section 2 donne un aperçu général de notre approche et indique les pré-requis nécessaires à sa compréhension avec une présentation succincte de l'apprentissage auto-supervisé et des réseaux d'attention. La section 3 décrit le traitement du flux des données, tandis que la section 4 présente notre modèle d'apprentissage profond et la représentation à base de réseaux d'attention. Dans la section 5, nous interprétons les résultats expérimentaux obtenus. Enfin, la section 6 conclut cet article en évoquant les pistes de travail en cours concernant l'application de l'apprentissage auto-supervisé à d'autres problématiques de la surveillance du trafic maritime.

## 2 Aperçu de notre approche

### 2.1 Schéma de fonctionnement général

La figure 1 présente le schéma général de notre méthode d'apprentissage auto-supervisé de déconnexions volontaires AIS. Le schéma se décompose en trois blocs distincts, intitulés *Statsat*, *Entraînement* et *Prédiction*. Le bloc *Statsat* correspond au travail de collecte et stockage des données S-AIS en provenance de satellites d'observation géo-marins gérés par l'opérateur étatique Norvégien StatSat AS. Le bloc *Entraînement* correspond au travail de sélection et de préparation des données et de calcul des annotations supplémentaires afin de créer un jeu de données propice à l'entraînement d'un modèle d'apprentissage auto-supervisé. Ce jeu de données est composé, comme à l'accoutumée, de trois sous-ensembles distincts : les données d'entraînement, les données de validation et les données de

test. Une fois le modèle entraîné, celui-ci est déployé dans le bloc *Prédiction* qui prend en entrée l'historique en messages AIS de navires qui n'ont pas envoyé de messages depuis plus de 10 minutes. Après une phase de pré-traitement, les trajectoires de ces navires sont classifiées par le modèle entre ceux présentant un risque d'activités illégales (risque de déconnexions AIS volontaires) et les autres. Dans le premier cas, une alerte spécifique peut-être retournée au block *StatSat* qui se charge d'alerter la garde-côtière si la menace se confirme. À titre d'exemple, la figure 2 montre la trajectoire d'un navire identifié par notre méthode. La figure montre la distance parcourue et l'irrégularité entre les messages. En pleine mer, au franchissement d'une ligne territoriale, le navire cesse d'émettre les messages AIS, ce qui le rend suspect d'activités illégales.

### 2.2 Apprentissage auto-supervisé

Détecter les déconnexions AIS volontaires se heurte à un obstacle d'importance : il n'existe pas de moyen parfaitement adéquat d'annoter les jeux de données. En effet, comme indiqué plus haut, une solution consisterait à entraîner un modèle avec des données annotées par les garde-côtes, mais un tel travail est peu réaliste en pratique, du fait du temps exigé pour des personnels très qualifiés et très occupés. De plus, dans la mesure où de très nombreuses déconnexions AIS sont involontaires, il s'agirait peu ou prou de rechercher des aiguilles dans une botte de foin et la constitution d'un jeu de données équilibré serait extrêmement fastidieuse. Notre solution pour ce problème a donc consisté à extraire l'annotation depuis les données elles-mêmes dans une approche d'apprentissage auto-supervisé. Du fait que la très grande majorité des exemples où un bateau n'envoie pas de messages AIS pendant un laps de

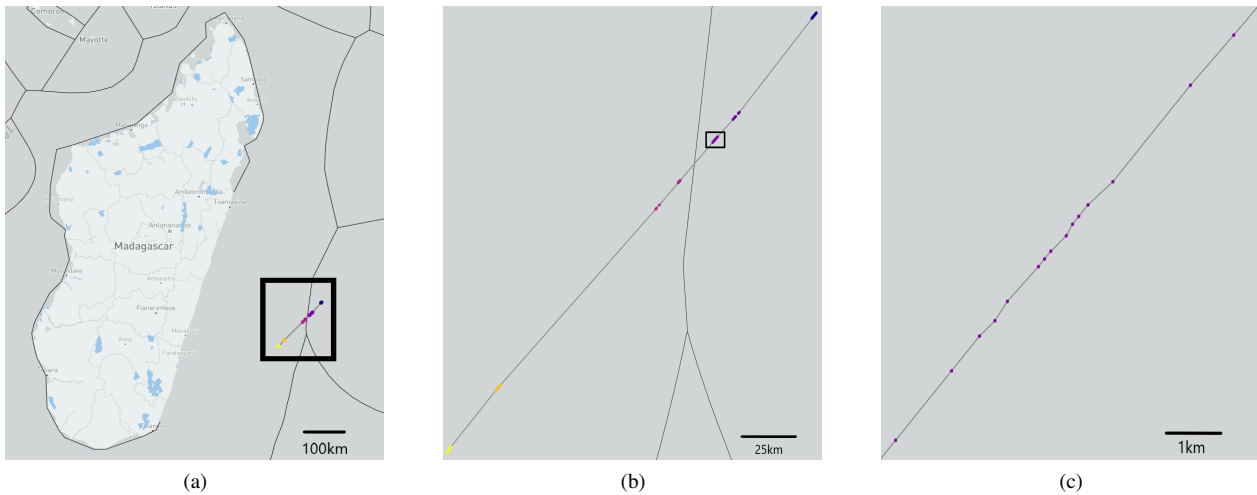


FIGURE 2 – Exemple d’une trajectoire d’un navire à différentes échelles (de la plus grande échelle (a) à la plus petite (c)). La trajectoire est ici composée de 100 messages AIS, tandis que le navire parcourt une distance d’environ 250km.

temps suffisant est due à une perte de connexion AIS involontaire, notre modèle peut être entraîné avec l’objectif d’apprendre la tâche qui consiste à déterminer si un message doit être reçu dans ce laps de temps ou non. Si notre modèle prédit qu’un message doit être reçu dans ce laps de temps et qu’il n’arrive pas, alors le navire correspondant peut être classifié comme suspect. Ce laps de temps significatif, à la vue des données S-AIS dont nous disposons, a été fixé à 10 minutes.

Cette approche a l’avantage de permettre la construction d’un large jeu de données annotées à partir de l’historique des messages AIS, sans nécessiter une annotation manuelle. En effet, il suffit d’extraire des exemples de disparition de plus de 10 minutes ainsi que des exemples de réception de messages dans les 10 minutes pour construire un large jeu de données bien équilibré.

### 2.3 Réseaux d’attention

Dans le domaine de l’apprentissage auto-supervisé, les réseaux d’attention et les "transformer models" ont montré des résultats spectaculaires, en particulier dans le traitement automatique du langage naturel [13]. Introduits en 2015 dans [2], l’objectif des réseaux d’attention est d’améliorer l’apprentissage des parties les plus importantes des données d’entrée, en consacrant plus de puissance de calcul à celles-ci. Ces réseaux se sont montrés capables de corriger certains défauts inhérents des "Long Short Term Memory". Une variante des premiers réseaux d’attention, nommée "self-attention network" (ou bien "intra-attention network") permet de mettre en relation différents éléments d’une même séquence d’entrée pour en créer ce qui s’appelle une représentation [3]. Ces réseaux se sont révélés très utiles dans la création automatique de résumés de texte. Les "transformer models" sont une variante des "self-attention networks" construits sans utiliser d’architecture récurrente. Ceci est possible grâce au mécanisme "multi-head self-attention" [13].

Dans notre approche, nous utilisons la partie encoder du

"transformer models" composé de deux blocs "transformer". Cet encodeur construit une représentation de la trajectoire du navire basée sur l’attention avec l’objectif de trouver des relations entre les différents messages d’un même navire, plus ou moins proches, et ainsi créer une représentation, comme détaillé dans la section 4.

## 3 Traitement du flux de données

### 3.1 Messages AIS

Les messages AIS sont composés d’informations statiques et dynamiques. Les champs statiques incluent les identifiants internationaux normalisés du navire, i.e., MMSI<sup>3</sup> and IMO, le nom du navire, le signe d’appel, la longueur, la largeur et le type de navire. Ces éléments statiques, qui sont saisis manuellement par le capitaine du navire dans l’émetteur AIS, sont automatiquement transmis sur un canal de diffusion, toutes les 6 minutes. Les émetteurs AIS envoient également des informations dynamiques toutes les 2 à 10 secondes selon la vitesse du bateau, ou toutes les 3 minutes si le bateau est au mouillage. Les informations dynamiques comprennent l’état de navigation (par exemple, "au mouillage", "pêche", etc.), la position du navire (latitude LAT, longitude LON), la vitesse du navire (SOG)<sup>4</sup>, sa direction par rapport au pôle Nord (COG)<sup>5</sup>, son cap qui est la direction (par rapport au pôle nord magnétique ou au pôle nord géographique) et les horodatages. Tous les messages AIS ne contiennent pas les mêmes informations et ne sont pas toujours envoyés à des horodatages réguliers. En règle générale, les messages AIS ont une portée d’environ 20 à 40km. La limitation de cette portée est due à la courbure de la terre et à la hauteur à laquelle l’antenne est installée sur les navires.

Depuis une dizaine d’années environ, le corps de garde-côtes utilise des satellites pour capter les messages AIS en

3. Maritime Mobile Service Identity

4. Speed Over Ground

5. Course Over Ground

TABLE 1 – Jeux de données utilisés pour le bloc Entraînement.

Jeux de données	Nombre de messages	Uniques navire	Nombre d'exemples extraits	Date
Entraînement	101 037 023	132 486	80 000	24/04 -> 26/05
Validation	12 497 753	88 489	10 000	26/05 -> 30/05
Test	12 945 336	91 023	10 000	30/05 -> 03/06

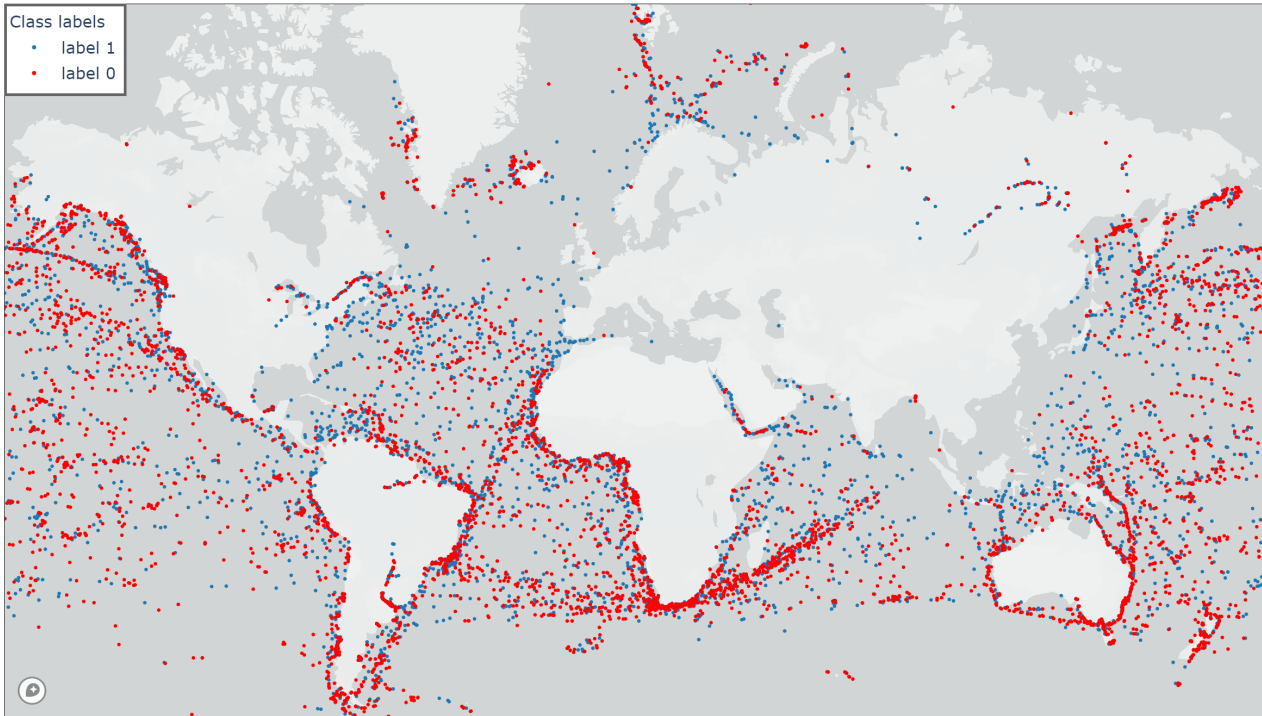


FIGURE 3 – Répartition des échantillons sur la surface du globe

dehors des zones couvertes par les balises. Cependant, il existe un problème inhérent à la norme AIS : le schéma d'accès radio défini dans la norme crée seulement 4500 plages horaires disponibles à chaque minute et les récepteurs peuvent être facilement submergés par de grandes empreintes de réception AIS. De part le nombre croissant d'émetteurs-récepteurs AIS et l'élargissement de la zone de captation des satellites, des collisions de messages peuvent avoir lieu et conduire à la disparition de certains d'entre eux. Certains ports de la mer Méditerranée, de la mer du Nord et des côtes chinoises sont très difficiles à surveiller à cause de cela.

### 3.2 Collection des données

Statsat est une compagnie norvégienne qui développe des infrastructures spatiales et opère des satellites de surveillance maritime pour le compte du gouvernement norvégien. La compagnie est responsable de la gestion des satellites *AISSAT 1 et 2* et *NORSAT 1 et 2* qui sont dédiés à la supervision des transmissions AIS au nom de l'administration côtière norvégienne et de l'agence spatiale norvégienne. Ces satellites sont positionnés en orbite polaire à une hauteur orbitale de 600-650km, de telle sorte qu'à chaque rotation autour de la terre, ils se décalent pour couvrir toutes les latitudes. Le choix de cette orbite conduit

à une bien meilleure couverture des pôles par rapport à l'équateur. Cela conduit à une meilleure efficacité du modèle au niveau des pôles. Le jeu de données S-AIS proposé par Statsat pour le bloc Entraînement est composé de 126 480 112 (environ 126 millions) de messages AIS provenant de l'ensemble du globe, ce qui correspond aux messages collectés entre le 24 avril 2019 et 3 juin 2019 (environ 6 semaines de données).

### 3.3 Sélection et augmentation de caractéristiques

Notre approche vise la détection des déconnexions AIS volontaires pour les navires en mer (ce qui est illégal). Ainsi, partant du jeu de données S-AIS fourni par le bloc Statsat, pour chaque message AIS, nous sélectionnons les caractéristiques pertinentes que sont les champs de position ( $lat, lon$ ), l'horodatage ( $t$ ), et la vitesse ( $s$ ).

Par ailleurs, nous enrichissons les caractéristiques avec  $\Delta t$  la différence de temps par rapport au message précédent du même navire,  $\Delta D_V$  la différence en mètres sur l'axe vertical avec le message précédant du même navire,  $\Delta D_H$  la différence en mètres sur l'axe horizontal,  $D_P$  la distance au port,  $T_D$  la seconde du jour (entre 0 et 86 400). Il est à noter que :

- $\Delta t$  et  $T_D$  permettent d'avoir une meilleure précision

sur la dimension temporelle que celle donnée par l'horodatage  $t$  ;

- $\Delta D_V$  et  $\Delta D_H$  améliorent la compréhension de la position sur de petites distances ;
- $D_P$  permet de filtrer les échantillons pertinents ;

Ainsi, le vecteur de caractéristiques correspondant à un message prend la forme suivante :

$$x = [t, lat, lon, s, \Delta t, \Delta D_V, \Delta D_H, D_P, T_D]$$

### 3.4 Trajectoire ou Échantillon

Un échantillon ( $\mathcal{E}$ ) est une séquence temporelle de  $\mathcal{T}$  messages successifs d'un navire  $\mathcal{E} = [x_1, x_2, \dots, x_T]$ . Nous l'appelons également *trajectoire* car cela correspond à la trajectoire d'un navire dans une fenêtre temporelle. Il est à noter qu'il n'y a aucune condition posée sur les valeurs maximales de  $\Delta t$  et  $\Delta D$ , entre deux messages successifs au sein d'une trajectoire.

### 3.5 Séparation des jeux de données

Les jeux de données utilisés pour l'entraînement, la validation et le test du modèle sont séparés en fonction de leur date. Nous aurions pu aussi séparer par navire mais la séparation temporelle a l'avantage de se rapprocher de l'utilisation qui est faite du modèle lors de son déploiement. L'objectif est d'effectuer des prédictions en s'appuyant sur les données passées de la trajectoire d'un navire. La table 1 décrit le volume des jeux de données ainsi que leur division temporelle.

### 3.6 Sélection et traitement des trajectoires

La figure 3 indique la répartition des messages AIS à la surface du globe. Pour atteindre un bon équilibre du jeu de données à partir de l'ensemble de ces messages, nous extrayons aléatoirement et de manière équitable 1) des trajectoires de navires présentant une continuité de transmission, et 2) des trajectoires de navires présentant une déconnexion AIS. Un seuil arbitraire de continuité a été fixé à 10 minutes pour l'entraînement de notre modèle car nous avons constaté que les déconnexions AIS sont assez courantes dans une période inférieure mais qu'elles deviennent anormales au-delà de 10 minutes. De manière empirique, nous avons constaté que ce seuil de 10 minutes est pertinent pour l'entraînement mais rien n'empêche, lors de la mise en production, de se concentrer sur des déconnexions AIS plus longues (e.g., 1 heure).

Pour la sélection des trajectoires dans le jeu de données, nous posons deux conditions :

1. l'historique des messages AIS du navire concerné doit être composé d'au moins 50 messages. Ceci exclut les bateaux qui viennent de rentrer dans la zone d'observation temporelle de notre jeu de données et qui sont peu pertinents pour l'entraînement du modèle. La limite de 50 messages a été choisie en utilisant les statistiques de la figure 5, on s'aperçoit que de 50 à 100 messages un navire peut parcourir une grande distance et que l'on rencontre régulièrement une longue déconnexion permettant ainsi au modèle de comprendre des situations de déconnexions ;

2. la trajectoire du navire doit être à plus de 5 kilomètres d'un port. Cela permet d'éliminer les exemples de déconnexions AIS volontaires (légal) qui ont lieu dans les ports. De plus, il est relativement facile pour la garde côtière de contrôler les bateaux sans transmission AIS qui se situent dans les ports. Pour déterminer la distance au port le plus proche, nous calculons la distance de l'arc entre deux points sur une sphère<sup>6</sup> avec une base de données de 20 756 ports fournie par l'organisation "global fishing watch"<sup>7</sup>.

Étant donné  $R$  le rayon de la terre (6371m), deux points  $p_1$  (resp.  $p_2$ ) ayant pour latitude  $lat_1$  (resp.  $lat_2$ ) et longitude  $lon_1$  (resp.  $lon_2$ ) (en radians), la formule utilisée pour le calcul de la distance  $d$  de l'arc sur un grand cercle est la formule d'Harvesine :

$$d(p_1, p_2) = 2R * \arcsin(\sqrt{\alpha + \beta})$$

$$\alpha = \sin^2\left(\frac{lat_1 - lat_2}{2}\right)$$

$$\beta = \cos(lat_1) * \cos(lat_2) * \sin^2\left(\frac{lon_1 - lon_2}{2}\right)$$

## 4 Modèle d'apprentissage et entraînement

### 4.1 Architecture du modèle

La figure 4 donne un aperçu de l'architecture générale utilisée pour notre modèle d'apprentissage auto-supervisé. Tout d'abord, la partie pré-traitement divise l'entrée  $\mathcal{E}$  en deux vecteurs,  $\mathcal{V}_H$  l'historique des messages AIS qui contient les informations relatives au précédent message et  $\mathcal{V}_L$  la position la plus récente. Ensuite,  $\mathcal{V}_H$  et  $\mathcal{V}_L$  sont normalisés par la couche  $\mathcal{N}$ . La division de l'entrée a pour objectif d'avoir un encodage qui prend en compte les possibles faibles distances que l'on peut retrouver entre deux messages, tout en conservant une très grande précision sur la position de la déconnexion sur le globe.  $\mathcal{V}_H$  est donné en entrée à deux blocs "transformers" successifs avec l'objectif de faire des relations entre des messages plus ou moins lointains et d'extraire les informations importantes de la trajectoire. En revanche,  $\mathcal{V}_L$  est mis de côté avant d'être concaténé avec la représentation  $\mathcal{R}$  de la trajectoire en sortie du transformer. Un ensemble de couches denses détermine si oui ou non, un message devrait être reçu dans les dix prochaines minutes. Au total, Le modèle  $\mathcal{M}$  est composé de 4 690 021 de paramètres entraînaibles.

Dans la figure 4,  $\mathcal{B}$  représente la taille des batchs (groupe de données) et  $\mathcal{W}$  la fenêtre de messages utilisés pour l'entraînement. Les résultats optimaux ont été obtenus avec  $\mathcal{B} = 128$ ,  $\mathcal{W} = 100$  et  $\mathcal{R} = 64$ .

6. La distance de l'arc sur une sphère n'est pas aussi précise que les formules de Vincenty en géodésie puisque la terre n'est pas une sphère parfaite, mais son calcul a l'avantage d'être vectorisable, ce qui est nécessaire dans le cas où le jeu de données contient un très grand nombre de trajectoires.

7. <https://globalfishingwatch.org/datasets-and-code/anchorages/>

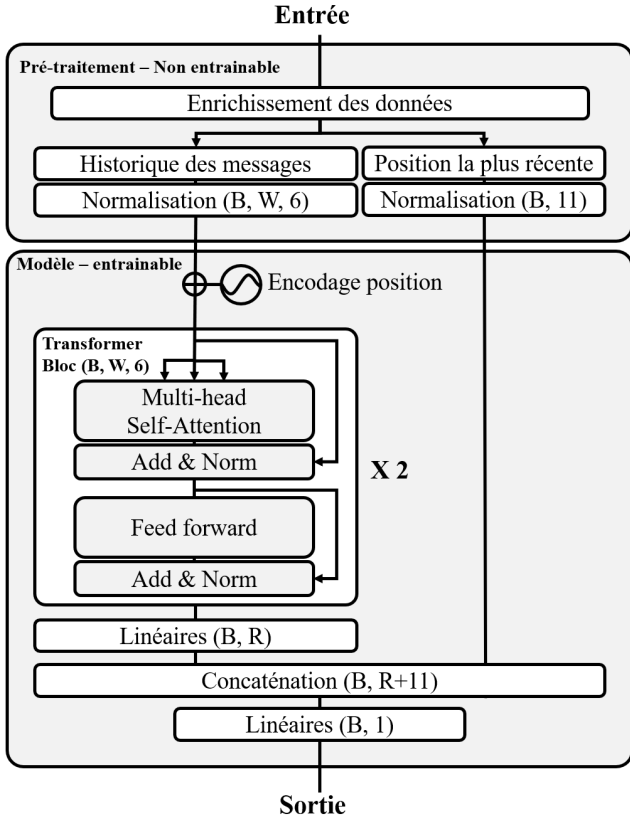


FIGURE 4 – Architecture du modèle

## 4.2 Préparation des données

Une partie importante de notre travail a été de trouver un encodage et une normalisation, nous permettant de gérer à la fois les grandes distances qui peuvent séparer deux messages tout en gardant une précision suffisante pour les messages espacés seulement de quelques secondes. La figure 5 montre la diversité des trajectoires et des messages. On peut noter par exemple que 10% des trajectoires dure moins de 100 minutes quand 10% des trajectoires durent plus de 40 heures. Dans ce dernier cas, les satellites n'ont été capable de capturer que 100 messages dans les 40 dernières heures. Cette grande durée est principalement due à l'écart entre quelques messages. En effet, on constate que 10% des trajectoires contiennent au moins un coupure de plus de 600 minutes (10h). La même analyse peut être effectuée sur les distances. De plus, nous avons travaillé à rendre le modèle générique pour qu'il puisse être utilisé pour analyser les trajectoires de n'importe quel instant et région du globe.

### 4.2.1 Historique des messages

Pour rendre le traitement de  $\mathcal{V}_H$  le plus générique possible,  $\mathcal{N}$  supprime la position absolue représentée par la latitude et la longitude ainsi que le temps absolu représenté par l'horodatage. Le "transformer model" s'appuie sur la différence temporelle et la différence de distance avec le message précédent pour construire la représentation de la trajectoire. De plus, la seconde de la journée  $\mathcal{S}_D$  est ajoutée pour renforcer la détection de motifs temporels. Une normalisation cy-

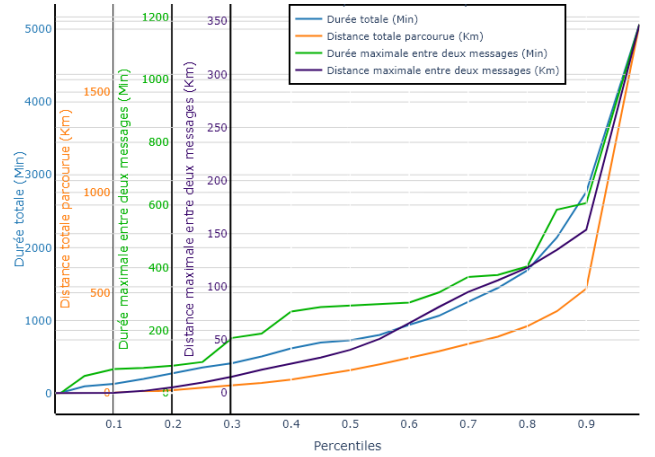


FIGURE 5 – Statistique des échantillons

clique  $\mathcal{N}_C$  (éq. 1) est appliquée sur les champs cycliques tels que la  $\mathcal{S}_D$ , une normalisation linéaire  $\mathcal{N}_L$  (éq. 2) est appliquée sur les valeurs limites telles que la vitesse  $\mathcal{V}$ . Pour le  $\Delta t$ ,  $\Delta D_V$  et  $\Delta D_H$ , un logarithme est appliqué (éq. 3).

### 4.2.2 Position la plus récente

Le vecteur  $\mathcal{N}(\mathcal{V}_P)$  est composé de 11 valeurs permettant au modèle d'avoir une précision maximale sur la position du navire lors de la déconnexion.  $\mathcal{N}$  décompose la latitude et la longitude en Degré-Minute-Seconde puis les normalise cycliquement pour conserver une continuité (éq. 4), par exemple lorsqu'un navire passe de la longitude 180 à -180. À noter que le degré de la latitude est encodé linéairement puisque un navire ne peut pas passer de la latitude -90 à 90.

$$\mathcal{N}_C = \left[ \sin \frac{2 * \pi * (x - \min x)}{\max x - \min x}, \cos \frac{2 * \pi * (x - \min x)}{\max x - \min x} \right] \quad (1)$$

$$\mathcal{N}_L = \frac{(x - \min x)}{\max x - \min x} \quad (2)$$

$$\mathcal{N}(\mathcal{V}_H) = [\log(\Delta t), \log(\Delta D_V), \log(\Delta D_H), \mathcal{N}_C(\mathcal{S}_D), \mathcal{N}_L(\mathcal{V})] \quad (3)$$

$$\mathcal{N}(\mathcal{V}_P) = [\mathcal{N}_L(lat_{Deg}), \mathcal{N}_C(lat_{Min}), \mathcal{N}_C(lat_{Sec}), \mathcal{N}_C(lon_{Deg}), \mathcal{N}_C(lon_{Min}), \mathcal{N}_C(lon_{Sec}), \dots] \quad (4)$$

## 4.3 Détection de déconnexions AIS volontaires

Notre modèle prédit si la réception d'un message AIS du navire est attendue ou non, dans les 10 prochaines minutes. Pour détecter les situations suspectes, nous devons comparer les sorties attendues et les prédictions du modèle, avec deux situations identifiées :

1. Les trajectoires où le navire émet un message AIS dans les 10 minutes sont utiles pour entraîner le modèle mais pas pour la prédiction. En effet, si le modèle prédit qu'un message est attendu, alors le modèle est conforme et si le modèle prédit qu'aucun message n'est attendu, alors une erreur (certaine) du modèle est présente.



- Les trajectoires où le navire n'émet pas de message AIS dans les 10 minutes sont intéressantes pour la prédiction. En effet, si le modèle prédit qu'aucun message n'est attendu alors le modèle est conforme et la déconnexion AIS est soit involontaire (e.g. panne du transpondeur), soit légale (e.g., navire au mouillage). Par contre, si le modèle prédit qu'un message est attendu alors on a soit une erreur du modèle, soit une déconnexion AIS illégale. Ce sont ces cas-là qui sont les plus pertinents pour la garde-côtière, qui peuvent alors procéder à une analyse détaillée de confirmation ou infirmation. Dans ce dernier cas, l'historique des positions du navire est rapporté.

Dans tous les cas, il est crucial de disposer d'un modèle présentant une très haute précision afin de ne retourner que des trajectoires suspectes intéressantes aux garde-côtes.

## 5 Résultats et discussion

### 5.1 Entraînement du modèle

Une chaîne d'intégration continue a été mise en place entre le serveur de pré-traitement, celui d'entraînement et celui de prédictions en utilisant gitlab-ci pour l'orchestration et DVC<sup>8</sup> pour la mise en version des jeux de données. L'infrastructure expérimentale de recherche eX3<sup>9</sup> nous fournit un environnement de calcul haute performance. L'infrastructure inclut un système DGX-2<sup>10</sup> composé de 16 cartes graphiques NVIDIA Tesla V100. Pour notre entraînement, une seule de ces cartes nous permet d'entraîner le modèle à une vitesse de une époque toutes les 7s, chaque époque étant composée de 625 pas. Le modèle atteint sa précision maximale après 100 époques mais, afin d'assurer au modèle une convergence garantie, nous l'entraînons sur 1000 époques. Cela amène le temps d'entraînement à 7100s ±150s (soit environ 2h). Aucun sur-apprentissage n'est observé et cela est particulièrement important dans le contexte d'une mise en production avec un apprentissage continu. En effet, l'objectif est d'utiliser les données collectées au jour le jour pour améliorer la précision du modèle.

### 5.2 Résultats

Les performances obtenues par le modèle pour la prédiction de la réception d'un message dans les dix prochaines minutes sont excellentes. La figure 6 permet d'observer l'évolution de la précision jusqu'à ce que celle-ci atteigne 99% après les 100 premières époques. Pour une meilleure évaluation, le modèle a été entraîné 10 fois sur 1000 époques avec à chaque fois une initialisation aléatoire des paramètres. La figure 7 décrit les précisions finales obtenues sur le jeu d'entraînement, de validations et de test. La précision sur le jeu de test est la plus importante, et nous pouvons

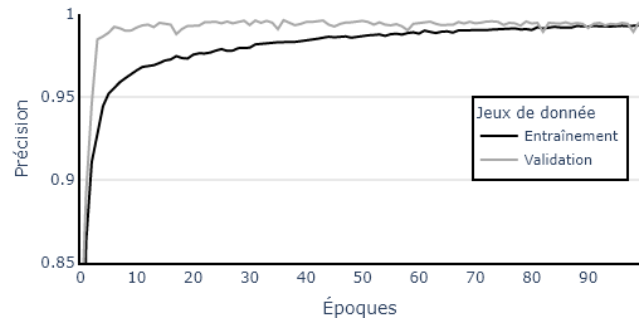


FIGURE 6 – Précision du modèle sur les 100 premières époques d'entraînement

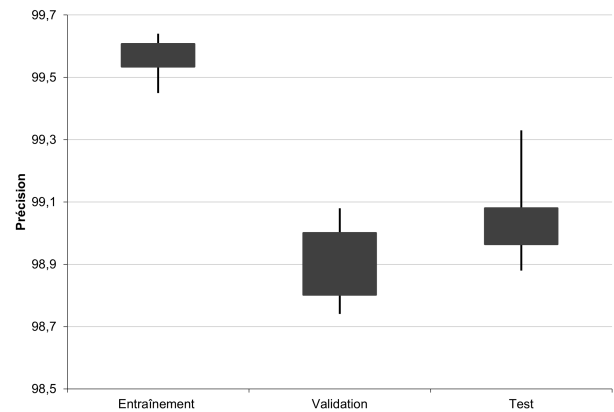


FIGURE 7 – Statistiques de précision sur 10 exécutions

observer une précision de 99% en moyenne avec un maximum atteint de 99.3%. Il est important de rappeler ici que la précision du modèle correspond à la prédiction de réception d'un message dans les 10 minutes et non pas à la prédiction de déconnexion AIS volontaires. Il n'existe pas, à notre connaissance, de jeux de données permettant de tester notre modèle précisément sur cette tâche. Par contre, l'analyse de la matrice de confusion (Table 2) confirme les bonnes performances du modèle. En effet, l'écart entre les faux négatifs et les faux positifs est significatif.

- **Faux négatif (F. nég.)** : le modèle prédit la non-réception d'un message et pourtant un message est reçu ;
- **Faux positif (F. pos.)** : le modèle prédit la réception d'un message mais aucun message n'est reçu.

TABLE 2 – Matrice de confusion, jeu de données de test - Doit-on recevoir un message du navire dans les 10 minutes ?

Prédiction \ Annotation	Oui	Non
	Oui	4887
Non	1 (F. nég.)	4992

Sur le jeu de test, un seul faux négatif est détecté. Cette erreur de classification ne peut provenir que d'une erreur du modèle. Par contre, les 104 faux positifs détectés peuvent provenir d'une erreur du modèle mais aussi d'une coupure

8. Data Version Control

9. <https://www.ex3.simula.no/>

10. <https://www.nvidia.com/content/dam/en-zz/Solutions/Data-Center/dgx-2/dgx-2-print-datasheet-738070-nvidia-a4-web-uk.pdf>

anormale du signal. Cet écart consolide les résultats.

### 5.3 Détection des déconnexions AIS suspectes sur le jeu de données de tests

Sur la courte période de tests dont nous disposons (1 semaine), nous avons extrait tous les cas où la connexion avec un navire est perdue pendant plus de 2h, cela représente 160 574 cas. Le seuil de 2h a été choisi pour limiter le nombre d'exemples à tester. De plus, 2h de coupure sont suffisantes pour effectuer des infractions mais restent un évènement fréquent dans le contexte d'une capture par satellite. Sur ces 160 574 cas, le modèle a permis de classer 12 navires qui aurait dû envoyer un message dans les 10 premières minutes de la déconnexion. Nous pensons que ce filtrage est une aide précieuse pour les garde-côtes qui peuvent concentrer leur attention sur ces cas suspects. Parmi ces 12 déconnexions suspectes, 10 ont eu lieu après avoir jeté l'ancre dans de petits ports, non répertoriés dans le fichier que nous utilisons pour le filtrage des ports. La figure 8 montre une une de ces dix trajectoires qui sont détectées par notre modèle, mais non illégales. Par contre, les deux autres trajectoires sont suspectes, comme indiquées dans la figure 9. Les deux trajectoires entrent dans une zone économique exclusive avant que l'on perde leur trace. Bien sûr, rien ne nous permet d'affirmer avec certitude que ces déconnexions sont volontaires mais elles nécessitent de lancer des investigations approfondies de la part de la garde-côtière.

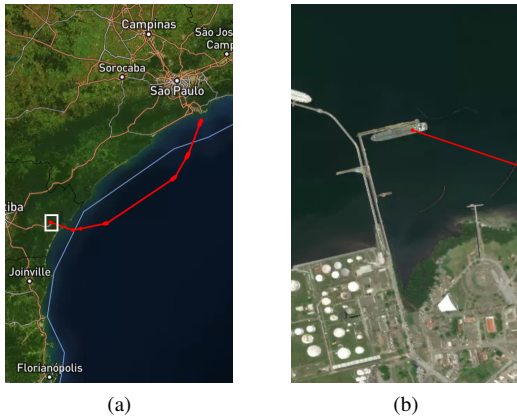


FIGURE 8 – Un exemple de déconnexion volontaire détectée mais non illégale. Image (b) est le zoom de l'image (a)

### 5.4 Discussion

Les déconnexions AIS volontaires (illégales) sont des évènements très rares et la quantité de données utilisée pour le test permet de confirmer que le modèle est bien capable de prédire si un navire doit recevoir un message dans les 10 prochaines minutes. Cependant, à ce stade de nos travaux, il ne nous est pas encore possible de fournir un score de précision pour la détection des déconnexions anormales. La matrice de confusion donne une bonne idée de la validité du modèle mais il faut tester le modèle sur un jeu de données plus grand (c'est-à-dire sur une plage de temps plus

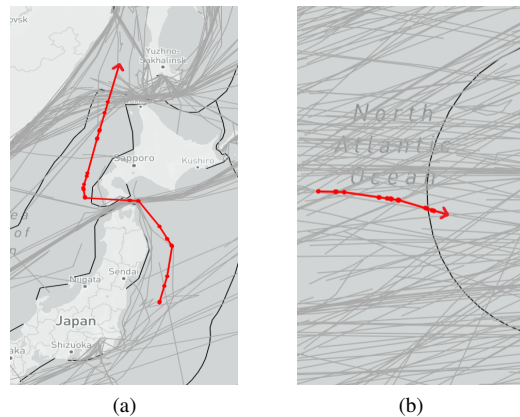


FIGURE 9 – Deux exemples de déconnexion suspecte d'après notre modèle

longue) pour apprendre des motifs de déconnexion AIS. Pour l'instant, le modèle n'est entraîné que sur des données entre avril et juin 2020, c'est-à-dire des données du printemps, ce qui peut créer un biais par rapport à des données d'autres saisons. Nous travaillons à intégrer une année complète de données dans l'expérience. De plus, le modèle ne détecte que les déconnexions AIS qui ont lieu au moment où un des satellites survole la zone. Cela rend le modèle très efficace près des pôles mais moins à l'équateur car les satellites couvrent moins souvent cette zone. Cette faiblesse inhérente à notre jeu de données doit faire l'objet d'une attention particulière.

## 6 Conclusion et perspectives

Les résultats sur la prédiction de réception de messages sont très encourageants sur nos données de test (99% de précision), ce qui rend cette approche par apprentissage auto-supervisé particulièrement intéressante pour détecter les déconnexions AIS volontaires en temps réel. La mise à disposition prochaine d'un jeu de données sur une année entière nous permettra d'expérimenter notre modèle sur d'autres périodes de l'année, ce qui réduira le biais potentiel lié à l'utilisation de données du printemps uniquement. Par ailleurs, l'apprentissage de motifs de déconnexions AIS telles que le franchissement de frontière d'une zone d'exclusion économique sera possible. En parallèle, nous visons également l'entraînement de modèles auto-supervisés pour d'autres tâches de détection de déconnexions illégales. En particulier, nous nous intéressons aux rendez-vous en mer pour le transbordement de cargaisons illégales et à l'identification de navires en mer.

## Remerciements

Les résultats présentés dans cet article ont été financés par le "Research Council of Norway" (RCN) dans le cadre du projet "AI-driven testing of false data injection attacks against transport infrastructure (TSAR)" [#287893]. En outre, les résultats ont été obtenus grâce à l'infrastructure expérimentale eX3 (exploration du calcul exascale), qui est également financée par le RCN [# 270053]. Les données S-AIS utili-

sées ont été généreusement mis à disposition par l'administration de la garde-côte Norvégienne.

## Références

- [1] Bryan Auslander, Kalyan Moy Gupta, and David William Aha. Maritime Threat Detection Using Probabilistic Graphical Models. In *Proc. of the 25th Int. FLAIRS Conf.*, May 2012.
- [2] Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. Neural machine translation by jointly learning to align and translate. In Yoshua Bengio and Yann LeCun, editors, *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*, 2015.
- [3] Jianpeng Cheng, Li Dong, and Mirella Lapata. Long short-term memory-networks for machine reading. In Jian Su, Xavier Carreras, and Kevin Duh, editors, *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, EMNLP 2016, Austin, Texas, USA, November 1-4, 2016*, pages 551–561. The Association for Computational Linguistics, 2016.
- [4] Dominik Filipiak, Milena Stróżyńska, Krzysztof Wecel, and Witold Abramowicz. Big data for anomaly detection in maritime surveillance : spatial AIS data analysis for tankers. *Scientific Journal of Polish Naval Academy*, 215(4) :5–28, 2018.
- [5] Natalie Fridman, Doron Amir, Yinon Douchan, and Noa Agmon. Satellite Detection of Moving Vessels in Marine Environments. In *Proc. of the AAAI Conf.*, volume 33, pages 9452–9459, July 2019.
- [6] Kwang-II Kim and Keon Myung Lee. Deep Learning-Based Caution Area Traffic Prediction with Automatic Identification System Sensor Data. *Sensors*, 18(9) :3172, September 2018.
- [7] Lacey Malarky and Beth Lowell. Avoiding detection : Global case studies of possible AIS avoidance. Technical report, OCEANA, Oct. 2020.
- [8] Fabio Mazzarella, Michele Vespe, Dario Tarchi, Giuseppe Aulicino, and Antonio Vollero. AIS reception characterisation for AIS on/off anomaly detection. In *19th Int. Conf. on Information Fusion (FUSION 2016), Heidelberg, Germany, July 5-8, 2016. IEEE 2016*, 2016.
- [9] Duong Nguyen, Rodolphe Vadaine, Guillaume Hajdich, René Garello, and Ronan Fablet. A Multi-task Deep Learning Architecture for Maritime Surveillance using AIS Data Streams. *IEEE 5th Int. Conf. on Data Science and Advanced Analytics (DSAA)*, pages 331–340, October 2018.
- [10] Duong Nguyen, Rodolphe Vadaine, Guillaume Hajdich, René Garello, and Ronan Fablet. GeoTrackNet-A Maritime Anomaly Detector using Probabilistic Neural Network Representation of AIS Tracks and A Contrario Detection. *IEEE Transactions on Intelligent Transportation Systems*, 2021.
- [11] Jaeyoon Park, Jungsam Lee, Katherine Seto, Timothy Hochberg, Brian A. Wong, Nathan A. Miller, Kenji Takasaki, Hiroshi Kubota, Yoshioki Oozeki, Sejal Doshi, Maya Midzik, Quentin Hanich, Brian Sullivan, Paul Woods, and David A. Kroodsma. Illuminating dark fishing fleets in north korea. 6(30) :eabb1197. Publisher : American Association for the Advancement of Science Section : Research Article.
- [12] Lokukaluge P. Perera, Paulo Oliveira, and C. Guedes Soares. Maritime Traffic Monitoring Based on Vessel Detection, Tracking, State Estimation, and Trajectory Prediction. *IEEE Transactions on Intelligent Transportation Systems*, 13(3) :1188–1200, September 2012.
- [13] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is All you Need. In *NeurIPS 2017*, pages 5998–6008, 2017.
- [14] Henri Weimerskirch, Julien Collet, Alexandre Corbeau, Adrien Pajot, Floran Hoarau, Cédric Marteau, Dominique Filippi, and Samantha C. Patrick. Ocean sentinel albatrosses locate illegal vessels and provide the first estimate of the extent of nondeclared fishing. 117(6) :3006–3014. Publisher : National Academy of Sciences Section : Biological Sciences.