

PRISME-CT

Vers une meilleure prévision de l'intensification des cyclones tropicaux grâce à l'intelligence artificielle

Adrien Gazo - DSM/CS/AGRO - Mickaël Lagavardan - DSO/DOA/IED

Parmi les techniques d'intelligence artificielle, le développement de modèles d'apprentissage profond permet d'envisager leur utilisation opérationnelle pour la prévision immédiate de l'intensification rapide des cyclones dans le cadre de la mission de surveillance cyclonique du CMRS de la Réunion.

Introduction

Parmi les phénomènes météorologiques les plus redoutables, les cyclones posent un défi considérable aux populations et aux autorités en raison de leur potentiel destructeur, en particulier lorsqu'ils subissent une "**intensification rapide**" (RI). Ce terme désigne une augmentation soudaine de la vitesse moyenne des vents maximums de plus de 30 nœuds (kt) en moins de 24 heures (1 kt = 1,8 km/h). Cette intensification rapide rend la gestion des catastrophes particulièrement complexe lorsqu'elle se produit près des terres (Figure 1).

Dans le bassin Sud-Ouest de l'Océan Indien, le **CMRS** (Centre Météorologique Régional Spécialisé) des Cyclones Tropicaux de La Réunion est chargé de surveiller ces systèmes à l'aide de divers outils. Ces dernières années, les avancées en intelligence artificielle et en particulier en apprentissage profond offrent de nouvelles perspectives pour améliorer la prévision des cyclones.



Figure 1. Photo d'une rue de Roseau (Dominique) après le passage de l'ouragan Maria en 2017. Source : Roosevelt Skerrit.

Le projet et les outils

Le jeu de données **TC-PRIMED**, récemment mis à disposition par la NOAA et l'Université d'État du Colorado (CSU) regroupe de nombreuses données sur les cyclones et en particulier des images satellites associées à des données environnementales sur une période de 1998 à 2021 (Figure 2). Cette base de données permet une application facilitée de ces nouveaux outils dans la classification des changements d'intensité cyclonique.

Le projet **PRISME-CT** (PRévision à base d'Images Satellites Micro-onDEs de Cyclones Tropicaux), développé durant un peu plus de trois mois dans le cadre d'un stage de fin d'études au sein du **LACY** (Laboratoire de l'Atmosphère et des CYclones), vise à exploiter ce potentiel pour développer des outils de prévision avancés, avec l'objectif de renforcer les capacités opérationnelles du CMRS. Ce projet se base d'une part sur les images satellites micro-ondes à 89,0 GHz qui représentent la structure interne des cyclones, mais aussi sur de multiples données environnementales (intensité du cyclone, cisaillement vertical, humidité relative...).

Une fois récupérées, les données ont été traitées (centrage sur le système, uniformisation de la taille des images, filtrage de valeurs aberrantes, correction par interpolation, conversion en valeurs numériques...) puis normalisées ; ces traitements préliminaires constituent une étape importante en apprentissage auto-

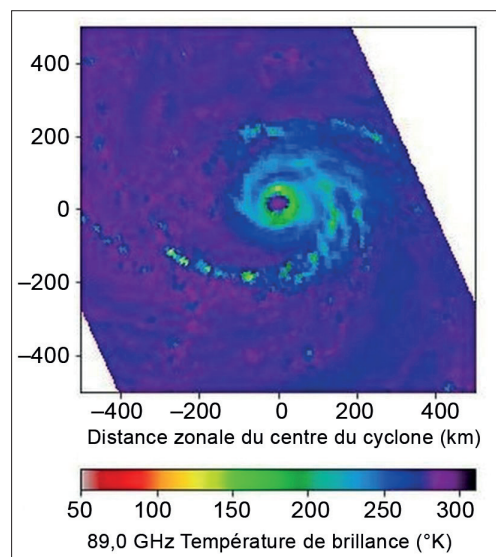


Figure 2. Exemple de données TC-PRIMED (Rayonnement micro-ondes à 89,0 GHz). Source : TC-PRIMED, Colorado State University.

matique pour améliorer la convergence des algorithmes et les performances des modèles. Enfin, des séquences de 5 images par pas de 6 h ont été formées ainsi que les séquences de données environnementales associées. Ces séquences ont ensuite été labellisées d'une classe de changement d'intensité à + 24 h parmi 7 (Figure 3, Tableau 1).

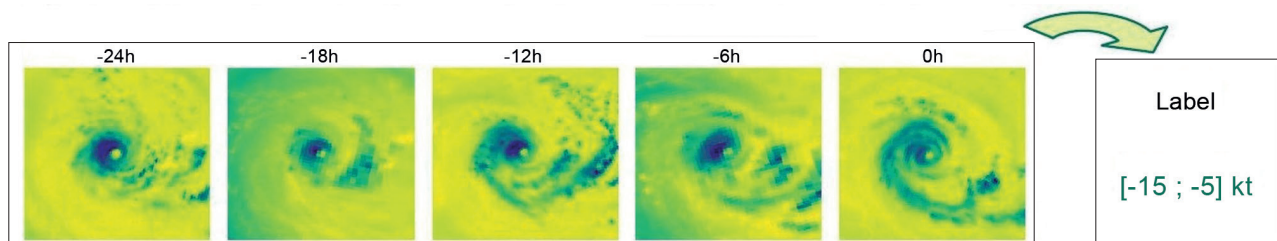


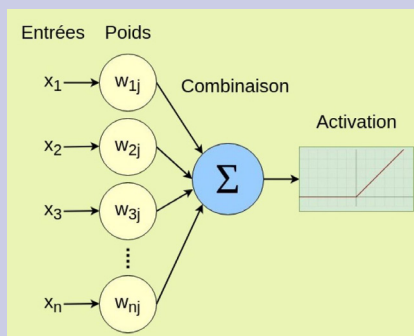
Figure 3. Exemple de séquence d'images créée et labellisée.

Pour entraîner et évaluer un réseau de neurones, il est nécessaire d'avoir à disposition un jeu de données d'entraînement et un jeu d'évaluation. Ces deux jeux doivent absolument être indépendants afin que l'évaluation se déroule correctement, sans que le modèle ne fasse usage de données déjà apprises. Pour que chaque classe de changement d'intensité soit bien représentée dans le modèle, étant donné que la classe d'intensification rapide ne contient que peu d'éléments, il a été décidé d'avoir la même quantité de données pour chaque classe dans les deux jeux de données.

< - 22 kt	> 30 kt
- 22 kt à - 15 kt	15 kt à 30 kt
- 15 kt à - 5 kt	5 kt à 15 kt
< - 5 kt à 5 kt	

Tableau 1. Les 7 classes de changement d'intensité en 24 h. En vert : affaiblissement ou stagnation ; en orange : intensification ; en rouge : intensification rapide (RI).

Qu'est-ce qu'un réseau de neurones ? Comment apprend-t-il ?



Représentation du mode de fonctionnement d'un neurone artificiel.

Un **neurone** artificiel est l'élément de base des réseaux de neurones. Chaque neurone reçoit des signaux d'entrée et leur affecte des poids modifiant leur importance. Les entrées sont multipliées par ces poids puis additionnées. La somme pondérée obtenue passe à travers une fonction d'activation non-linéaire qui détermine la valeur en sortie du neurone.

Une **couche de neurones** est constituée d'un regroupement de plusieurs neurones. Il est possible de superposer ces couches et de les relier en connectant les sorties de tous les neurones d'une couche à chaque neurone de la couche suivante. On les appelle alors des couches denses de neurones.

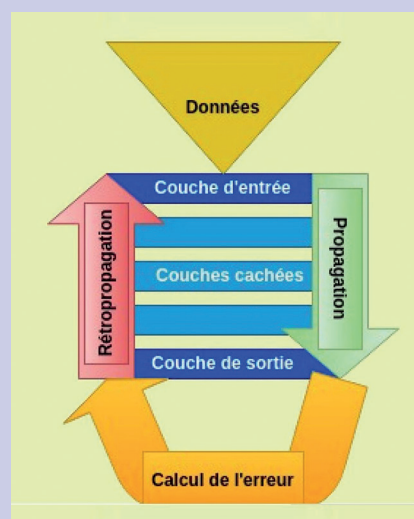
Le **réseau de neurones** ainsi obtenu est capable d'apprendre et de modéliser des relations complexes dans les données, en particulier les relations non-linéaires.

Durant la première phase de l'apprentissage, la couche d'entrée reçoit les données. Puis chaque neurone des couches suivantes reçoit les signaux de la couche précédente (couches cachées), jusqu'à atteindre la couche de sortie qui effectue la prévision. Cette phase est appelée la phase de **propagation**.

L'**erreur** entre la classe prévue et la classe réelle est calculée puis propagée dans le sens inverse à travers le réseau (c'est la phase de **rétropropagation**). Les poids sont alors ajustés en fonction de l'erreur.

Ces étapes sont répétées pour de nombreux **cycles**, chacun utilisant l'ensemble des données d'entraînement. L'erreur est ainsi minimisée au fil des cycles.

Au fur et à mesure de l'apprentissage, les performances du modèle sont **estimées** sur les données d'évaluation à l'aide de plusieurs métriques, ce qui permet de terminer l'apprentissage au moment où le modèle arrive le mieux à **généraliser** ses prévisions.



Fonctionnement de l'apprentissage d'un réseau de neurones.

La méthodologie

La démarche suivie s'articule autour de trois axes, à savoir l'étude de modèles de classification à partir des images, des données d'environnement, et enfin des deux combinées. Les résultats du modèle combinant images et données environnementales se sont révélés insatisfaisants ; seuls les résultats du meilleur modèle pour les images d'une part, et pour les données environnementales d'autre part, sont présentés dans cet article. Chacun de ces modèles est comparé à un modèle de référence permettant d'évaluer et quantifier l'apport des modèles plus complexes.

Le modèle de référence pour les images est un CNN (Convolutional Neural Network), modèle d'apprentissage profond classique pour la classification d'images basé sur des filtres de convolution pouvant extraire les structures caractéristiques d'une image. Le modèle de référence pour les données environnementales est un modèle de régression logistique, (modèle d'apprentissage automatique similaire à un modèle de régression linéaire mais appliqué à la classification de données).

Avant de présenter les modèles et de les comparer avec les modèles de référence, il est important de définir les quatre scores utilisés pour évaluer leurs performances. La précision globale (ratio du nombre de bonnes prévisions sur le nombre total de prévisions) est utilisée pour déterminer la performance globale du modèle. Trois autres scores sont utilisés pour déterminer les performances du modèle concernant les prévisions d'intensification rapide :

- Le score F1-RI, moyenne harmonique de la précision (capacité du modèle à distinguer la classe RI des autres classes) et du rappel (capacité du modèle à bien prévoir la classe RI lorsqu'il y a une RI).
- Le taux de fausses alarmes RI (FAR-RI), lorsqu'il est faible, indique la capacité du modèle à écarter le scénario de RI lorsqu'il n'y a effectivement pas de RI.
- Le taux de non détection RI (NDR-RI), lorsqu'il est faible, indique la capacité du modèle à bien prévoir la classe RI lorsqu'il y a effectivement une RI.

Pour chacun de ces scores, un pourcentage d'amélioration par rapport aux scores du modèle de référence sera calculé (Skill) et permettra une comparaison précise des modèles.

Traitement des images avec le modèle ResNet18-LSTM

Le modèle comparé au CNN est aussi un réseau de neurones de type ResNet18, basé sur des filtres de convolution conçu pour faciliter l'apprentissage des modèles possédant un nombre important de couches (ici 18). Ces modèles dits profonds tendent à avoir une baisse de performance due à une mauvaise optimisation des poids lors de la rétropropagation. Mais les ResNet, qui utilisent des chemins de raccourci, permettant aux couches de copier leurs entrées directement à leurs sorties, offrent une solution à ce problème. Ce modèle est alors couplé à un composant nommé Long Short Term Memory (LSTM), permettant au modèle de traiter la séquence d'images de façon chronologique en détectant l'influence de chaque image sur les prochaines.

Ce modèle permet d'obtenir une amélioration de tous les scores par rapport au modèle de référence (tableau 2), en particulier pour le taux de non-détection : 60,9 % des RI réelles sont détectées. avec un taux de fausses alarmes relativement bas (17,1 % des séquences en réalité non-RI sont prévues RI par le modèle).

En évaluant le modèle sur sa capacité à discriminer les cas d'intensification des cas d'affaiblissement/stagnation (Figure 4a), on constate que le modèle détecte 72 % des intensifications correctement et prévoit de manière erronée une intensification dans seulement 17,6 % des cas réels d'affaiblissement/stagnation.

Scores (%) ResNet-LSTM		Skill / CNN%
Précision globale	36,3	+ 8,2
F1-RI	46,3	+ 7,6
FAR-RI	17,3	+ 0,6
NDR-RI	39,1	+ 15,9

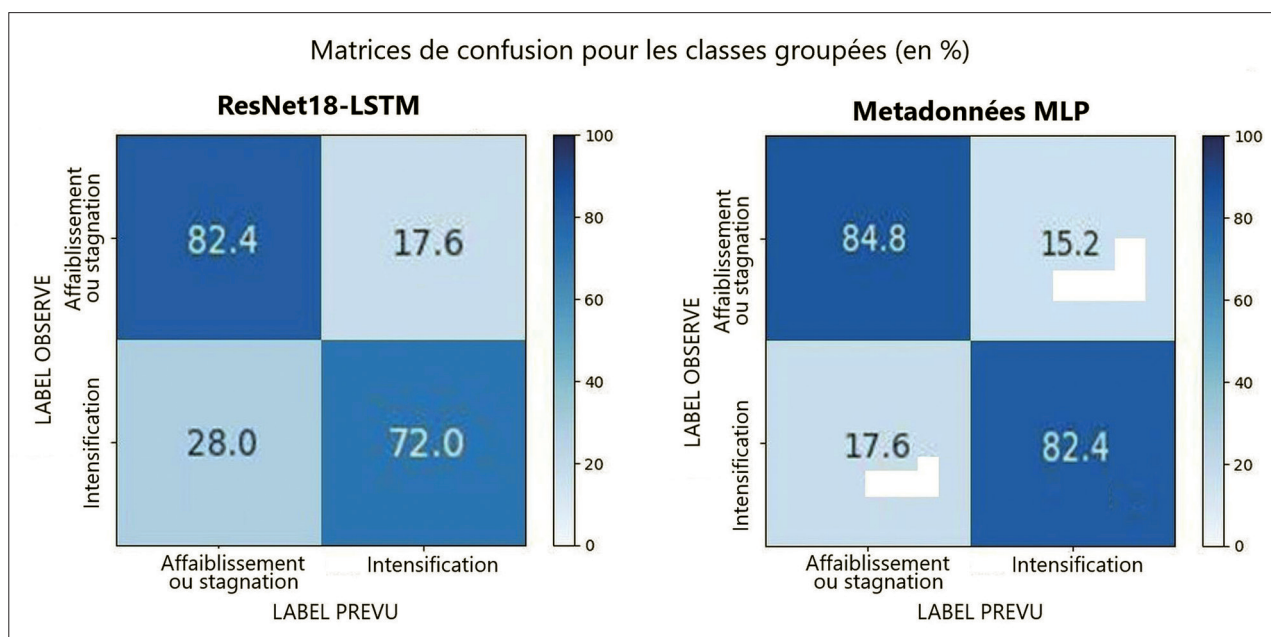
Tableau 2. Scores du modèle ResNet18-LSTM et pourcentage d'amélioration par rapport au modèle de référence CNN.

Traitement des données environnementales avec le modèle MLP

Le modèle MLP (Multi Layer Perceptron) comparé au modèle de régression logistique est un réseau de neurones à couches denses tel que décrit précédemment, possédant 5 couches de taille variable. Ce modèle permet une nette amélioration de tous les scores par rapport au modèle de référence (tableau 3), en particulier pour le taux de non-détection : 71,9 % des RI réelles sont détectées, ce qui améliore de 31,2 % les performances par rapport à la régression, avec un taux de fausses alarmes bas (10,9 % des séquences en réalité non-RI sont prévues RI par le modèle). Le modèle détecte (Figure 4b) 82,4 % des intensifications correctement et prévoit de manière erronée une intensification dans seulement 15,2 % des cas réels d'affaiblissement /stagnation.

Scores (%) MLP		Skill / Reglog %
Précision globale	46,5	+ 8,7
F1-RI	60,7	+ 19,8
FAR-RI	10,9	+ 9,9
NDR-RI	28,1	+ 31,2

Tableau 3. Scores du modèle MLP et pourcentage d'amélioration par rapport au modèle de régression logistique Reglog.



Figures 4. Matrice de confusion « intensification vs affaiblissement/stagnation » des modèles ResNet18-LSTM (4a à gauche) et MLP (4b à droite). Chaque case (i,j) représente pour une catégorie réelle i donnée le pourcentage de fois où le modèle a prévu la catégorie j.

Les modèles d'apprentissage profond se comportant comme une boîte noire, il est difficile d'interpréter la manière dont ils font leurs prévisions. Il est néanmoins possible d'utiliser des moyens permettant de fournir une part d'explicabilité. On peut par exemple récupérer pour les différents modèles basés sur les données environnementales tels que le MLP les données en moyenne les plus importantes pour effectuer leurs prévisions. Les résultats sont similaires pour ces modèles : la donnée la plus importante est l'intensité du cyclone en elle-même (plus particulièrement celle de la dernière image), suivie du cisaillement vertical (facteur inhibant l'intensification), de la température de la surface de la mer (qui peut être considérée comme une réserve d'énergie pour l'intensification des cyclones), et enfin l'humidité relative des différentes couches de la troposphère. Concernant les modèles basés sur les images, il serait inutile de chercher à déterminer l'importance de chaque pixel car les systèmes évoluent et diffèrent spatialement. Il est cependant possible pour une séquence donnée d'observer les endroits des images qui se sont révélés être importants pour la prévision du modèle. On appelle cet outil une carte de saillance (saliency map en anglais) (Figure 5). Cette carte représente la température de brillance normalisée dans le canal micro-ondes (à 89,0 GHz) recouverte par l'importance de chaque pixel qui est déterminée en récupérant et en normalisant les valeurs des gradients de l'image en entrée du modèle entraîné après propagation et rétropropagation.

Cet outil présente un fort potentiel opérationnel car ces cartes peuvent être utilisées par le prévisionniste pour le conforter dans son analyse et sa prévision, pour détecter des structures complexes (zones de convection profonde, intrusion d'une masse d'air sec dans le cyclone...) conduisant à une potentielle intensification rapide (RI) et donner ainsi une forme d'interprétabilité à la prévision du modèle.

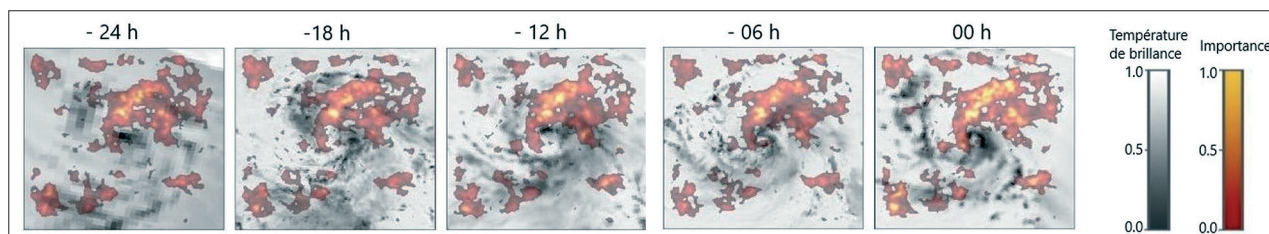


Figure 5. Carte de saillance superposée à une séquence RI bien prévue.

Bilan et perspectives

Ce projet aura abouti à l'élaboration de deux modèles d'apprentissage profond performants pour prévoir si un système cyclonique va s'intensifier rapidement durant les prochaines 24 h : un modèle ResNet18 LSTM classifiant des séquences d'images satellite micro-ondes, et un réseau de neurones MLP classifiant un ensemble de données environnementales telles que l'intensité du cyclone, le cisaillement vertical, la température de surface de la mer et l'humidité relative.

Bien que le MLP obtienne de meilleurs résultats que le ResNet18-LSTM, le délai plus rapide d'obtention des images satellites par rapport aux données environnementales provenant de réanalyses (ERA-5) tend à favoriser le modèle ResNet18-LSTM pour une utilisation en prévision à courte échéance.

Afin d'atteindre l'objectif d'utilisation opérationnelle, le développement de ce projet s'est fait de manière à ce que l'ensemble du code soit aisément réutilisable, modifiable et améliorable par quiconque ayant accès au GitLab de Météo-France (site de dépôt de code collaboratif). Ce projet présente plusieurs axes d'amélioration parmi lesquels :

- L'amélioration des modèles pour des données complexes, notamment les séquences RI où l'intensification n'est pas encore notablement amorcée, l'évaluation des modèles sur ce type de séquence ayant montré une baisse considérable des performances.
- L'exploration d'autres types de modèles plus récents pouvant être plus performants pour traiter des séquences temporelles (le code déposé sur GitLab contient certains modèles pour l'instant non évalués).
- La réalisation d'une classification à plus longue échéance (36 h, 48 h) afin de potentiellement utiliser les données environnementales pour la prévision.
- La comparaison avec les performances du modèle AROME. Des premières analyses ont montré que les modèles développés ici seraient plus performants, notamment sur le taux de fausses alarmes (FAR-RI) et le taux de non détection (NDR RI).

Remerciements

Ce stage n'aurait pas été le même et n'aurait pas abouti aux mêmes résultats sans l'aide et l'accompagnement de nombreuses personnes que nous tenons à remercier chaleureusement, en particulier nos encadrants Adrien Colomb et Quoc-Phi Duong.

Bibliographie

Razin, M. N., C. J. Slocum, J. A. Knaff, P. J. Brown, and M. M. Bell, 2023: Tropical Cyclone Precipitation, Infrared, Microwave, and Environmental Dataset (TC PRIMED). *Bull. Amer. Meteor. Soc.*, **104**, E1980-E1998, <https://doi.org/10.1175/BAMS-D-21-0052.1>.