



République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

Université Abderrahmane Mira – Béjaïa
Faculté des Sciences Exactes
Département d'Informatique

Mémoire de Fin d'Études

En vue de l'obtention du diplôme de Master en Informatique

Systeme d'aide à la surveillance sismique

Présenté par :
AIT HAMMOUDA Wissam

Encadré par :
YAICI Malika

Membres du jury :
AMROUNE Kamal
KHAMMARI Mohammed
BOUKERRAM Samira
ADEL Karima

Année universitaire : 2024-2025

Remerciements

Je tiens tout d'abord à exprimer ma profonde gratitude à mon encadrante, Mme Yaici Malika, pour sa disponibilité, son accompagnement rigoureux, et ses précieux conseils tout au long de ce travail. Son expertise et sa bienveillance ont grandement contribué à l'aboutissement de ce mémoire. J'adresse également mes remerciements aux membres jurys, pour avoir accepté d'évaluer ce travail, et pour leurs remarques constructives.

Je n'oublie pas ma famille, en particulier mes parents, pour leur amour, leurs encouragements constants, et leur soutien moral indéfectible, qui ont été essentiels à la réussite de mon cursus. Enfin, je remercie mes amies et toutes personnes qui, de près ou de loin, ont contribué à la réalisation de ce travail.

Contents

Remerciements	iii
List of Figures	ix
List of Tables	xi
INTRODUCTION GENERALE	1
1 Généralités sur les tremblements de terre	3
1.1 Introduction :	3
1.2 Qu'est-ce que un tremblement de terre ?	3
1.3 Caractéristiques principales :	3
1.3.1 Détermination de l'épicentre	4
1.3.2 Magnitude	4
1.3.3 Calcul de la magnitude :	4
1.3.4 Intensités d'un séisme	5
1.3.5 Mesure d'un tremblement de terre selon RICHTER	5
1.3.6 Propagation des ondes sismiques :	5
1.4 Appareils de mesure :	7
1.4.1 Le Sismographe :	7
1.4.2 Le sismomètre :	7
1.4.3 Capteur sismique :	8
1.4.4 Réseaux de capteurs GNSS (GPS) :	9
1.4.5 Historique des plus grands séismes	9
1.5 Méthode de Surveillance	9
1.5.1 Ancienne méthode chinoise	9
1.5.2 Méthode moderne	9
1.5.3 Méthode de prédiction	10
1.5.4 Prédiction à long terme (sur plusieurs années)	10
1.5.5 Prédiction à moyen terme (sur plusieurs mois)	10
1.5.6 Prédiction à court terme (inférieur à quelques jours)	10
1.5.7 Prédiction à moyen terme (sur plusieurs mois)	10
1.5.8 Prédiction à court terme (inférieur à quelques jours)	11
1.6 Conclusion	11

CONTENTS

2	IA appliquée à la surveillance et à la détection des séismes	13
2.1	L'utilisation de L'IA dans la prédiction de tremblement de terre	13
2.2	Effets et prévention :	13
2.3	Intégration aux systèmes d'alerte :	13
2.3.1	LIDAR :	14
2.3.2	Drones :	14
2.3.3	Robots sous marins :	16
2.3.4	Robots terrestres	17
2.4	Méthodes de Deep Learning pour la Prédiction Sismique	18
2.4.1	Introduction	18
2.4.2	Prédiction des Répliques Sismiques avec des LSTM	18
2.4.3	Détection d'Anomalies Sismiques avec des Auto encodeurs	19
2.4.4	Prédiction de la Magnitude des Séismes avec des CNN	20
2.4.5	Modélisation des Interactions entre Failles avec des GNN	20
2.4.6	Prédiction des Séismes avec des Modèles Hybrides (CNN + LSTM)	21
2.4.7	Apprentissage par Transfert pour la Prédiction Sismique	21
2.4.8	Gated Recurrent Unit	21
2.5	Mesures d'évaluation	22
2.5.1	Accuracy :	22
2.5.2	Matrice de confusion :	22
2.5.3	Recall :	22
2.5.4	Le F1-score	22
2.5.5	La courbe ROC (Receiver Operating Characteristic)	22
2.5.6	Entropie de Shannon :	23
2.6	Conclusion	23
3	État de l'art	25
3.1	Introduction	25
3.2	Interpretability and spatial efficacy of a deep learning based on site early warning framework using explainable artificial intelligence and geographically weighted random forests [34]	25
3.3	2. The role of artificial intelligence and IoT in prediction of earthquakes[35]	26
3.4	3. Forecasting future earthquakes with deep neural networks: application to California[36]	27
3.5	Small Earthquakes Can Help Predict Large Earthquakes[37]	27
3.6	Application of Artificial Intelligence in Predicting Earthquakes: State-of-the- Art and Future Challenges [38	28
3.7	Does the Catalog of California Earthquakes, With Aftershocks Included, Contain Information about Future Large Earthquakes?[39]	29
3.8	Étude critique	30
4	Solution de prédiction sismique basée sur l'IA	31
4.1	Problématique	31
4.2	Données utilisées	32
4.3	Méthodologie	33
4.3.1	Environnement de développement	33

CONTENTS

4.3.2	Bibliothèques et outils utilisés	33
4.3.3	Prétraitement des données	33
4.3.4	Équilibrage des classes(augmentation des données minoritaires	34
4.3.5	Séparation des données	34
4.4	Modèle d'apprentissage utilisé	34
4.4.1	Choix de l'algorithme LightGBM	34
4.4.2	Comparaison succincte LightGBM - Random Forest	35
4.5	Évaluation et résultats	35
4.6	Visualisations réalisées	35
4.7	Interprétation des résultats	38
	Conclusion Générale	39
	Références	39
	Bibliography	41
	Résumé	45

CONTENTS

List of Figures

1.1	Épicentre, hypocentre (foyer) et faille [2].	4
1.2	les différentes ondes du séisme [40].	7
1.3	Exemple de l'enregistrement du mouvement du sol en fonction du temps (= un sismogramme) d'un séisme au Cachemire (Pakistan) de magnitude 6 enregistré par une station sismologique située à plus de 1000 km de distance [41]	8
1.4	capteur sismique	8
2.1	la faille de San Andreas a été cartographiée par un drone [42].	15
2.2	(a) Drone DJI Inspire 1 v2.0 équipé d'une caméra de 12 mégapixels (capteur CMOS Sony Exmor 1/2.3) ; (b) UgCS pour un plan de vol tridimensionnel (3D) utilisant Google Earth ; (c) Pix4D pour la planification d'un relevé aérien ; (d) affleurement virtuel créé avec Agisoft Metashape [43].	15
2.3	robot sous marin [44].	17
2.4	Architecture de LSTM [45].	18
2.5	Schéma d'un auto-encodeur profond[18].	19
2.6	Architecture de CNN[19].	20
2.7	Architecture de GRU [46].	22
4.1	Système d'aide à la prédiction et la surveillance sismique	32
4.2	Extrait du dataset	33
4.3	Courbe ROC	36
4.4	Courbe précision-rappel	36
4.5	Comparaison séries temporelles	37
4.6	Erreurs de classification par mois	37
4.7	Matrice de confusion	38

LIST OF FIGURES

List of Tables

1.1	Séismes enregistrés en Algérie en 2024	9
4.1	Comparaison entre Random Forest et LightGBM	35

LIST OF TABLES

INTRODUCTION GENERALE

Les tremblements de terre représentent l'un des phénomènes naturels les plus destructeurs, tant par leurs conséquences humaines que matérielles. Malgré les progrès réalisés dans la compréhension des mécanismes géophysiques à l'origine des séismes, la prédiction précise de ces événements demeure un défi scientifique majeur. La nature imprévisible des mouvements tectoniques, combinée à la complexité des données sismologiques, rend difficile l'élaboration de modèles fiables capables d'anticiper ces catastrophes.

Dans ce contexte, l'intelligence artificielle (IA), et plus particulièrement les techniques d'apprentissage automatique, suscite un intérêt croissant. Grâce à leur capacité à traiter de grandes quantités de données et à détecter des motifs complexes, ces approches offrent de nouvelles perspectives pour l'analyse, la détection et la prédiction des séismes. En s'appuyant sur des données historiques et en apprenant à identifier les signaux faibles annonciateurs d'un événement sismique, les modèles d'IA peuvent potentiellement contribuer à l'amélioration des systèmes d'alerte précoce.

Ce mémoire s'inscrit dans cette dynamique. Il vise à explorer l'apport des méthodes d'apprentissage supervisé, et plus particulièrement de l'algorithme LightGBM, dans la classification et la prédiction des tremblements de terre. Après avoir présenté les bases des phénomènes sismiques et les outils de mesure associés, nous aborderons les concepts fondamentaux de l'intelligence artificielle appliquée à ce domaine. Un état de l'art permettra ensuite de situer notre travail par rapport aux recherches existantes. Enfin, une partie expérimentale décrira la mise en œuvre concrète du modèle, du prétraitement des données à l'évaluation des performances, en passant par la gestion du déséquilibre des classes.

Ce travail a pour objectif de montrer comment l'intelligence artificielle peut être mise au service de la réduction des risques sismiques, en proposant une approche pratique, reproductible et évolutive, susceptible de s'intégrer à terme dans des systèmes de surveillance opérationnels.

LIST OF TABLES

Chapter 1

Généralités sur les tremblements de terre

1.1 Introduction :

Les tremblements de terre figurent parmi les catastrophes naturelles les plus destructrices, entraînant des pertes humaines considérables, des dégâts matériels importants et des répercussions économiques majeures. Bien que les avancées scientifiques aient permis des progrès significatifs, la prévision précise des séismes demeure un défi. Les méthodes traditionnelles, basées sur l'analyse géologique et géophysique, permettent d'évaluer les risques sismiques à long terme dans certaines régions, mais elles ne peuvent pas prédire avec exactitude le moment, le lieu ou la magnitude des secousses. L'essor des technologies d'intelligence artificielle (IA) ouvre de nouvelles perspectives pour analyser les vastes ensembles de données sismiques complexes. Grâce au machine Learning et au deep learning, l'IA pourrait identifier des motifs et des signaux précurseurs imperceptibles à l'œil humain, améliorant ainsi les capacités de détection et d'anticipation des séismes. Ce travail vise à explorer l'application de l'IA dans la prévision des tremblements de terre. En analysant des données historiques et en utilisant des modèles d'apprentissage automatique, l'objectif est d'évaluer comment ces technologies peuvent optimiser la détection précoce des séismes et renforcer les systèmes d'alerte.

1.2 Qu'est-ce que un tremblement de terre ?

Un séisme ou tremblement de terre se traduit en surface par des vibrations du sol. Il provient de la fracturation des roches en profondeur; celle-ci est due à l'accumulation d'une grande énergie qui se libère, créant des failles, au moment où le seuil de rupture mécanique des roches est atteint. Les dégâts observés en surface sont fonction de l'amplitude, la fréquence et la durée des vibrations[1].

1.3 Caractéristiques principales :

Le point d'origine d'un séisme est appelé hypocentre ou foyer sismique. Il peut se trouver entre la surface et jusqu'à sept cents kilomètres de profondeur (limite du manteau

1. GÉNÉRALITÉS SUR LES TREMBLEMENTS DE TERRE

supérieur) pour les événements les plus profonds. On parle de l'épicentre du séisme pour désigner le point de la surface de la terre qui se trouve directement au-dessus de l'hypocentre [2].

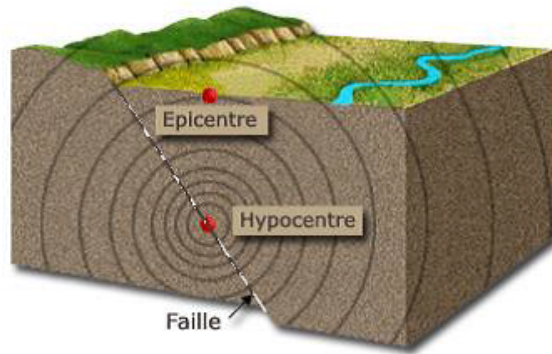


Figure 1.1: Épicentre, hypocentre (foyer) et faille [2].

1.3.1 Détermination de l'épicentre

Les stations sismiques enregistrent le temps d'arrivée des ondes P et S. En comparant ces données entre plusieurs stations, on peut localiser l'épicentre en traçant des cercles dont l'intersection indique sa position [3].

1.3.2 Magnitude

La puissance d'un tremblement de terre peut être quantifiée par sa magnitude, notion introduite en 1935 par le sismologue Charles Francis Richter. La magnitude se calcule à partir des différents types d'ondes sismiques en tenant compte de paramètres comme la distance à l'épicentre, la profondeur de l'hypocentre, la fréquence du signal, le type de sismographe utilisé, etc. La magnitude est une fonction continue logarithmique: lorsque l'amplitude des ondes sismiques est multipliée par 10, la magnitude augmente d'une unité. Ainsi, un séisme de magnitude 7 provoquera une amplitude dix fois plus importante qu'un événement de magnitude 6, cent fois plus importante qu'un de magnitude 5 [3].

1.3.3 Calcul de la magnitude :

Elle mesure l'énergie libérée par un séisme à son point d'origine elle est évaluée via :

Magnitude de moment (M_w) :

Est une des échelles logarithmiques qui mesurent la magnitude d'un séisme, c'est-à-dire la taille d'un séisme proportionnelle à l'énergie sismique dégagée. Basé sur les basses fréquences des ondes sismiques, elle quantifie précisément l'énergie émise par le séisme.

Magnitude de Richter (M_L) :

Basée sur l'amplitude maximale des ondes sismiques enregistrées par un sismographe local. Elle est efficace pour les séismes de faible à moyenne magnitude et proches de la station.

1.3 Caractéristiques principales :

Magnitude de surface (Ms) :

Basée sur les amplitudes des ondes de surface (ondes Rayleigh). Elle était souvent utilisée avant la généralisation de la magnitude de moment.

Magnitude de volume ou de corps (Mb) :

Calculée à partir des ondes P (ondes de volume), utile pour les séismes profonds.

Amplitude des ondes(A) :

correspond à la hauteur maximale atteinte par l'onde par rapport à sa position au repos.

1.3.4 Intensités d'un séisme

Caractérisation de la force destructrice d'un séisme par une cotation (de I à XII) des effets d'une secousse sismique en un site donné sur la population, les bâtiments et l'environnement. L'intensité traduit les effets et dommages induits par le séisme en un lieu donné. Son échelle est fermée et varie de I (non ressenti) à XII (presque tous les bâtiments détruits). A ne pas confondre avec la magnitude qui traduit l'énergie libérée par les ondes sismiques, qui est mesurée sur une échelle ouverte et dont les plus forts séismes sont de l'ordre de magnitude 9.

Intensité : Effets

I : Non ressenti

II : A peine perceptible

III : Ressenti par certains

IV : Ressenti par la plupart, objets vibrent

V : Frayeur, chutes d'objets

VI : Dégâts légers (fissuration plâtres)

VII : Dégâts (chutes cheminées, fissures murs)

VIII : Dégâts importants (effondrements murs)

IX : Destructures

X et + : Catastrophe

1.3.5 Mesure d'un tremblement de terre selon RICHTER

L'échelle de Richter a été instaurée en 1935. Elle nous fournit ce qu'on appelle la magnitude d'un séisme, calculée à partir de la quantité d'énergie dégagée au foyer. Elle se mesure sur une échelle logarithmique ouverte ; à ce jour, le plus fort séisme a atteint 9,5 sur l'échelle de Richter (Chili). Cette fois, il s'agit d'une valeur qu'on peut qualifier d'objective : il n'y a qu'une seule valeur pour un séisme donné [2].

1.3.6 Propagation des ondes sismiques :

La propagation de la rupture s'arrête lorsque l'énergie potentielle de déformation est totalement dissipée. Dans une région frappée par un séisme majeur, des secousses plus petites, appelées répliques, surviennent pendant un certain temps. Ces répliques sont souvent responsables de la destruction finale des structures déjà endommagées par la secousse principale. Il existe deux types d'ondes sismiques : les ondes de volume, qui se déplacent à l'intérieur de la Terre, et les ondes de surface. [2].

1. GÉNÉRALITÉS SUR LES TREMBLEMENTS DE TERRE

Ondes de volume :

Elles prennent naissance dans le foyer et se propagent à l'intérieure de la terre sous deux formes :

- *Ondes primaires P :*

Ou ondes longitudinales, ces types d'ondes font basculer les ouvrages d'avant en arrière dans la même direction que leur propagation, elles induisent donc, un mouvement vertical dans les ouvrages ce mouvement sera d'autant plus important que l'ouvrage se trouve à proximité de l'épicentre ; elles se propagent dans la croûte terrestre avec une vitesse de 6 à 8km/s et s'accompagnent d'un changement de volume (compression et dilatation alternées [2]).

- *Ondes secondaires S:*

Ou ondes de cisaillement, elles se propagent dans la croûte terrestre, avec une vitesse d'environ 3 à 5km/s et s'accompagnent d'une distorsion dans le plan perpendiculaire à la direction de propagation, provoquant un cisaillement sans changement de volume. C'est bien ce type d'ondes qui produit un mouvement horizontal dans les ouvrages. Les ondes P se propagent toujours à plus grande vitesse que les ondes S ; ainsi, quand un tremblement de terre se produit, les ondes P sont les premières à se manifester et à être enregistrées [2].

Ondes de surface :

Les ondes de volume qui arrivent à la surface de la terre donnant naissance à des ondes de surface ne concernant le sol que sur une profondeur extrêmement faible ; les ondes de surface les plus importantes sont les suivantes :

Ondes de Rayleigh:

ce sont des ondes pour lesquelles les points du sol décrivent des ellipses dans le plan vertical et dans le sens rétrograde. Ce mouvement entraîne des compressions ou des tractions, ainsi que des cisaillements dans le sol.

Ondes de Love:

Ce sont des ondes pour lesquelles les points du sol se déplacent dans un plan tangent à la surface et perpendiculairement à la propagation. Leur mouvement n'engendre que des contraintes de cisaillement [2].

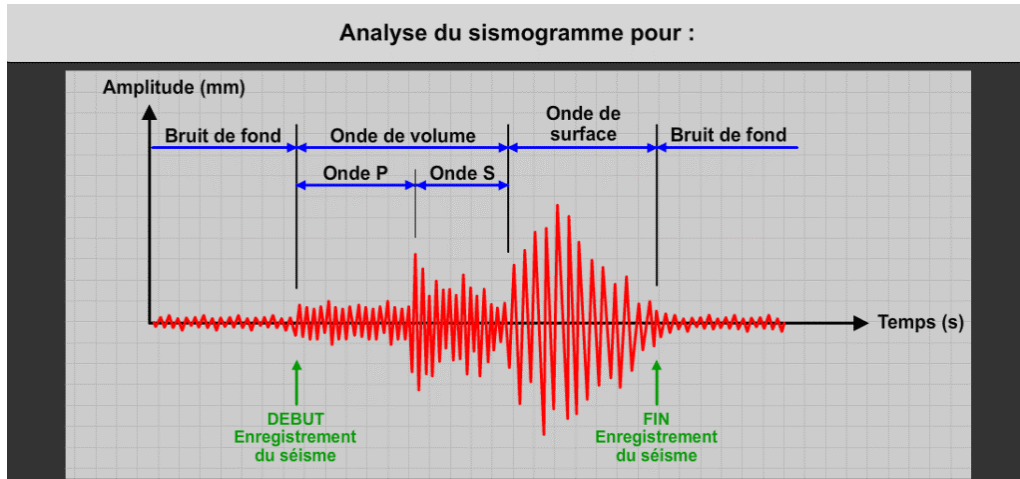


Figure 1.2: les différentes ondes du séisme [40].

1.4 Appareils de mesure :

1.4.1 Le Sismographe :

Un sismographe est un appareil qui enregistre et mesure les tremblements de terre. Au cours d'un séisme, les vibrations provoquées par la cassure de la croûte terrestre se transmettent à partir du point de rupture. Ces appareils captent et enregistrent ces vibrations, qui sont ensuite étudiées. L'enregistrement visuel produit par les sismographes s'appelle un sismogramme [4]

1.4.2 Le sismomètre :

Est un capteur qui enregistre le mouvement du support sur lequel il se trouve fixé. Ce terme est le plus souvent employé en sismologie et désigne l'instrument qui enregistre les mouvements du sol.

1. GÉNÉRALITÉS SUR LES TREMBLEMENTS DE TERRE

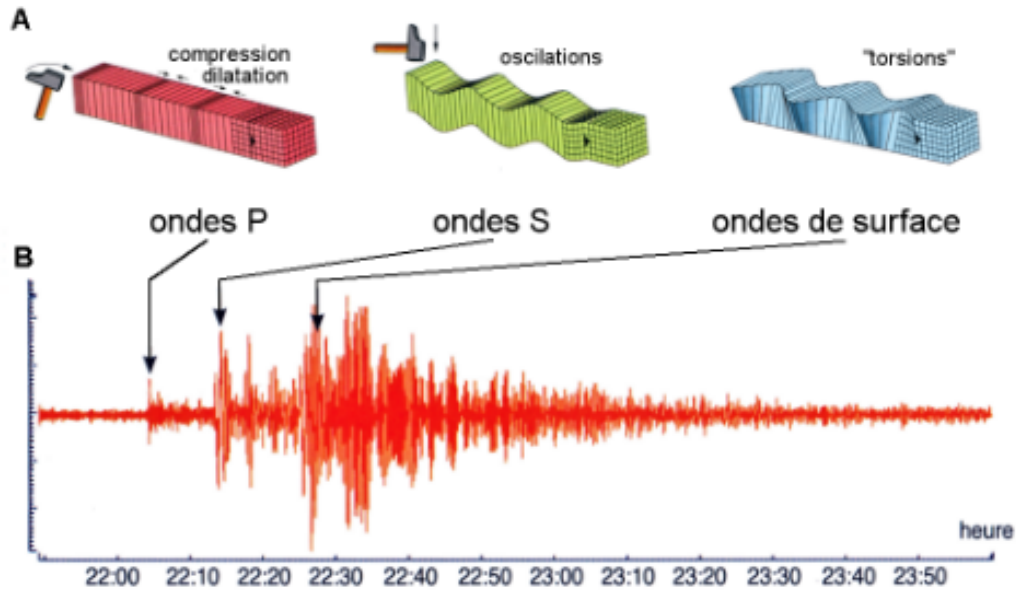


Figure 1.3: Exemple de l'enregistrement du mouvement du sol en fonction du temps (= un sismogramme) d'un séisme au Cachemire (Pakistan) de magnitude 6 enregistré par une station sismologique située à plus de 1000 km de distance [41]

1.4.3 Capteur sismique :

Les capteurs sismiques, initialement conçus dans les années 1860 pour détecter les tremblements de terre, fonctionnent comme des capteurs de vitesse ou des accéléromètres en mesurant les vibrations du sol. Aujourd'hui, ils sont utilisés pour l'exploration des ressources et la surveillance de la sécurité. Dans le domaine de la sécurité, ces capteurs analysent en temps réel les vibrations détectées afin d'alerter en cas d'intrusion. Selon leur sophistication, ils peuvent distinguer l'origine des vibrations (personnes, véhicules) et localiser leur source.



Figure 1.4: capteur sismique

1.4.4 Réseaux de capteurs GNSS (GPS) :

Les réseaux de capteurs GNSS (GPS) sont utilisés pour mesurer les déplacements lents des plaques tectoniques avant, pendant et après un séisme, avec une précision pouvant atteindre quelques millimètres (mm) par an, et pouvant aller jusqu'à plusieurs centimètres (cm) ou même mètres (m) lors d'un séisme.

1.4.5 Historique des plus grands séismes

Voici un historique des plus grands séismes enregistrés depuis 1868.

Table 1.1: Séismes enregistrés en Algérie en 2024

Date	Heure	Mag	Localisation
16/12/2024	01:09	4.9	03 Km Nord de Zeboudja W.Chlef
10/12/2024	05:02	4.0	07 Km Sud Est de Bourkika W.Tipaza
10/10/2024	14:59	4.3	09 Km Nord Est de Azeffoun W.Tizi Ouzou
03/08/2024	09:51	4.3	04 Km Nord Est de Derdrag W.Medea
24/07/2024	20:16	3.4	03 Km Ouest de Hadjedj W.Chlef
25/06/2024	04:53	4.0	06 Km Sud Est de Ziamamansouriah W.Jijel
20/04/2024	13:44	4.2	06 Km Sud Ouest de M'cif W.M'sila
09/03/2024	14:54	4.3	13 Km Sud Ouest de Ain Errich W.M'sila

1.5 Méthode de Surveillance

1.5.1 Ancienne méthode chinoise

L'ancienne méthode chinoise consistait en un vase de bronze comportant huit dragons sur le contour, Une bille était placée dans la gueule de chaque dragon, prête à tomber dans la gueule d'un crapaud. Lorsqu'un séisme se produisait, la bille d'un des dragons (dépendant de l'endroit où se produisait le séisme) tombait dans la gueule d'un des crapauds. Cela indiquait la direction de l'épicentre du tremblement de terre, et vers où il fallait envoyer les secours [3].

1.5.2 Méthode moderne

La localisation de l'épicentre par des moyens modernes se fait à l'aide de plusieurs stations sismiques (locale, régionale, globale...), et un calcul tridimensionnel. Les capteurs modernes permettent de détecter des événements très sensibles, tels qu'une explosion nucléaire. Le Centre sismologique euro-méditerranéen a quant à lui développé un processus

1. GÉNÉRALITÉS SUR LES TREMBLEMENTS DE TERRE

de détection sismique basé sur l'analyse du trafic web et des contenus sur Twitter. La collecte de témoignages et de photos permet en outre de connaître l'intensité des séismes ressentis, et d'apprécier et géolocaliser les dégâts matériels [3].

1.5.3 Méthode de prédiction

Les méthodes de prédiction reposent sur une prévision qui spécifie, avec leur incertitude, la position, la taille, la date du séisme, et donne une estimation de la probabilité de son propre succès. La possibilité de la prédiction sismique repose sur l'existence, et la reconnaissance des « précurseurs », signes avant-coureurs d'un séisme. En l'absence de précurseurs fiables, ces méthodes sont accompagnées de non-détections qui entraînent des procès pour les spécialistes et des fausses alarmes qui provoquent une perte de confiance des populations alertées. On peut distinguer trois types de prévisions : la prévision à long terme (sur plusieurs années), à moyen terme (sur plusieurs mois) et à court terme (inférieur à quelques jours).

1.5.4 Prédiction à long terme (sur plusieurs années)

Les méthodes de prévision à long terme des séismes ne permettent pas de prédire précisément leur moment ou leur magnitude, mais elles évaluent le risque sismique. L'une de ces méthodes repose sur l'étude de la récurrence des séismes et leur périodicité, liée à la tectonique des plaques, qui se déplacent lentement et génèrent des contraintes sur la croûte terrestre, provoquant des ruptures. La méthode des lacunes sismiques identifie les zones de faille où les séismes sont absents depuis longtemps, indiquant un risque accru.

1.5.5 Prédiction à moyen terme (sur plusieurs mois)

Les méthodes de prévision à moyen terme des séismes reposent sur des signes précurseurs. Dans les années 1960, des chercheurs soviétiques ont noté que la vitesse des ondes P diminue avant un séisme, pour ensuite revenir à la normale juste avant la secousse, un phénomène lié aux fractures dans les roches (méthode de la dilatance). Bien que prometteur, ce phénomène n'est pas systématique. La concentration en radon dans les eaux souterraines augmente également avant un séisme, mais d'autres facteurs naturels compliquent l'interprétation. Des scientifiques japonais ont aussi observé des soulèvements du sol avant un séisme, bien que ce phénomène soit aléatoire. Enfin, de petits séismes précèdent souvent les séismes plus violents, mais ces précurseurs ne sont pas fiables car les phases d'activité sismique peuvent être entrecoupées de périodes calmes.

1.5.6 Prédiction à court terme (inférieur à quelques jours)

Plusieurs méthodes ont été explorées pour prédire les séismes, bien que leur efficacité reste débattue. La méthode VAN (Val

1.5.7 Prédiction à moyen terme (sur plusieurs mois)

Les méthodes de prévision à moyen terme des séismes reposent sur des signes précurseurs. Dans les années 1960, des chercheurs soviétiques ont noté que la vitesse des ondes P diminue avant un séisme, pour ensuite revenir à la normale juste avant la secousse, un phénomène lié aux fractures dans les roches (méthode de la dilatance). Bien que prometteur, ce phénomène n'est pas systématique. La concentration en radon dans les eaux souterraines augmente également avant un séisme, mais d'autres facteurs naturels compliquent l'interprétation. Des scientifiques japonais ont aussi observé des soulèvements du sol avant un séisme, bien que ce phénomène soit aléatoire. Enfin, de petits séismes

précèdent souvent les séismes plus violents, mais ces précurseurs ne sont pas fiables car les phases d'activité sismique peuvent être entrecoupées de périodes calmes.

1.5.8 Prédiction à court terme (inférieur à quelques jours)

Plusieurs méthodes ont été explorées pour prédire les séismes, bien que leur efficacité reste débattue. La méthode VAN (Valeur Actuelle Net), développée par des scientifiques grecs dans les années 1980, consiste à enregistrer les anomalies dans les courants électriques naturels de surface (Valeur Actuelle Net), développée par des scientifiques grecs dans les années 1980, consiste à enregistrer les anomalies dans les courants électriques naturels du sous-sol, appelées SES (Signaux Électro-Sismiques), qui se produisent avant un séisme. Cependant, cette méthode est critiquée en raison de l'incertitude sur l'origine et la propagation de ces signaux. Une autre approche, la "méthode chinoise", impliquait la population pour surveiller divers signes précurseurs, tels que les comportements des animaux et les variations dans les puits. Bien qu'elle ait permis de prédire un séisme en 1975, elle a échoué en 1976 avec une prédiction ratée. Enfin, des satellites comme Quakesat et Demeter ont été lancés pour analyser les variations du champ magnétique terrestre, un phénomène supposé se produire à l'approche d'un séisme, mais les résultats sont encore en qtest.

1.6 Conclusion

L'étude des mécanismes sismiques met en évidence la complexité des phénomènes qui sous-tendent la survenue des tremblements de terre. Malgré les avancées importantes en sismologie, la prévision exacte des séismes reste aujourd'hui un défi majeur. Toutefois, la compréhension approfondie des processus tectoniques, des types d'ondes et des paramètres influençant l'occurrence des séismes constitue une base indispensable pour envisager de nouvelles approches prédictives. C'est sur cette base que s'appuieront les travaux présentés dans les prochains chapitres, où seront explorées les contributions récentes de l'intelligence artificielle à la prédiction sismique.

1. GÉNÉRALITÉS SUR LES TREMBLEMENTS DE TERRE

Chapter 2

IA appliquée à la surveillance et à la détection des séismes

2.1 L'utilisation de L'IA dans la prédiction de tremblement de terre

Le Big Data peut collecter bien plus facilement des informations provenant de différents événements pour déduire la probabilité d'un tremblement de terre. La nature donne toujours des signaux, ou il faut les comprendre afin de pouvoir agir en conséquence. Puisque ces indices sont éparpillés, la meilleure solution est d'utiliser le Big Data afin de les collecter et de les regrouper dans un site central pour obtenir d'eux des résultats plus précis.

2.2 Effets et prévention :

- Déplacements du sol : Le déplacement du sol après un séisme est mesuré à l'aide de GPS et d'observation satellites.
- Analyse des dégâts : permet d'améliorer les techniques de construction pour mieux résister aux séismes futurs.
- Surveillance des répliques : Aide à prévenir les risques secondaires après un séisme principale.

2.3 Intégration aux systèmes d'alerte :

Les données sismiques alimentent des systèmes comme ShakeAlert(shakealert.org) pour détecter et prévenir les séismes futurs, améliorer la modélisation des effets et renforcer les infrastructures dans les zones à risque, si la magnitude dépasse un certain seuil critique, ou peu profond, une alerte est générée. Exemple : Séisme de Ridgecrest (Californie, 2019).

Le 4 juillet 2019, un premier séisme de magnitude 6.4 frappe Ridgecrest. Les capteurs ShakeAlert détectent les ondes P et envoient une alerte.

2. IA APPLIQUÉE À LA SURVEILLANCE ET À LA DÉTECTION DES SÉISMES

Le 5 juillet, un séisme encore plus puissant (7.1) se produit. ShakeAlert envoie des alertes jusqu'à Los Angeles et San Francisco, donnant plusieurs secondes aux habitants pour se protéger.

Les hôpitaux, trains et centrales électriques activent des protocoles d'urgence (ex : ralentissement des trains, coupure automatique de gaz). L'efficacité du système ShakeAlert lors de ce séisme a démontré l'importance des capteurs sismiques et des protocoles d'alerte pour la protection des populations et infrastructures .

2.3.1 LIDAR :

Un acronyme anglais de « détection et télémétrie par la lumière », est une technologie de télédétection qui utilise des faisceaux laser pour mesurer des distances et des mouvements précis en temps réel.

2.3.2 Drones :

Les drones jouent un rôle croissant dans la surveillance sismique et la gestion des catastrophes naturelles. Ils sont utilisés pour cartographier les failles sismiques, évaluer les dommages post-séisme et surveiller les volcans. Par exemple, En Californie, des scientifiques ont équipé un drone d'une caméra RGB et d'un système LIDAR pour cartographier avec précision la faille de San Andreas (figure 5), une fracture de 1300 km entre les plaques Pacifique et Nord-américaine. Cette méthode, combinant photogrammétrie et LIDAR, permet de créer des modèles 3D ultra-détaillés du sol essentiels pour analyser les microdéformations de la faille et anticiper les risques sismiques. Le drone utilisé, basé sur le DJI Matrice M600 Pro, est équipé de technologies Velodyne et Geodetics, offrant une solution efficace et mobile pour l'étude des zones sismiquement actives. Grâce à cette technologie, les chercheurs peuvent détecter les moindres déplacements de terrain qui pourraient indiquer une activité tectonique anormale. En complément, d'autres drones et capteurs GNSS (Global Navigation Satellite System) sont également utilisés pour la surveillance des failles en temps réel, améliorant ainsi la précision des simulations sismiques et la prévention des catastrophes naturelles .

Le DJI Matrice M600 : est un drone professionnel utilisé pour transporter des capteurs comme LIDAR ou GNSS dans des missions de cartographie ou de surveillance de terrain

Velodyne et Geodetics : Deux entreprises l'une fournit des capteurs LIDAR pour la cartographie 3D tandis que l'autre pour le positionnement précis .

2.3 Intégration aux systèmes d'alerte :



Figure 2.1: la faille de San Andreas a été cartographiée par un drone [42].

Dans une autre étude menée en Italie, un levé aérien a été réalisé à l'aide d'un drone (UAV - Unmanned Aerial Vehicle) DJI Inspire 1 v2.0, équipé d'une caméra de 12 méga pixels (capteur CMOS Sony Exmor 1/2.3). Ce drone est capable de capturer des images nettes et lumineuses, même dans des conditions difficiles. Sa structure géométrique offre une excellente résistance au vent, un atout essentiel pour les opérations dans les zones montagneuses. Les vols ont été planifiés et contrôlés à l'aide de deux logiciels (i) UgCS : Ce logiciel permet de maintenir le drone à une altitude constante au-dessus du sol en utilisant un modèle numérique de terrain (DEM) ou Google Earth. (ii) Pix4D : Principalement utilisé pour les levés verticaux, il permet de prédéfinir les positions de prise de vue horizontales et verticales.

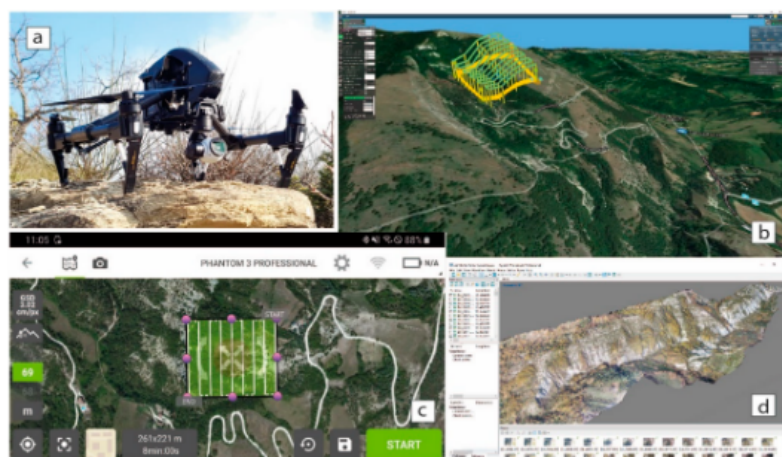


Figure 2.2: (a) Drone DJI Inspire 1 v2.0 équipé d'une caméra de 12 mégapixels (capteur CMOS Sony Exmor 1/2.3) ; (b) UgCS pour un plan de vol tridimensionnel (3D) utilisant Google Earth ; (c) Pix4D pour la planification d'un relevé aérien ; (d) affleurement virtuel créé avec Agisoft Metashape [43].

Les photographies capturées, représentant environ 20 gigaoctets de données, ont ensuite

2. IA APPLIQUÉE À LA SURVEILLANCE ET À LA DÉTECTION DES SÉISMES

été traitées avec le logiciel Agisoft Metashape. Ce logiciel permet de créer des modèles 3D d'affleurements virtuels à partir d'une séquence de photos prises avec un chevauchement de 80 pourcent . Le processus comprend plusieurs étapes :

- Importation des images.
- Alignement des photos (en identifiant les points communs au sol).
- Création d'un nuage de points dense.
- Génération d'un maillage 3D.
- Ajout de texture au maillage pour obtenir un modèle réaliste.

Le résultat final est un modèle numérique de terrain (DEM) recouvert de photos géoréférencées, permettant de créer un affleurement virtuel en 3D. Cette méthode offre une représentation précise et détaillée des structures géologiques, facilitant ainsi l'analyse des failles et des déformations du sol. mais, Jusqu'à présent, aucune alerte précoce n'a été déclenchée uniquement sur la base des observations faites par drones. Toutefois, ces appareils jouent un rôle essentiel dans la cartographie des failles et l'évaluation des risques sismiques à long terme.

2.3.3 Robots sous marins :

Un projet innovant à Ayia Napa, Chypre, prévoit l'utilisation d'un essaim de sous-marins autonomes pour surveiller, protéger et analyser l'évolution de récifs artificiels offshore (figure 3). Ces récifs, conçus pour stimuler la biodiversité marine, seront installés dans des zones maritimes auparavant désertes, ont annoncé les responsables lundi. Ces véhicules sous-marins autonomes (AUVs), présentés comme "l'avenir de la surveillance sous-marine", seront équipés de capteurs et de caméras haute définition. Ils fourniront aux chercheurs des données précises sur l'efficacité des récifs artificiels, conçus et fabriqués grâce à l'impression 3D. Fabriqués à partir d'un ciment écologique, ces récifs intégreront une station d'accueil permettant aux AUVs de se recharger et de transmettre leurs relevés environnementaux et vidéo. Le principal atout de ces AUVs réside dans leur capacité à patrouiller sous l'eau pendant un mois, garantissant une surveillance continue. En plus de collecter des données, ils assureront la protection des récifs artificiels en détectant toute activité suspecte, telle que la pêche illégale ou la présence de navires dans des zones protégées. En cas d'anomalie, les scientifiques pourront rapidement alerter les autorités compétentes. En gros Ce projet montre que les robots sous-marins peuvent déjà générer des alertes, mais pour l'instant, elles concernent la surveillance environnementale. À l'avenir, avec des capteurs adaptés, ces robots pourraient aider à détecter des séismes sous-marins et alerter sur des tsunamis potentiels.



Figure 2.3: robot sous marin [44].

2.3.4 Robots terrestres

Les robots terrestres jouent un rôle crucial dans l'analyse et la surveillance sismique. Ils sont équipés de capteurs avancés pour détecter les vibrations du sol, analyser les failles actives et transmettre des données en temps réel aux chercheurs. Parmi ces robots, Quakebot est un exemple marquant d'outil automatisé dans la détection des séismes.

1. **Exemple :** Séisme de Ridgecrest (Californie, 2019) :

Quakebot est un robot logiciel développé par le Los Angeles Times pour générer automatiquement des articles dès qu'un séisme est détecté en Californie. Il fonctionne en surveillant en temps réel les données sismiques fournies par l'United States Geological Survey (USGS), l'agence officielle de surveillance des séismes aux États-Unis. Lorsqu'un tremblement de terre survient, Quakebot récupère immédiatement des informations clés telles que la magnitude, la localisation, la profondeur et l'heure de l'événement, puis génère un article automatisé qu'il transmet aux journalistes pour une validation rapide avant publication. Un exemple marquant de son efficacité est le séisme de Ridgecrest survenu les 4 et 5 juillet 2019, avec des secousses de magnitude 6.4 puis 7.1. Lors du premier séisme, Quakebot a rédigé et publié un article en quelques secondes. Le lendemain, lors d'un second séisme plus puissant, le système d'alerte officiel ShakeAlert a diffusé des notifications jusqu'à 48 secondes avant les secousses à Los Angeles et San Francisco, permettant ainsi aux citoyens, aux hôpitaux, aux trains et aux infrastructures d'activer leurs protocoles de sécurité. Tandis que ShakeAlert s'occupait de la diffusion des alertes préventives, Quakebot a assuré une communication rapide et efficace de l'information dans les médias.,ShakeAlert

2.4 Méthodes de Deep Learning pour la Prédiction Sismique

2.4.1 Introduction

Le deep learning (ou apprentissage profond) constitue une approche prometteuse pour la prédiction sismique, en offrant des outils avancés pour l'analyse des données et l'amélioration des prévisions de séismes. Des architectures telles que les réseaux de neurones récurrents (RNN), notamment les modèles LSTM (Long Short-Term Memory), sont utilisées pour traiter les séquences temporelles d'activités sismiques. Les autoencodateurs permettent de détecter les anomalies dans les signaux, tandis que les réseaux de neurones convolutifs (CNN) sont adaptés à l'analyse spatiale et à la prédiction de la magnitude des séismes. De plus, les réseaux de neurones graphiques (GNN) modélisent les interactions complexes entre failles tectoniques, et les modèles hybrides combinant CNN et LSTM améliorent la précision en intégrant des dimensions spatiales et temporelles. Enfin, l'apprentissage par transfert facilite l'adaptation de modèles à des régions faiblement documentées. Ensemble, ces techniques ouvrent la voie à une meilleure fiabilité des systèmes de prévision, bien qu'une validation approfondie reste nécessaire pour leur intégration opérationnelle.

2.4.2 Prédiction des Répliques Sismiques avec des LSTM

C'est une méthode permettant d'entraîner les réseaux neuronaux et de stocker des informations importantes sur le long terme. L'utilisation des réseaux LSTM pour la prédiction des répliques sismiques s'avère particulièrement efficace, car ces modèles permettent de modéliser les séquences temporelles d'événements sismiques et d'améliorer la localisation des répliques, surpassant ainsi les méthodes traditionnelles basées sur les lois empiriques comme la loi d'Omori qui définit la décroissance du nombre de répliques qui se produisent après un grand séisme.

Les LSTM utilisent des équations récurrentes pour capturer les dépendances temporelles des données sismiques.

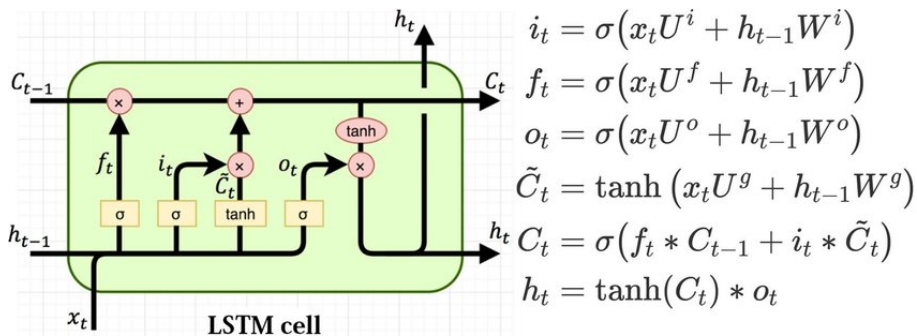


Figure 2.4: Architecture de LSTM [45].

Tel que:

C:memory, h: Hiddenstate, f<t>: Forget gate, i<t>: Input gate, c:Candidate memory, o:Output gate

2.4.3 Détection d'Anomalies Sismiques avec des Auto encodeurs

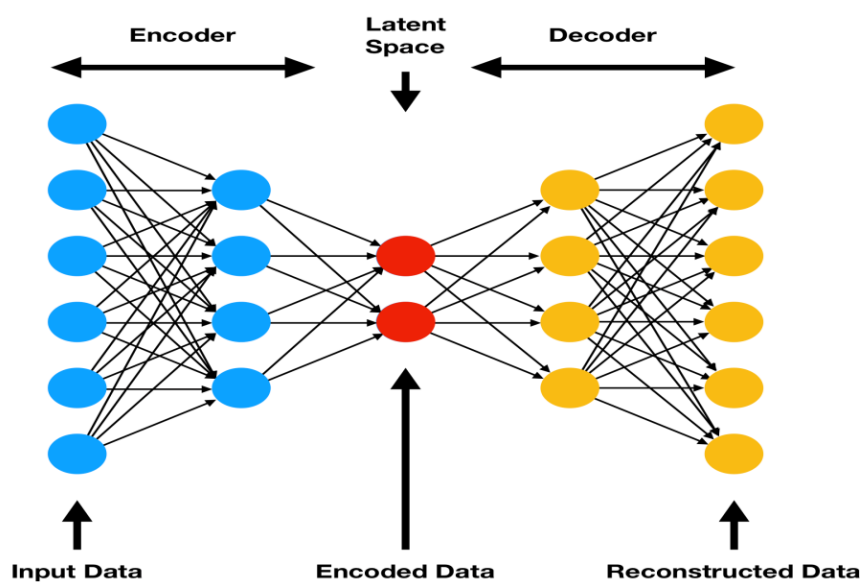


Figure 2.5: Schéma d'un auto-encodeur profond[18].

Les auto encodeurs sont des architectures de réseaux neuronaux conçues pour apprendre une représentation comprimée (encodée) des données d'entrée, puis reconstruire l'entrée d'origine à partir de cette représentation réduite. Dans le cadre de la détection d'anomalies sismiques, les auto encodeurs ont montré un taux de détection de 85 pourcent , ce qui permet d'identifier des signaux faibles qui pourraient être liés à des précurseurs de séismes .

2. IA APPLIQUÉE À LA SURVEILLANCE ET À LA DÉTECTION DES SÉISMES

2.4.4 Prédiction de la Magnitude des Séismes avec des CNN

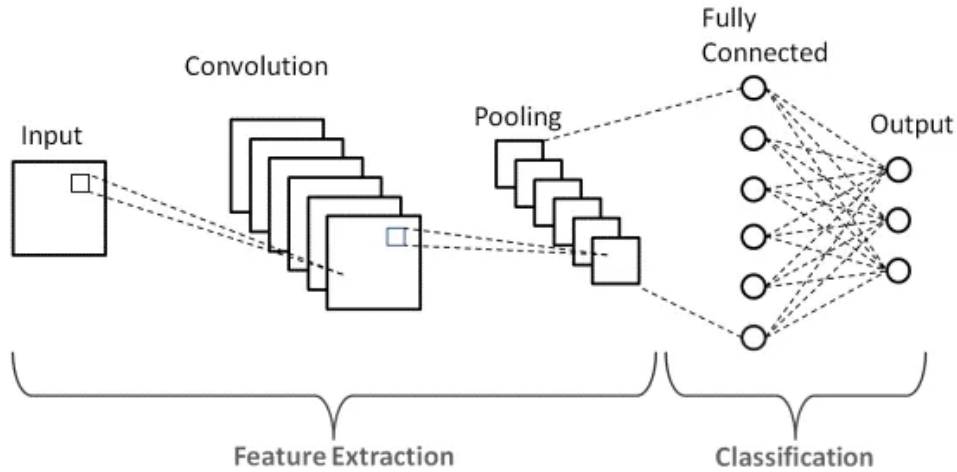


Figure 2.6: Architecture de CNN[19].

La prédiction de la magnitude des séismes avec des CNN (Convolutional Neural Network) a également montré une amélioration significative par rapport aux méthodes classiques, avec une erreur moyenne de ± 0.2 sur l'échelle de Richter. Les CNN extraient des caractéristiques spatiales en appliquant des convolutions :

$$Y=f(WX+b)$$

Où :

X est l'entrée (les données sismiques sous forme de tenseurs),

W est le filtre (ou noyau de convolution),

b est le biais,

f est une fonction d'activation (telle que ReLU ou Leaky ReLU), qui introduit la non-linéarité dans le réseau pour capturer des motifs complexes dans les données.

2.4.5 Modélisation des Interactions entre Failles avec des GNN

L'utilisation des GNN (Graph neuronal network) pour modéliser les interactions entre failles a permis une meilleure compréhension des relations spatiales complexes et une précision de 75 pourcent pour la prédiction des zones à haut risque sismique (Ross et al., 2021). Un GNN met à jour l'état d'un nœud en fonction de ses voisins :

$$h_v(l+1) = f\left(W \sum_{u \in N(v)} h_u(l) + b\right) \quad (2.1)$$

Où:

$h_v(l)$ représente l'état du nœud v à la couche l,

$N(v)$ est l'ensemble des voisins du nœud v,

W est une matrice de poids apprise par le modèle,

b est un biais,

2.4 Méthodes de Deep Learning pour la Prédiction Sismique

f est une fonction d'activation (telle que ReLU ou Sigmoid)

L'opération de sommation permet d'agréger les informations des nœuds voisins pour mettre à jour l'état du nœud central

Ce processus est répété sur plusieurs couches du GNN pour capturer des relations complexes dans le graphe des failles sismiques.

2.4.6 Prédiction des Séismes avec des Modèles Hybrides (CNN + LSTM)

Les modèles hybrides combinant des réseaux de neurones convolutifs (CNN) et des réseaux de neurones à mémoire longue courte (LSTM) se révèlent particulièrement efficaces pour la prédiction des séismes. Les CNN permettent d'extraire automatiquement des caractéristiques spatiales pertinentes à partir des signaux sismiques bruts, tandis que les LSTM sont capables de capturer les dépendances temporelles à long terme dans les données.

2.4.7 Apprentissage par Transfert pour la Prédiction Sismique

L'apprentissage par transfert (transfer learning) est une technique de plus en plus utilisée en prédiction sismique pour tirer parti de modèles pré-entraînés sur de grandes quantités de données sismiques, et les adapter à des régions ou types de séismes spécifiques où les données sont plus limitées. Cette méthode permet de réduire considérablement le temps d'entraînement et d'améliorer la précision, surtout dans des contextes où les événements sismiques sont rares ou peu documentés

2.4.8 Gated Recurrent Unit

Une GRU (Gated Recurrent Unit) est un type de réseau de neurones récurrent (RNN) conçu pour traiter des données séquentielles, comme des séries temporelles ou des séquences de texte. Elle améliore les RNN classiques en utilisant des mécanismes de portes qui contrôlent le flux d'information dans le réseau. Contrairement aux LSTM, les GRU ont une structure plus simple avec seulement deux portes principales : la porte de mise à jour (update gate) et la porte de réinitialisation (reset gate). Cela leur permet de mieux gérer la mémoire à long terme tout en étant plus rapides à entraîner.

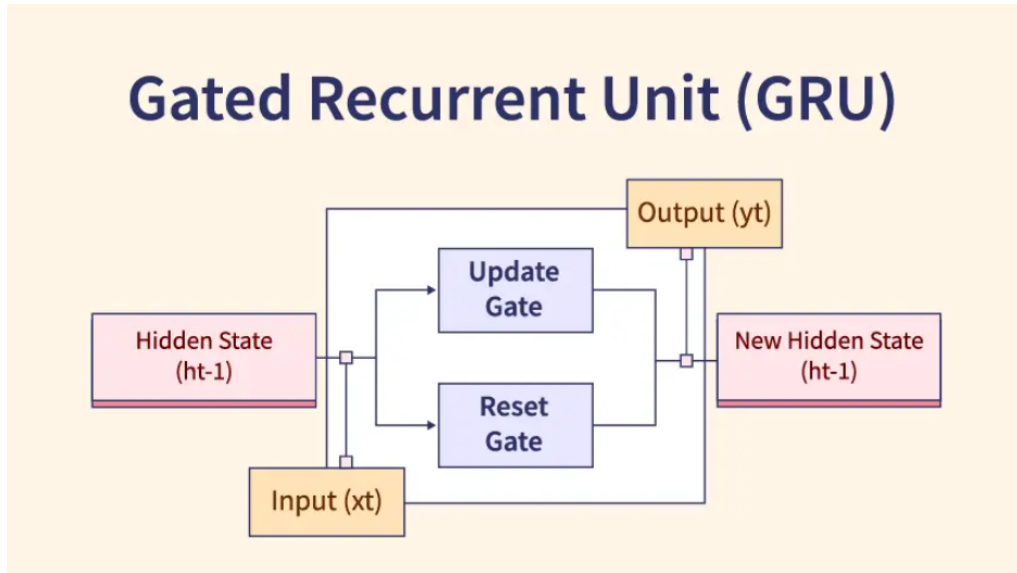


Figure 2.7: Architecture de GRU [46].

2.5 Mesures d'évaluation

2.5.1 Accuracy :

ou exactitude en français est une métrique de classification l'exactitude est la proportion de prediction correctes c'est a dire les vrais positifs et vrais négatifs parmi le nombre total de prédictions ce nombre est donc compris entre 0 parfaite inexactitude et 1 inclus parfaite exactitude [21]

2.5.2 Matrice de confusion :

Une matrice de confusion est un tableau qui permet de visualiser les performances d'un modèle de classification en comparant les valeurs réelles et les valeurs prédites par le modèle. Elle montre combien de prédictions étaient correctes et combien étaient incorrectes, organisées selon le type d'erreur. [22]

2.5.3 Recall :

c'est le pourcentage des vrais positifs détectés correctement parmi tous les vrais cas positifs existants et sa formule est la suivante [23] $\text{Recall} = \frac{\text{Vrais positifs (VP)}}{\text{Vrais positifs (VP)} + \text{Faux négatifs (FN)}}$

2.5.4 Le F1-score

est une mesure qui permet de synthétiser la précision et le rappel en une seule valeur, donnant un aperçu fiable de la performance du modèle, surtout lorsque les erreurs n'ont pas le même impact [24]. $F1 = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$

2.5.5 La courbe ROC (Receiver Operating Characteristic)

est une représentation graphique qui montre la performance d'un modèle de classification binaire pour différents seuils de décision. Elle trace le taux de vrais positifs (Recall) en

fonction du taux de faux positifs (FPR). Plus la courbe est proche du coin supérieur gauche, meilleur est le modèle. [25]

2.5.6 Entropie de Shannon :

une mesure de l'incertitude ou du désordre dans un système. Si l'entropie est élevée, cela veut dire qu'on reste dans le flou (beaucoup d'incertitude). Si elle est basse, cela veut dire que le modèle arrive à donner des prévisions plus précises et ordonnées. [26]

2.6 Conclusion

L'intelligence artificielle offre aujourd'hui des outils performants pour améliorer la prédiction et la surveillance sismique grâce à sa capacité d'analyse de données massives et complexes. Malgré certaines limites liées notamment à la disponibilité et à la qualité des données, les résultats obtenus encouragent à approfondir l'application de ces méthodes dans ce domaine.

Dans le chapitre suivant, nous examinerons plusieurs travaux de recherche récents qui ont mis en œuvre différentes approches d'IA pour la prédiction sismique. Cette revue permettra d'identifier les principales méthodes utilisées, les résultats obtenus, ainsi que les perspectives et limites relevées par les auteurs.

2. IA APPLIQUÉE À LA SURVEILLANCE ET À LA DÉTECTION DES SÉISMES

Chapter 3

État de l'art

3.1 Introduction

Les avancées récentes dans le domaine de la prédiction sismique témoignent d'une convergence croissante entre les géosciences, l'intelligence artificielle (IA) et l'ingénierie des capteurs. Plusieurs études analysées partagent une base méthodologique commun fondé sur l'exploitation de grands ensembles de données sismiques historiques et en temps réel, sur lesquels sont appliquées des techniques de machine learning ou d'apprentissage profond. Après avoir présenté les enjeux majeurs et les approches générales employées dans le domaine de la prédiction sismique par intelligence artificielle, nous allons maintenant détailler les contributions spécifiques et les résultats clés des articles analysés.

3.2 Interpretability and spatial efficacy of a deep learning based on site early warning framework using explainable artificial intelligence and geographically weighted random forests [34]

Les auteurs de cet article présentent ROSERS (Real-time On-Site Earthquake Response Spectra estimation), un cadre innovant d'alerte précoce aux tremblements de terre (EEW) fondé sur l'apprentissage profond, capable d'estimer les spectres de réponse sismique dès les premières secondes suivant l'arrivée des ondes sismiques. Ce système repose sur une architecture hybride combinant un variational autoencoder (VAE) et des réseaux de neurones profonds optimisés pour prédire avec précision les mouvements sismiques futurs à partir d'un signal initial limité. Afin de renforcer la transparence du modèle, les auteurs intègrent des méthodes d'intelligence artificielle explicable (XAI), en particulier les SHAP values (Shapley Additive Explanations), permettant d'interpréter les contributions des différentes variables latentes dans les prédictions du spectre de réponse. Parallèlement, une approche de modélisation spatiale avancée, basée sur les forêts aléatoires géographiquement pondérées (GWRF), est proposée pour évaluer l'efficacité géographique des stations d'EEW en tenant compte des variations régionales des paramètres sismiques. Les expérimentations s'appuient sur un vaste ensemble de données comprenant 11 000 enregistrements bidirectionnels (nord-sud et est-ouest) issus des réseaux sismologiques K-Net et Kik-Net au Japon,

3. ÉTAT DE L'ART

couvrant 1 500 séismes survenus entre 1996 et 2022, avec des magnitudes allant de M4 à M9 et des distances épacentrales atteignant 300 km.

3.3 2. The role of artificial intelligence and IoT in prediction of earthquakes[35]

Dans cet article les auteurs examinent le rôle croissant de l'intelligence artificielle (IA) et de l'Internet des objets (IoT) dans la prédiction des séismes. Face aux défis persistants de la prévision sismique — manque de données, complexité des signaux sismiques et diversité géologique — les auteurs passent en revue les méthodes conventionnelles et présentent des approches innovantes associant IA et IoT pour améliorer la précision et la réactivité des prédictions. Ils soulignent les limites actuelles des systèmes classiques de surveillance (sismomètres, télédétection) et discutent de l'apport potentiel des technologies modernes. L'objectif principal est d'évaluer comment l'IA et l'IoT peuvent être intégrés pour pallier les insuffisances des approches traditionnelles et fournir un cadre d'amélioration continue pour les systèmes de prévision. L'étude se concentre sur plusieurs disciplines complémentaires : géosciences (mécanismes des séismes), IA (réseaux de neurones récurrents et convolutifs, LSTM, transformers), IoT (capteurs connectés) et télédétection (GPS, radar satellitaire InSAR). Méthodologiquement, les auteurs proposent une architecture combinée où des modèles d'IA avancés analysent les données issues de capteurs IoT pour estimer la magnitude et la localisation probables des événements sismiques. Les données sont d'abord traitées en local via des edgenodes (unités de calcul proches des capteurs) afin de limiter la latence, notamment dans les systèmes d'alerte précoce. Elles sont ensuite centralisées dans des infrastructures cloud pour des analyses plus approfondies par des modèles complexes. Les sources incluent des capteurs terrestres, sous-marins et des données satellitaires, renforçant ainsi la fiabilité et la diversité des observations. Plusieurs expériences ont été réalisées pour évaluer cette approche. Des modèles IA comme BiKCONN, qui combine réseaux de neurones convolutifs (CNN) et réseaux à graphes (GNN), ont été appliqués à des catalogues sismiques du Japon et de l'Alaska. Les résultats démontrent une réduction significative des erreurs de prédiction grâce à la capacité de ces modèles à capturer des motifs complexes et non linéaires. En parallèle, des réseaux de capteurs sans fil et des dispositifs RFID ont été testés pour la détection en temps réel des mouvements du sol, avec des résultats encourageants en termes de réactivité et de précision. D'une part, BiKCONN (Bidirectional K-Connected Neural Network) est une architecture de réseau neuronal conçue pour l'analyse de séquences temporelles complexes comme les signaux sismiques. Il combine un réseau RNN bidirectionnel (traitant les données dans les deux sens temporels) et un système de K-connexion, où chaque couche du réseau peut être reliée à plusieurs couches précédentes. Cette structure favorise la rétropropagation efficace du signal et l'apprentissage de motifs profonds et complexes, particulièrement utiles dans la prévision de séismes. Les jeux de données utilisés proviennent de catalogues sismiques historiques (Japon, Alaska, région Himalayenne), de capteurs IoT en temps réel et de données satellitaires (GPS, InSAR). Les outils utilisés incluent des bibliothèques IA (LSTM, GRU, CNN, transformers), des plateformes IoT (edge computing, cloud) et des logiciels de visualisation géospatiale. Sur le plan critique, l'étude présente plusieurs points forts : une approche interdisciplinaire intégrant IA, IoT et géosciences, des résultats prometteurs en

3.4 3. Forecasting future earthquakes with deep neural networks: application to California[36]

matière de réduction des erreurs de prédiction, et une revue approfondie des méthodes existantes et de leurs limites. Toutefois, des limitations subsistent, notamment le manque de données pour entraîner des modèles robustes, la complexité des signaux sismiques rendant la modélisation difficile et les contraintes logistiques et financières liées au déploiement massif de capteurs, surtout en zones inaccessibles.

3.4 3. Forecasting future earthquakes with deep neural networks: application to California[36]

Zhang et al proposent une approche de prévision sismique pour la région de la Californie au USA utilisant un réseau pleinement convolutionnel (FCN) pour prédire les séismes (M 3.0–5.0) sur des fenêtres de 15 à 90 jours, en comparaison avec le modèle ETAS. Le FCN, basé sur une architecture résiduelle (16 à 2 filtres), prend en entrée une grille spatiale (0.1×0.1) du logarithme de l'énergie cumulative des séismes passés, transformée via des lois empiriques (Kanamori, Kasahara), évitant ainsi les hypothèses rigides de l'ETAS. Pour gérer le déséquilibre des données, ils pondèrent les échantillons positifs ($= 1/\text{proportion de séismes}$) et convertissent les probabilités en niveaux d'alerte (I à IX). Les tests pseudo-prospectifs (2010–2020), évalués via des diagrammes de Molchan, montrent que le FCN atteint des performances comparables à l'ETAS (skill score 0.89 pour M 3.0) tout en étant nettement plus rapide (3 min vs 4.5 h), comme en témoigne sa prédiction des répliques du séisme de Ridgecrest (2019, M_w 7.1) avec des alertes de niveaux VII–VIII. Le catalogue ANSS (1980–2020, M 0, profondeur 40 km) couvrant la zone Californie et les outils comme TensorFlow (v1.13.1) sur GPU RTX2080Ti ont été utilisés. Cependant, le FCN présente des limites : il surestime les probabilités (nécessitant un post-traitement), manque d'interprétabilité physique (lois d'Omori/Gutenberg-Richter non intégrées), et sa performance sur des régions moins instrumentées reste à valider.

3.5 Small Earthquakes Can Help Predict Large Earthquakes[37]

L'article explore l'hypothèse selon laquelle les petits séismes contiennent des informations précieuses permettant de prédire des séismes de plus grande magnitude. En s'appuyant sur des méthodes d'intelligence artificielle et plus précisément sur le machine learning, les auteurs testent différentes approches pour estimer la probabilité de survenue d'un grand séisme (M 6.0) dans l'année suivante ainsi que la magnitude maximale attendue sur cette période. Ce travail s'inscrit dans le cadre de la prédiction sismique à moyen terme, de l'application du machine learning à la géophysique et de l'évaluation des risques dans les zones sismiques actives. Deux grandes approches ont été mises en œuvre : des méthodes classiques de machine learning telles que Random Forest, Decision Tree, SVM et régression logistique, et de l'apprentissage profond à travers des réseaux de neurones LSTM pour estimer la magnitude maximale future. Les données d'entrée sont transformées en 22 variables sismiques calculées à l'aide d'une fenêtre glissante. Les expériences ont été menées sur la région du Sichuan–Yunnan, en Chine, active sismiquement, avec une période d'étude allant de 1970 à 2021 et des prédictions effectuées sur des fenêtres d'observation glissantes. Le modèle Random Forest atteint une AUC de 0.98 pour la classification, tandis

3. ÉTAT DE L'ART

que le modèle LSTM montre une bonne tendance mais surestime fréquemment les petites magnitudes et sous-estime les grandes. Les données utilisées proviennent du catalogue sismique du China Earthquakes Data Center et portent sur les événements de magnitude supérieure ou égale à 3.0 dans la région de Chuandian entre 1970 et 2021. Les logiciels et outils employés incluent Python ainsi que les bibliothèques sklearn [27], pandas [28], numpy[29], matplotlib[30] et torch [31], avec MinMaxScaler pour la normalisation des variables et des optimisations d'hyperparamètres accompagnées d'une séparation des ensembles d'entraînement, de validation et de test. L'étude confirme que les petits séismes peuvent enrichir la prévision des grands, mais présente plusieurs limites, notamment un manque de précision du modèle LSTM pour estimer la magnitude des événements rares, une forte sensibilité au déséquilibre des classes du fait du faible nombre de séismes majeurs disponibles, un volume de données relativement limité pour les grands séismes et l'absence de prise en compte de la localisation spatiale, ce qui réduit la portée opérationnelle du modèle.

3.6 Application of Artificial Intelligence in Predicting Earthquakes: State-of-the-Art and Future Challenges [38]

L'article propose une étude complète de 84 études scientifiques traitant de l'utilisation des techniques d'intelligence artificielle pour la prédiction des séismes. Il met en lumière les performances des modèles, les approches méthodologiques employées et les défis rencontrés dans ce domaine. Les méthodes abordées couvrent un large éventail de techniques allant des approches à base de règles, telles que la logique floue, aux algorithmes d'apprentissage automatique classiques comme SVM et Random Forest, en passant par des réseaux de neurones profonds tels que LSTM et DNN. L'étude constate que les approches hybrides et les modèles bio-inspirés offrent des performances prometteuses, même si certaines limites persistent, notamment le manque de données sur les séismes majeurs et la difficulté à interpréter les résultats de certains modèles complexes. L'objectif de cette étude bibliographique est de fournir une vue d'ensemble comparative des différentes méthodes d'intelligence artificielle appliquées à la prédiction sismique, d'évaluer leurs performances à l'aide de métriques telles que RMSE, MAE et AUC, et d'identifier les principales limites méthodologiques pour proposer des pistes de recherche future. Les expériences analysées comparent les résultats des modèles IA sur des tâches variées, telles que la prédiction de la magnitude, la localisation de l'épicentre et la prévision du moment d'occurrence, en utilisant des métriques comme RMSE, MAE, AUC, sensibilité et spécificité. Des approches hybrides combinant, par exemple, des systèmes ANFIS à des méthodes d'optimisation ont également été évaluées. Les études examinées s'appuient principalement sur des catalogues sismiques publics tels que ceux de l'USGS et de l'EMSC, intégrant des données historiques sur la localisation, la magnitude, la profondeur et l'énergie des séismes, et parfois sur des données satellitaires comme DEMETER ou des signaux électromagnétiques. Les logiciels et outils utilisés incluent fréquemment MATLAB pour les systèmes ANFIS et le clustering, ArcGIS pour la visualisation, ainsi que des bibliothèques Python telles que TensorFlow et Scikit-learn.

1

¹RMSE (Root Mean Square Error) et MAE (Mean Absolute Error) sont deux métriques classiques utilisées pour évaluer la performance des modèles de régression, notamment dans la prédiction de valeurs

3.7 Does the Catalog of California Earthquakes, With Aftershocks Included, Contain Information about Future Large Earthquakes?[39]

3.7 Does the Catalog of California Earthquakes, With Aftershocks Included, Contain Information about Future Large Earthquakes?[39]

L'occurrence des grands séismes est souvent modélisée par une loi de Poisson, considérée aléatoire et indépendante. Toutefois, certaines études suggèrent qu'il pourrait exister des motifs liés à l'accumulation de contraintes tectoniques, provoquant des séismes en grappes ou selon des cycles. La prévision reste difficile, mais l'idée d'une stricte distribution aléatoire est de plus en plus remise en question. Dans ce contexte, cet article vise à quantifier l'information contenue dans les catalogues sismiques bruts incluant les répliques, à développer un indicateur de risque temporel et à évaluer la prédictibilité à l'aide de métriques rigoureuses comme les courbes ROC et l'entropie de Shannon. Les données utilisées proviennent du catalogue USGS pour la Californie couvrant la période de 1970 à 2022 sur une zone de $10^\circ \times 10^\circ$ centrée sur Los Angeles, avec une sélection de séismes de magnitude supérieure ou égale à 3.29 représentant environ 51 000 événements. Le prétraitement repose sur une moyenne mobile exponentielle appliquée aux taux de petits séismes, associée à une correction par le taux moyen $\mu(t)$ afin d'ajuster le nombre moyen d'événements et compenser les variations naturelles du catalogue. Les performances prédictives sont ensuite évaluées à l'aide de courbes ROC pour des fenêtres temporelles comprises entre 0.125 et 8.5 ans, ce qui permet de mesurer la capacité du modèle à anticiper la survenue d'un grand séisme à différentes échéances. L'incertitude associée aux prévisions est quantifiée grâce à l'entropie de Shannon calculée à partir des probabilités issues des courbes ROC. L'ensemble de ces analyses a été réalisé en Python, en utilisant les bibliothèques SciPy [31] pour les optimisations et Scikit-learn [32] pour le calcul des métriques statistiques comme ROC, AUC et PPV (Positive predictive value). L'étude met en évidence qu'une phase de quiescence sismique, caractérisée par une diminution de la variable $\Theta(t)$, précède souvent la survenue de méga-séismes. $\Theta(t)$ est une variable d'état temporelle utilisée pour suivre l'évolution de l'activité sismique locale, en particulier celle des petits séismes. Elle est construite de manière à mettre en évidence les variations anormales de l'activité sismique, notamment les périodes de calme relatif dites de quiescence qui peuvent constituer des signaux précurseurs de séismes majeurs. Ainsi, une baisse significative de $\Theta(t)$ traduit une diminution de l'occurrence des petits événements, pouvant annoncer une accumulation de contraintes en profondeur susceptible de précéder un méga-séisme. Parmi les avantages de cette démarche figurent l'innovation méthodologique reposant sur une combinaison inédite d'EMA (Exponential Moving Average), ROC et théorie de l'information, ainsi que la reproductibilité des résultats rendue possible grâce à l'utilisation d'un code open-source et de données publiques. Toutefois, certaines limites subsistent, notamment des biais potentiels liés à l'hétérogénéité des réseaux sismiques avant et après 1990, ainsi qu'une généralisation incertaine des résultats en dehors de la Californie. Enfin, le pouvoir prédictif du modèle décline au-delà de 3 ans et nécessite un recalage fréquent de la variable $\Theta(t)$ pour maintenir la qualité des prévisions.

continues comme la magnitude ou le moment d'occurrence d'un séisme.

3.8 Étude critique

Les articles étudiés mettent en évidence la diversité des approches d'intelligence artificielle appliquées à la prédiction sismique, notamment les réseaux de neurones profonds, les algorithmes de Boosting, les forêts aléatoires pondérées et les systèmes IoT. Que ce soit dans l'approche ROSERS développée dans [34] ou dans l'architecture BiKCONN évoquée dans [35], les réseaux de neurones convolutifs, récurrents ou hybrides sont mobilisés pour extraire des motifs temporels complexes à partir des signaux sismiques. L'article [36] confirme cette tendance avec l'emploi d'un Fully Convolutional Network (FCN), tandis que les travaux présentés dans [37] et [38] intègrent également des réseaux LSTM, des forêts aléatoires ou encore des modèles hybrides, illustrant une diversité algorithmique au service d'un même objectif prédictif. Par ailleurs, la question de l'interprétabilité des modèles, cruciale pour leur acceptabilité opérationnelle, revient fréquemment dans ces études. Certaines approches, comme l'utilisation des SHAP values dans le cadre de ROSERS, ou l'analyse conjointe des courbes ROC et de l'entropie de Shannon dans [39], témoignent d'une volonté partagée de rendre les résultats plus transparents et les performances mieux évaluées. Sur le plan des données, la plupart des recherches s'appuient sur des catalogues sismiques officiels tels que ceux de l'USGS, de K-Net ou de l'EMSC. Certaines études y ajoutent des sources complémentaires issues de la télédétection spatiale (InSAR, GPS) ou de capteurs IoT, montrant ainsi une intégration croissante entre observation terrestre et spatiale pour enrichir les entrées des modèles. Ces avancées permettent d'améliorer significativement la détection des signaux sismiques et la surveillance en temps réel. Toutefois, malgré ces résultats prometteurs, des limites persistent : la qualité et la disponibilité des données, ainsi que le déséquilibre des classes, freinent parfois la performance des modèles et leur capacité de généralisation. Ainsi, bien que l'IA renforce les systèmes d'alerte précoce, la prédiction précise du lieu, du moment et de la magnitude des séismes reste encore un défi majeur.

Chapter 4

Solution de prédiction sismique basée sur l'IA

4.1 Problématique

La région de Béjaïa, située au nord-est de l'Algérie, est régulièrement exposée à une activité sismique notable en raison de sa position sur la frontière entre la plaque africaine et la plaque eurasiatique. Plusieurs séismes significatifs ont marqué l'histoire de la région, notamment celui de mars 2021, qui a causé d'importants dégâts matériels et suscité un regain d'intérêt pour les systèmes de surveillance et d'alerte sismique. Dans ce contexte, la prévision des séismes constitue un défi scientifique et sociétal majeur, en raison de la nature imprévisible et multifactorielle des phénomènes géodynamiques. L'objectif de ce projet est d'explorer l'utilisation de l'intelligence artificielle, et plus spécifiquement des techniques de machine learning, afin de prédire la magnitude des séismes à partir de leurs caractéristiques géographiques et temporelles. Le projet s'articulait initialement autour de deux sources de données complémentaires :

4. SOLUTION DE PRÉDICTION SISMIQUE BASÉE SUR L'IA

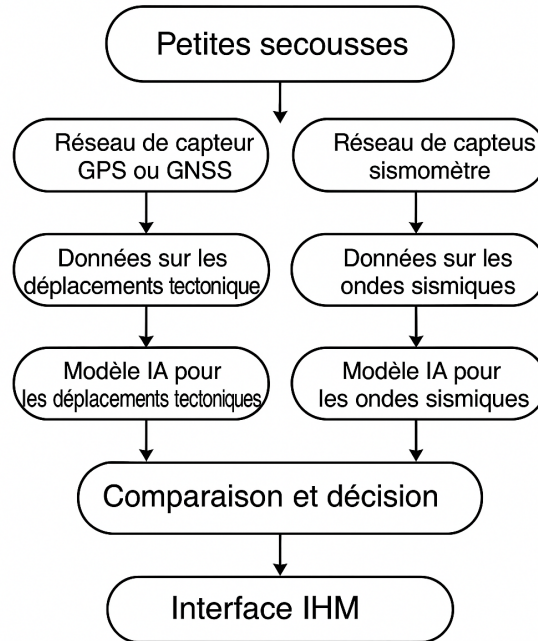


Figure 4.1: Système d'aide à la prédiction et la surveillance sismique

Cependant, en raison de l'indisponibilité de données géophysiques exploitables pour la région de Bejaïa, l'étude a été recentrée sur l'analyse de données sismiques disponibles pour la Californie, une région également sujette à une forte activité tectonique. En l'absence de données exploitables sur les déplacements tectoniques, l'approche retenue s'est concentrée sur l'analyse des séismes enregistrés, dans le but de développer un modèle prédictif de la magnitude à partir de variables spatio-temporelles.

4.2 Données utilisées

Les données utilisées proviennent des bases de données publiques telles que l'USGS (United States Geological Survey). Le dataset final comporte les informations suivantes pour chaque événement sismique : • Latitude et longitude (localisation de l'épicentre), • Profondeur (en kilomètres), • Magnitude (valeur numérique), • Date et heure de l'événement.

time	latitude	longitude	depth	mag	magType	nst	gap	dmin	rms	net	id	updated	place	type	horizontalError	depthError	magError	mag1st
2025-05-1	16.5215	-96.218	57.051	4.3	mb	59	177	2.465	0.91	us	us7000qj	2025-05-112	km N of earthquake	11.13	11.653	0.039	18	
2025-05-1	31.665	-104.36	7.0217	2.8	ml	35	66	0.1	0.2	tx	tx2025jue	2025-05-526	km S of earthquake	0.83018	0.75713	0.2	17	
2025-05-1	35.048	-119.04	6.03	3.76	mw	80	23	0.15	0.28	ci	ci4115659	2025-05-115	km NW earthquake	0.2	1.09		1	
2025-05-1	32.0965	-115.11	15.89	2.99	ml	29	55	0.1439	0.3	ci	ci4115641	2025-05-115	km SW earthquake	0.37	0.68	0.254	13	
2025-05-1	40.3713	-125.1	0.33	2.74	md	22	271	0.5858	0.32	nc	nc751842	2025-05-169	km W earthquake	4.9	13.38	0.11	15	
2025-05-1	37.5217	-118.83	4.29	2.6	md	28	63	0.06824	0.05	nc	nc751841	2025-05-114	km W earthquake	0.23	0.88	0.14	31	
2025-05-1	10.9537	-86.71	10	4.4	mb	31	198	1.271	0.73	us	us7000qj	2025-05-194	km SSW earthquake	8	1.978	0.078	48	
2025-05-1	42.0394	-126.41	10	2.9	ml	25	257	1.555	0.89	us	us7000qz	2025-05-1168	km W earthquake	7.25	2.018	0.037	94	
2025-05-1	9.0947	-81.483	10	4.5	mb	35	112	1.616	0.93	us	us7000qz	2025-05-138	km E of earthquake	5.91	1.923	0.145	31	

Figure 4.2: Extrait du dataset

Les données ont été collectées sous format Excel, puis soumises à un processus de pré-traitement avant leur utilisation dans le cadre des modèles d'apprentissage.

4.3 Méthodologie

4.3.1 Environnement de développement

Le projet a été entièrement développé en langage Python, en s'appuyant sur l'environnement Google Colab. Ce choix a permis de bénéficier de ressources de calcul cloud performantes (CPU et GPU) facilitant l'entraînement des modèles.

4.3.2 Bibliothèques et outils utilisés

Pandas :

manipulation et traitement des données

Scikit-learn :

implémentation des modèles de machine learning

Matplotlib et Seaborn :

visualisation des résultats.

LIGHT GBM ,XGBoost,AdaBoost:

pour l'entraînement des modèles de classification basés sur le boosting.

4.3.3 Prétraitement des données

- Nettoyage des données (suppression des valeurs non besoins)
- Normalisation des variables numériques (magnitude, profondeur)
- Extraction des caractéristiques temporelles (année, mois) à partir de la date
- Encodage des types de magnitude (magType)
- Conversion des dates en timestamp numériques
- Regroupement et annotation des données en classes binaires : les données ont été organisées en intervalles (ou groupes) de 10 observations, puis annotées selon la règle suivante : si au moins une observation du groupe présente une magnitude supérieure ou égale à 4, le groupe est labellisé comme classe 1 (événement significatif), sinon classe 0 (absence de séisme significatif).

4. SOLUTION DE PRÉDICTION SISMIQUE BASÉE SUR L'IA

- Sélection des variables explicatives pertinentes. (latitude, la longitude, la profondeur du foyer sismique, ainsi que des variables temporelles)

4.3.4 Équilibrage des classes (augmentation des données minoritaires)

Compte tenu du déséquilibre naturel des classes (séismes majeurs beaucoup moins fréquents), la technique SMOTE (Synthetic Minority Over-sampling Technique) a été appliquée afin de générer artificiellement des exemples supplémentaires de la classe minoritaire.

4.3.5 Séparation des données

L'ensemble des données a été divisé en deux sous-ensembles :

- 80 % pour l'apprentissage du modèle ;
- 20 % pour l'évaluation et les tests.

4.4 Modèle d'apprentissage utilisé

4.4.1 Choix de l'algorithme LightGBM

L'algorithme LightGBM (Light Gradient Boosting Machine) a été choisi pour cette étude en raison de ses performances élevées dans les tâches de régression, en particulier dans le contexte de la prédiction sismique. LightGBM appartient à la famille des algorithmes de Boosting, une technique d'apprentissage automatique qui consiste à combiner plusieurs modèles faibles (généralement des arbres de décision peu profonds) pour créer un modèle globalement plus fort et plus précis. Contrairement aux méthodes comme le bagging (ex : Random Forest), le Boosting construit chaque nouvel arbre de manière séquentielle, en corrigeant les erreurs des arbres précédents. Parmi les algorithmes de Boosting existants, on peut citer AdaBoost, XGBoost, et LightGBM. Ce dernier présente plusieurs avantages qui justifient son choix dans cette étude :

- Capacité à modéliser des relations complexes et non linéaires entre les variables
- Gestion efficace des jeux de données déséquilibrés, notamment grâce à l'ajustement automatique des poids des classes
- Temps d'entraînement rapide, même sur des volumes de données importants
- Robustesse face aux valeurs manquantes, évitant ainsi un prétraitement intensif
- Moins coûteux en mémoire et plus rapide que XGBoost grâce à sa stratégie d'optimisation interne.

LightGBM construit plusieurs arbres pour corriger petit à petit les erreurs et se concentre sur les parties les plus difficiles à prédire en priorité. Il est plus performant que d'autres modèles classiques comme Random Forest, car il peut apprendre des relations complexes entre les variables et s'adapte mieux à des données déséquilibrées, tout en optimisant l'apprentissage pour réduire les erreurs efficacement.

4.4.2 Comparaison succincte LightGBM - Random Forest

Critère	Random Forest	LightGBM
Gestion du déséquilibre	Moyen	Très bon
Temps d'entraînement	Lent	Rapide
Modélisation complexe	Bon	Excellent
Sensibilité au sur-apprentissage	Faible	Moyenne
Adaptabilité à gros datasets	Moyen	Très bon

Table 4.1: Comparaison entre Random Forest et LightGBM

Avec une accuracy de 74,75pourcent, un F1-score de 0,85 pour la classe majoritaire, un rappel de 0,83, un taux d'erreur de 25,25 pourcent, et un temps d'exécution de 1,75s, le modèle Random Forest présente des performances acceptables. Toutefois, il reste inférieur au modèle LightGBM, qui atteint une accuracy de 85,65 pourcent un F1-score de 0,92, un rappel parfait de 1,00, un taux d'erreur réduit à 14,35pourcent, et un temps d'exécution nettement plus court de 0,15s.

Cette différence de performance s'explique en partie par la nature des deux algorithmes : Random Forest construit des arbres de décision en parallèle, ce qui lui confère une bonne robustesse, mais limite sa capacité à optimiser finement les décisions. En revanche, LightGBM repose sur un apprentissage par boosting séquentiel, plus adapté à la détection des motifs rares ou subtils dans des ensembles de données déséquilibrés.

Ces résultats mettent clairement en évidence la supériorité de LightGBM dans le contexte de la prédiction sismique, tant en termes de précision que de vitesse de traitement.

4.5 Évaluation et résultats

L'évaluation du modèle a été réalisée à l'aide des indicateurs suivants :

- Accuracy (précision globale),
- Matrice de confusion(répartition des bonnes et mauvaises prédictions),
- Rapport de classification (précision, rappel, F1-score),
- Courbes ROC et précision-rappel(évaluation des performances de séparation des classes)..

4.6 Visualisations réalisées

Courbe ROC : performance du modèle (vrais positifs vs faux positifs).

4. SOLUTION DE PRÉDICTION SISMIQUE BASÉE SUR L'IA

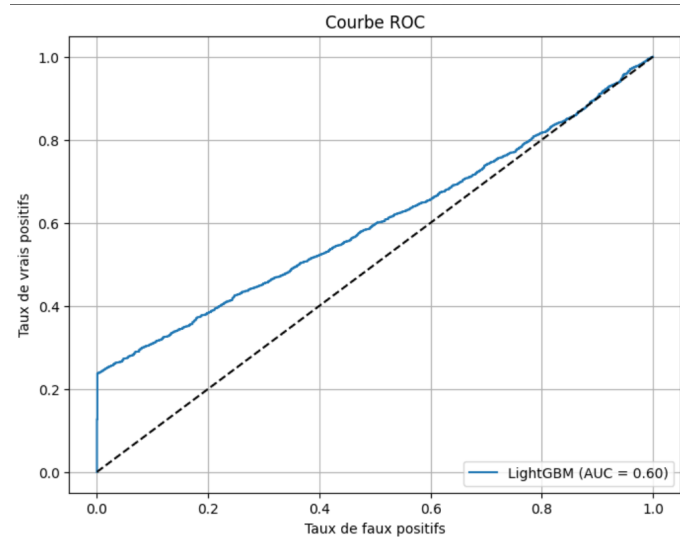


Figure 4.3: Courbe ROC

- Courbe précision-rappel permettant d'évaluer les performances sur la classe minoritaire. Cette courbe montre que plus le modèle essaie de détecter tous les séismes (rappel élevé), plus il risque de générer de fausses alertes (précision plus basse), ce qui reflète un compromis classique en classification déséquilibrée.

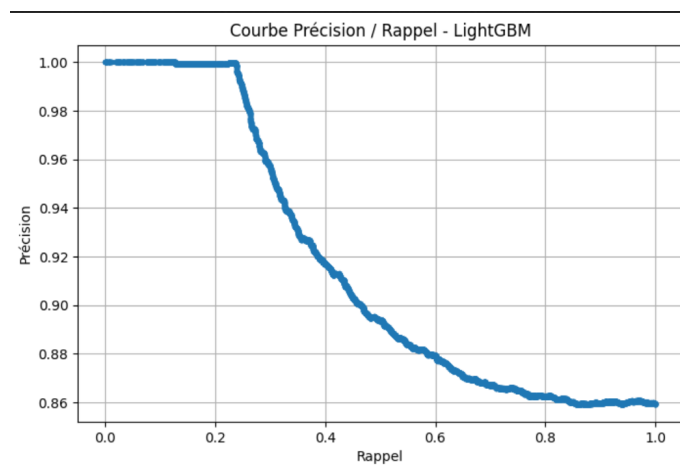


Figure 4.4: Courbe précision-rappel

- Comparaison des classes réelles et prédites sous forme de séries temporelles

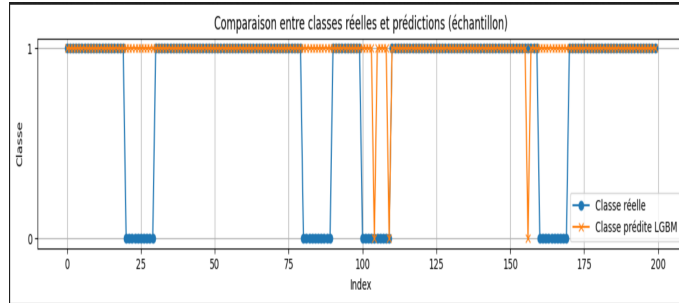


Figure 4.5: Comparaison séries temporelles

- Suivi mensuel des erreurs de classification pour détecter les périodes critiques

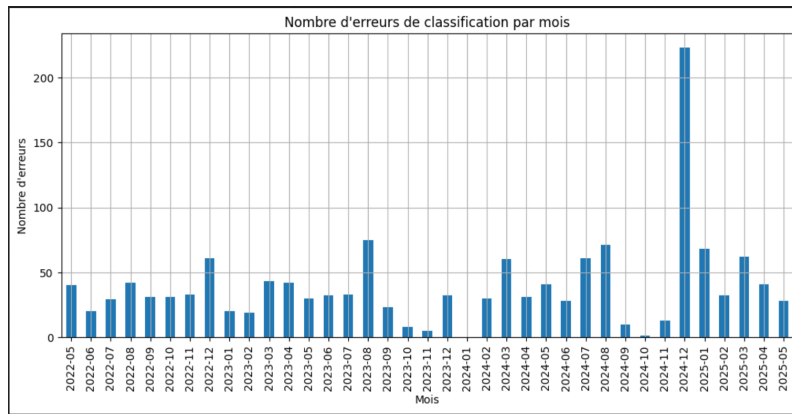


Figure 4.6: Erreurs de classification par mois

- Analyse des erreurs de classification via la matrice de confusion avec un taux d'erreur de : 0.1435 La matrice de confusion du modèle LightGBM montre qu'il a correctement prédit 8636 cas de séismes réels (classe 1) sur 8676, avec seulement 40 faux négatifs. En revanche, il a généré 1409 faux positifs, en classant à tort comme séismes des événements non sismiques (classe 0), et n'a détecté que 11 vrais négatifs. Cette configuration reflète un modèle très sensible (rappel élevé), capable de détecter presque tous les séismes, au prix de quelques fausses alertes, ce qui est acceptable dans un contexte où ne pas détecter un séisme est plus critique qu'une fausse alerte.

Matrice de confusion

4. SOLUTION DE PRÉDICTION SISMIQUE BASÉE SUR L'IA

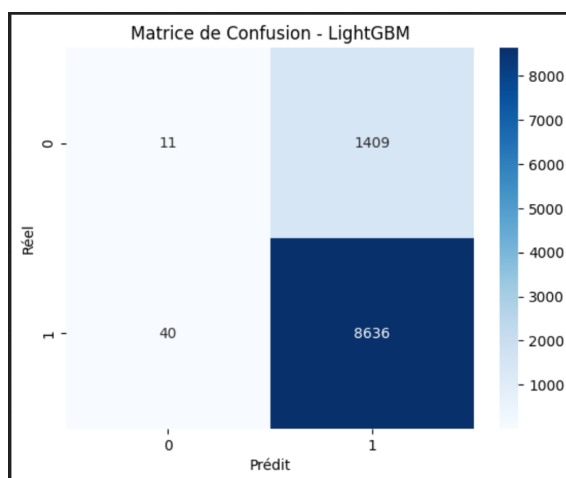


Figure 4.7: Matrice de confusion

4.7 Interprétation des résultats

Les résultats obtenus sont prometteurs et montrent que les approches basées sur le machine learning, et en particulier le modèle LightGBM, permettent d'améliorer la détection et la surveillance des séismes. Toutefois, la prédiction exacte de l'occurrence temporelle et spatiale d'un séisme reste encore un objectif difficile à atteindre en raison de la complexité des phénomènes géophysiques et de certaines limitations en matière de disponibilité de données.

Conclusion Générale

Ce travail nous a permis d'acquérir une expérience pratique dans le traitement de données sismiques, à travers l'application de techniques de prétraitement, l'équilibrage de classes par la méthode SMOTE, et l'évaluation rigoureuse des performances des modèles via des indicateurs adaptés (accuracy, F1-score, rappel, matrice de confusion, courbes ROC et précision-rappel). Ces étapes ont conduit à la construction d'un système de classification efficace des événements sismiques, démontrant que des algorithmes comme LightGBM peuvent fournir des résultats très satisfaisants même dans un contexte de classes déséquilibrées. La contribution essentielle de ce mémoire réside dans la mise en place d'une démarche complète, structurée et reproductible, allant de la préparation des données à la prise de décision. Elle a permis de proposer une solution performante pour la classification binaire des séismes, fondée sur des données tabulaires issues de sources réelles. Cette approche peut servir de base pour de futurs systèmes intelligents d'alerte précoce. En outre, ce travail ouvre de nombreuses perspectives. Il serait intéressant d'intégrer des sources de données supplémentaires, telles que des mesures géophysiques en temps réel, des données GNSS ou encore des images satellites (InSAR), afin d'enrichir les entrées du modèle. De plus, l'exploration de modèles plus avancés, notamment les réseaux de neurones profonds (CNN, LSTM, GRU), pourrait améliorer encore la précision et la robustesse du système. Enfin, une perspective particulièrement prometteuse serait le développement d'une interface IHM (interface homme-machine) permettant de visualiser les résultats des prédictions en temps réel. Une telle interface offrirait un outil pratique pour les chercheurs, les centres de surveillance et les autorités de gestion des risques, en affichant les probabilités d'occurrence sismique, les niveaux d'alerte et les informations géographiques associées de manière interactive et exploitable. En résumé, cette étude confirme l'intérêt croissant de l'intelligence artificielle dans le domaine de la surveillance sismique. Elle s'inscrit dans une dynamique de recherche appliquée visant à améliorer la détection précoce et la réduction des risques liés aux séismes, et constitue une base solide pour des travaux futurs plus ambitieux.

4. SOLUTION DE PRÉDICTION SISMIQUE BASÉE SUR L'IA

Bibliography

- [1] http://www.actu-environnement.com/ae/dictionnaire_environnement/definition/seisme_tremblement_de_terre.php4
- [2] B. Allaoua, *Simulation numérique de l'effet de tremblement de terre sur les barrages en terre*, Université de Biskra.
- [3] fr.wikipedia.org, consulté le 15 juin 2025.
- [4] <https://www.alloprof.qc.ca/fr/eleves/bv/physique/les-caracteristiques-des-ondes-p1109>
- [5] <https://www.georisques.gouv.fr/intensite-dun-seisme>
- [6] <https://www.seismescanada.rncan.gc.ca/info-gen/smeters-smetres/seismograph-fr.php>
- [7] <https://fr.wikipedia.org/wiki/Sismographe>
- [8] <https://www.cea.fr/multimedia/Documents/publications/clefs-cea/archives/fr/clefs200857p95.pdf>
- [9] <https://musee-sismologie.unistra.fr/comprendre-les-seismes/notions-pour-petits-et-grands/le-risque-sismique/methodes-de-prediction-sismique/>
- [10] <https://shakealert.org>
- [11] <https://www.ibm.com/fr-fr/topics/lidar>
- [12] <https://www.futura-sciences.com/tech/actualites/drone-faille-san-andreas-cartographiee-grace-drone-71956/>
- [13] <https://www.mdpi.com/2220-9964/9/11/616>
- [14] <https://apnews.com/article/cyprus-artificial-reefs-autonomous-underwater-vehicles-ua>
- [15] <https://www.latimes.com/california/story/2025-04-14/californias-earthquake-early-warning-system-let-many-know-about-temblor>
- [16] CNN-BiLSTM Model with Attention Mechanism for Earthquake Prediction, arXiv.

BIBLIOGRAPHY

- [17] <https://www.sciencedirect.com/science/article/abs/pii/S0267726122005085>
- [18] <https://doi.org/10.1109/LGRS.2019.2909218>
- [19] https://www.researchgate.net/publication/371079721_A_CNN-BiLSTM_model_with_attention_mechanism_for_earthquake_prediction
- [20] <https://www.intelligence-artificielle-school.com/ecole/technologies/gated-recurrent-unit/>
- [21] <https://blog.octo.com/latelier-matrice-derreur-demystifier-les-performances-du-ml-avec->
- [22] <https://www.ibm.com/fr-fr/think/topics/confusion-matrix>
- [23] <https://www.intelligence-artificielle-school.com/ecole/technologies/recall-definition/>
- [24] <https://www.ultralytics.com/fr/glossary/f1-score>
- [25] https://fr.wikipedia.org/wiki/Courbe_ROC
- [26] C. E. Shannon, *A Mathematical Theory of Communication*, Bell System Technical Journal, 27(3), 379–423, 1948.
- [27] <https://scikit-learn.org/stable/>
- [28] <https://pandas.pydata.org/>
- [29] <https://numpy.org/>
- [30] <https://matplotlib.org/>
- [31] <https://pytorch.org/>
- [32] <https://scipy.org/>
- [33] <https://scikit-learn.org/>
- [34] R. Wolfers *et al.*, "Interpretability and spatial efficacy of geographically weighted random forests...", *Engineering Geology*, vol. 329, 107010, 2023.
- [35] A. Rahmani *et al.*, "The role of artificial intelligence and IoT in prediction of earthquakes", *Measurement: Sensors*, 2022.
- [36] Y. Zhang *et al.*, "Forecasting future earthquakes with deep neural networks," *Geophys. J. Int.*, vol. 240, 2025.
- [37] X. Wang *et al.*, "Small Earthquakes Can Help Predict Large Earthquakes," *Applied Sciences*, vol. 13, no. 11, 6424, 2023.
- [38] M. H. Al Banna *et al.*, "Application of Artificial Intelligence in Predicting Earthquakes," *IEEE Access*, vol. 8, 2020.

BIBLIOGRAPHY

- [39] J. B. Rundle *et al.*, “Does the Catalog of California Earthquakes...”, *Earth and Space Science*, vol. 10, no. 2, 2023, doi: 10.1029/2022EA002521.
- [40] <https://www.evoluscience.fr/>
- [41] <https://www.profsvt71.fr/pages/premiere-s/1b-tectonique-des-plaques/de-la-surface-a-la-lithosphere-l-apport-des-seismes.html>
- [42] <fr.wikipedia.org>
- [43] <https://www.mdpi.com/2220-9964/9/11/616>
- [44] <https://apnews.com/article/cyprus-artificial-reefs-autonomous-underwater-vehicles-ua>
- [45] <https://aapm.onlinelibrary.wiley.com/doi/full/10.1002/acm2.13558>
- [46] <https://www.intelligence-artificielle-school.com/ecole/technologies/gated-recurrent-unit/>

BIBLIOGRAPHY

Résumé

Ce mémoire s'intéresse à l'application de l'intelligence artificielle pour la prédiction et la surveillance des tremblements de terre. Face aux limites des méthodes traditionnelles et à la difficulté de prévoir précisément ces phénomènes naturels, les techniques d'apprentissage automatique offrent des possibilités nouvelles grâce à leur capacité à analyser de grands volumes de données et à détecter des relations complexes. Dans ce travail, nous avons étudié plusieurs approches proposées dans la littérature, à travers l'analyse de six articles scientifiques spécialisés dans la prévision sismique par intelligence artificielle. Ensuite, une solution a été proposée en utilisant le modèle LightGBM, reconnu pour ses performances en classification. Les données sismiques historiques ont été prétraitées, normalisées et équilibrées à l'aide de la méthode SMOTE afin de corriger le déséquilibre des classes. Le modèle a ensuite été entraîné et évalué à l'aide d'indicateurs de performance tels que la précision et l'aire sous la courbe ROC. Les résultats obtenus ont montré que LightGBM est capable de capturer des relations complexes entre les variables et la survenue des séismes, avec des performances satisfaisantes. Ce travail confirme ainsi l'intérêt des techniques d'intelligence artificielle pour améliorer les systèmes de détection et d'alerte précoce, tout en ouvrant des perspectives pour l'intégration de nouvelles données et de modèles plus avancés dans les recherches futures.