Ministère de l'Enseignement Supérieur
et de la Recherche Scientifique
Université Virtuelle de Tunis

MEMOIRE

DE STAGE DE FIN D'ETUDES

Pour l'obtention du

Mastère Professionnel en Ingénierie Avancée des Systèmes Robotisés et Intelligence Artificielle (IASRIA)

Présenté par :

BEN AMMAR SIRINE

TITRE

Détection Sonore Véhicules NOMAD

Soutenu le : ..novembre 2025

Devant le jury :

Président : …..

Encadreur pédagogique : Mr. Pr. Hassene Seddik

Encadreur professionnel : MR Foued El Kamel

Rapporteur :…….

Année Universitaire : 2024 / 2025

Remerciements

A la fin de ce travail, je souhaite exprimer ma gratitude à toutes les personnes qui ont, de près ou de loin, contribué au succès de ce projet.

Je tiens, avant tout, à exprimer ma profonde reconnaissance à **Monsieur le Professeur Hassene Seddik**, mon encadrant pédagogique à l'ENSIT, pour l'attention et l'intérêt qu'il a portés à ce projet. Son suivi constant, ses orientations pertinentes et la qualité de son encadrement ont constitué pour moi une véritable source d'apprentissage et de motivation tout au long de cette expérience.

J'adresse également mes sincères remerciements à **Monsieur Foued El Kamel** mon encadrant professionnel au sein de l'entreprise **Avionav**, pour son accompagnement, sa disponibilité et ses conseils techniques avisés. Son expertise et son sens du professionnalisme ont été essentiels à la réussite de ce travail.

Je remercie également **l'ensemble des membres du jury** pour l'honneur qu'ils me font en acceptant d'évaluer ce travail. Leurs remarques et suggestions constitueront, sans aucun doute, une contribution précieuse à l'amélioration et à l'approfondissement de mes compétences.

Je tiens aussi à exprimer ma gratitude envers toute l'équipe de **l'entreprise Avionav** pour leur accueil chaleureux, leur collaboration et leur esprit d'équipe. Leur soutien et leur partage d'expérience ont rendu cette période de stage à la fois enrichissante et formatrice.

Enfin, j'adresse mes plus vifs remerciements à mes professeurs de l'ENSIT, ainsi qu'à ma famille et mes amis, pour leur soutien moral, leurs encouragements constants et leur présence bienveillante tout au long de mon parcours académique.

Ben Ammar Sirine

Dédicaces

À ma chère mère,
pour son amour inépuisable, sa douceur et ses prières constantes qui m'ont toujours accompagnée dans les moments de doute et d'effort.

À mon père,
pour sa sagesse, son soutien indéfectible et les valeurs de persévérance et d'honnêteté qu'il m'a transmises.

À mon époux,
pour sa compréhension, son encouragement permanent et sa présence bienveillante tout au long de mon parcours universitaire.

À mon enfant bien-aimé,
véritable source de bonheur et de motivation, qui m'inspire chaque jour à donner le meilleur de moi-même.

À ma belle-mère,
à qui je dois une profonde reconnaissance pour son aide précieuse et son soutien sans faille, grâce auxquels j'ai pu poursuivre mes études de master durant ces deux années.

Ce travail est le reflet de votre amour, de votre patience et de votre confiance.
Je vous le dédie avec tout mon cœur.

Résumé

Ce projet de fin d'études, réalisé dans le cadre du Master Professionnel à l'ENSIT et l'UVT et en collaboration avec l'entreprise Avionav, porte sur le développement d'un système de détection et de classification sonore destiné au rover NOMAD, conçu pour des missions de surveillance des zones frontalières et pour des usages à finalité militaire. L'étude vise à reconnaître non seulement la présence et le type de véhicules à partir de leurs signatures acoustiques, mais aussi un ensemble d'événements sonores pertinents en contexte opérationnel : communications radio, coups de feu, pas, bombardements et activités aériennes.

Pour atteindre ces objectifs, une base de données audio représentative a été constituée, regroupant des enregistrements issus de différentes sources et conditions d'environnement, afin d'alimenter l'apprentissage et l'évaluation des modèles. Le projet inclut l'extraction des caractéristiques audio, l'entraînement de modèles d'apprentissage profond (CNN et CRNN), ainsi que la comparaison de leurs performances pour sélectionner le modèle le plus adapté à une implémentation sur carte embarquée.

Les expérimentations réalisées démontrent que les architectures CNN et CRNN permettent de discriminer efficacement plusieurs classes sonores, même dans des environnements bruités, chaque modèle présentant des avantages spécifiques en termes de précision et de capture des informations temporelles. Ces résultats constituent une étape clé vers la mise en place d'une surveillance acoustique intelligente. Le travail ouvre également des perspectives pour l'enrichissement de la base de données, l'optimisation des performances des modèles et leur **intégration future dans des systèmes opérationnels embarqués**.

Mot clés : CNN,CRNN, Google Collab, Rover, Python

Table des matières

Introduction générale 9

Chapitre I : Présentation générale du projet 10

1.1.Organisme d'accueil..... 9
1.1.1 Présentation de la société..... 9
1.1.2 Domaine d'activité..... 9
1.1.3 Service de la société..... 9
1.1.4 Historique de la société..... 9
1.2. Contexte général..... 9
1.3. Présentation du projet..... 9
1.4. Problématique 4
1.5. Objectifs du projet .....
1.6. Plan du travail 6

Chapitre II : Étude de l'existant et état de l'art..... 7

Partie I : Étude de l'existant – Approche technologique
2.1. Robots mobiles et autonomes : Rover et Nomad……
2.1.1 Robots Rover……
2.1.2 Robot Nomad……
2.2 Systèmes de surveillance acoustique existants……..
2.2.1 Systèmes de détection acoustique de tirs
2.2.2 Systèmes de détection sous-marins (Sonar)
2.2.3 Systèmes de surveillance environnementale et industrielle
2.2.4 Systèmes mobile de surveillance acoustique
2.3 Analyse Critique de l'existant……..

Partie II : État de l'art – Approche scientifique
2.3 Approche de classification sonore……..
2.3.1 Approches classiques (Apprentissage automatique)
2.3.1.1 MFCC+SVM
2.3.1.2 Randon Forest
2.3.2 Approches récentes (Apprentissage profond)
2.3.2.1 CNN
2.3.2.2 RNN(LSTM+GRU)
2.3.2.3 CRNN
2.3.2.4 Yamnet
2.3.2.5 Audio MAE
2.3.2.6 Transformer AST
2.4 Comparaison des approches et justification du choix
2.5 Les bases des données existantes
2.6 Evaluation et choix de la base de données
2.7Synthèse et solution adoptée 13

Chapitre III : Conception et réalisation du système 15

3.1 Architecture globale et méthodologie……
3.2 Environnement de travail……..
3.2.1 Environnement matériel…….
3.2.2 Environnement logiciel…….
3.3 Présentation de la base de données……
3.3.1 Description générale de la base de données
3.3.2 Structure et classes de la base de données
3.3.3 Origine des données et processus de création
3.3.4 Préparation et format des fichiers
3.3.5 Répartition des données
3.4 Prétraitement des données…….
3.4.1 Modélisation des données (MFCC, Delta, Delta-delta)……
3.4.2 Augmentation des données……...
3.5 Conception des modèles IA ..... ……..
3.5.1 Conception du model CNN
3.5.1.1 architectures générales d'un CNN
3.5.1.2 architectures adoptées pour l'entrainement
3.5.2 Conception du model CRNN
3.5.2.1 architectures générales d'un CNN
3.5.2.2 architectures adoptées pour l'entrainement
3.6 Entrainement des modèles
3.6.1 Hyperparamètre de l'entrainement
3.6.2 Entrainement (nombre des paramètre entrainer et paragraphe al entrainement)

Chapitre V : Résultats expérimentaux et perspectives 31

4.1 Métriques d'évaluations
4.2 Résultat du modèle CNN
4.3 Résultat du modèle CRNN
4.4 Inférence
4.5 Comparaison et choix du modèle
4.6 Déploiement et test sur une carte (Rasberi pi ou jetson xavier)

5.1. Interprétation des résultats 32
5.2. Limites du travail réalisé 33
5.3. Améliorations et perspectives futures 34

Conclusion générale 36

Webographie.....

Références bibliographiques 38

Annexes 4

Table des Figures

FIGURE 1 : Logo de la société Avionav

FIGURE 2 : Avion Rally de l'entreprise tunisienne Avionav

FIGURE 3 : Avion Storm de l'entreprise tunisienne Avionav

FIGURE 4 : Le Rover Sojourner

FIGURE 5 : Le Rover Spirit

FIGURE 6 : Le Rover Perseverance

FIGURE 7 : Le Rover PackBot (iRobot)

FIGURE 8 : Le Rover TerraSentia

FIGURE 9 : Le Rover TALON

FIGURE 10 : Le Rover Ripsaw

FIGURE 11 : Le Rover Crusher

FIGURE 12 : Le robot Nomad Zoë

FIGURE 13 : Le robot Nomad

FIGURE 14 : Le robot Wave Glider

FIGURE 15 : Le robot GuardBot

FIGURE 16 : Boomerang

FIGURE 17 : Interface de l'application du ShotSpotter

FIGURE 18 : le diapositive AudioMoth

FIGURE 19 : les différentes classes de la base de données MAD

FIGURE 20 : Processus de création de la base de données MAD

FIGURE 21 : Répartition des données par classe et durée audio

FIGURE 22 : Histogramme illustrant la répartition des données par classe

FIGURE 23 : Signaux bruts de la base de données MAD

Liste des Tableaux

TABLE 1 : Tableau d'identité de la société AVIONAV

TABLE 2 : Comparaison entre les différentes approches pour la classification sonore

TABLE 3 : Comparaison des principales bases de données sonores

TABLE 4 : Répartition de la base de données MAD

TABLE 5 : Les étapes de modélisation des données

Listes des Abréviation

IA : Intelligence Artificielle.

CNN : Convolutional Neural Network.

CRNN : Convolutional Recurrent Neural Network.

UGV : Unmanned Ground Vehicle.

MFCC : Mel-Frequency Cepstral Coefficients.

ZRC : Zero Crossing Rate.

SVM : Support Vector Machine.

KNN : K-Nearest Neighbors.

RF : Randon Forest.

RNN : Recurrent Neural Networks.

LSTM : Long Short-Term Memory.

GRU : Gated Recurrent Unit.

Yamnet : Yet Another Mobile Network for audio.

Audio-MAE : Masked Autoencoder for Audio.

AST : Audio Spectrogram Transformer

MAD : Military Audio Dataset

Introduction générale

L'évolution rapide des technologies d'intelligence artificielle et de traitement du signal a profondément transformé les domaines de la surveillance, de la robotique et de la sécurité. Les systèmes autonomes capables d'analyser leur environnement et de prendre des décisions en temps réel constituent aujourd'hui un enjeu majeur dans les applications civiles et militaires.

Dans ce contexte, le présent projet s'inscrit dans le cadre du développement du rover NOMAD, un véhicule terrestre autonome conçu par l'entreprise Avionav. Contrairement aux approches classiques basées sur la vision artificielle, ce projet vise à doter ce rover d'un module intelligent de détection et de classification des sons, afin de lui permettre d'identifier et d'interpréter en temps réel la présence de véhicule militaire, de mouvement humains suspects ou toute activité potentiellement dangereuse à proximité de zone sensibles. Une telle capacité renforce l'autonomie du système et ouvre la voie à des applications variées, notamment dans les missions de surveillance des zones frontalières et la détection d'activités anormales à caractère militaire même dans des environnements où la visibilité est limitée (obscurité, fumée, obstacle).

Le projet repose sur l'analyse de signaux audio et l'exploitation d'algorithmes d'apprentissage profond pour reconnaître diverses classes sonores, telles que les bruits de véhicules, communications radio, coups de feu, pas, explosions, hélicoptères ou avions de combat. Une base de données audio a été élaborée et traitée afin de permettre l'entraînement, la validation et l'évaluation des modèles de classification.

Ce travail soulève plusieurs défis techniques, notamment la gestion du bruit ambiant, la diversité des signaux enregistrés et la nécessité de concevoir un modèle performant et généralisable. Ces contraintes ont guidé les choix méthodologiques à chaque étape du projet, de la préparation des données à la sélection des architectures de réseaux neuronaux.

Le rapport s'articule autour de quatre chapitres : le premier présente le cadre général et les objectifs du projet ; le second est consacré à l'état de l'art des techniques de détection et de classification sonore ; le troisième décrit la conception du système ; le quatrième expose la phase de réalisation et les résultats obtenus ; enfin, le cinquième propose une discussion des performances et les perspectives d'évolution.

Ainsi, ce projet constitue une contribution à la mise en œuvre d'un système de perception acoustique pour rover autonome, démontrant la synergie entre intelligence artificielle, traitement du signal et robotique appliquée à la sécurité.

Chapitre 1

Présentation général du Projet

Introduction

Ce premier chapitre est dédié à la présentation générale du projet. Il présente tout d'abord l'organisme d'accueil, puis expose le contexte général du projet, suivi de la description du sujet, de la problématique et des objectifs fixés. Enfin, il met en lumière les différentes tâches réalisées au cours de ce travail.

1.1 Organisme d'accueil

1.1.1 Présentation de la société

Notre projet de fin d'études a été réalisé au sein de la société **Avionav**, une entreprise tunisienne spécialisée dans la fabrication d'avions légers, dont le logo officiel est présenté dans l'image ci-dessous.

FIGURE 1 : Logo de la société Avionav[1]

Voici un tableau qui donne un aperçu des informations essentielles concernant l'entreprise AVIONAV, notamment son domaine d'activité, son statut juridique, les produits qu'elle vend et le nom de son manager.

Nomdel'entreprise	AVIONAV
Domained'activite´	Industrieaéronautique
Statutjuridique	SARL(Sociétéa`responsabilite´limitée)
Produitsvendus	Avionslégers
Siteinternet	www.avionav.net
Manager	ELKAMELFoued

TABLE 1 : Tableau d'identité de la société AVIONAV

1.1.2 Domaine d'activité

Au cours de son parcours dans le secteur de la construction aéronautique, Avionav s'est distinguée sur le marché grâce à son développement technologique innovant et à son positionnement avantageux en tant que seul acteur du marché tunisien dans ce domaine. L'atelier d'Avionav est spécialisé dans les matériaux composites à haute performance, la tôlerie fine et industrielle, l'usinage de précision, la peinture ainsi que l'assemblage.

1.1.3 Service de la société

Avionav est spécialisée dans la production de deux types d'avions distincts, offrant ainsi une gamme diversifiée de produits aéronautiques :

- Un modèle en composite de fibre de carbone baptisé Rally.

FIGURE 2 : Avion Rally de l'entreprise tunisienne Avionav **[1]**

- Un deuxième en aluminium baptisé Storm.

FIGURE 3 : Avion Storm de l'entreprise tunisienne Avionav**[1]**

1.1.4 Historique de la société

- **2007** : Fondation de l'entreprise Storm Aircraft par un groupe d'entrepreneurs italiens.
- **2011** : Les frères El Kamel lancent leur start-up Oxygène Aeronautics.
- **2014** : Acquisition complète de Storm Aircraft par les frères El Kamel, entraînant le changement de nom de l'entreprise en Avionav.
- **2015** : Exportation vers l'Italie d'un avion en aluminium à aile basse et d'un avion en composite de type Rally**.**
- **2016** : Engagement dans la fabrication d'un aéronef amphibie à quatre places en partenariat avec Evada Aircraft.
- **2017** : Exportation de plusieurs dizaines d'avions légers vers des pays d'Europe, d'Asie et d'Afrique.
- **2020** : Lancement d'un nouveau programme d'expansion visant à étendre les ateliers de fabrication et de programmation**,** ainsi qu'à recruter de nouvelles compétences**.**

1.2 Contexte général

Au cours des dernières années, les systèmes de surveillance et de reconnaissance autonomes ont connu un développement rapide, grâce aux progrès de la robotique, de l'intelligence artificielle et des capteurs embarqués, avec des applications allant de l'exploration scientifique à la surveillance militaire, en passant par la recherche environnementale et la sécurité civile. Ces technologies permettent aujourd'hui de mettre en place des systèmes intelligents capables de percevoir leur environnement via différentes modalités : vision, infrarouge, radar ou encore **acoustique**. Cette dernière, longtemps sous-exploitée, s'imposant progressivement comme un moyen efficace et complémentaire pour la détection et la classification d'événements dans des cas où la visibilité est limitée.

Dans ce cadre, les plateformes robotiques mobiles, telles que les **Rovers** et les **Nomad**, connus par leurs capacités à évoluer sur des terrains complexes, isolés ou inaccessibles, sont les meilleurs systèmes pour être doter de cette capacité de détection acoustique afin de permettre une meilleur surveillance et reconnaissance indépendamment de la visibilité et des conditions métrologiques.

1.3 Présentation du projet

Notre projet « Détection Sonore Véhicules NOMAD» consiste à mettre en place un rover acoustique mobile et autonome en intégrant un système intelligent de détection et de classification sonore basé sur des techniques d'intelligence artificielle. Ce véhicule terrestre autonome et intelligent doit être capable d'identifier et d'interpréter en temps réel des événements sonores d'intérêt tel que la présence de véhicule militaire, de mouvements humains suspects ou toutes autres activités dangereuses dans des milieux hostile et des environnements bruité et non structuré (nuit, brouillard, obstacles physiques).

Pour cela le prototype doit comprendre :

- Une plateforme mobile de type **Nomad** (robot mobile autonome tout- terrain),
- Capteurs acoustiques (microphones omnidirectionnels),
- Un modèle de Deep Learning (CNN ou CRNN) pour la détection et la classification des sons,
- Un système embarqué à ressources limitées (Raspberry Pi5 ou Jetson xavier)

1.4 Problématique

Malgré les progrès significatifs dans le domaine de la robotique et de l'intelligence artificielle, la plupart des systèmes de surveillance actuels reposent sur la vision artificielle et utilisent peu la perception acoustique. La question qui se pose est donc : comment mettre en place un tel système acoustique autonome et mobile tout en surmontant les défis scientifiques et techniques suivantes:

- Comment choisir ou constituer une base de données représentative du contexte de surveillance (sons militaires, bruits de moteurs, tirs..) pour entraîner efficacement les modèles ?
- Quelles architectures de modèles d'apprentissage doivent être adoptées pour assurer de bonnes performances de classification malgré la variabilité des conditions acoustiques ?
- Comment optimiser le modèle pour qu'il puisse être embarqué sur la carte du système tout en respectant les contraintes de mémoire, de calcul et de consommation énergétique ?

1.5 Objectif du projet

L'objectif global de notre projet est la mise en place d'un prototype de robot Nomad de surveillance acoustique, intégrant un module intelligent de détection et de classification des sons basé sur l'apprentissage profond.

Pour répondre à cet objectif global, on a définis plusieurs objectifs spécifiques:

- Choisir ou construire une base de données sonore représentative des sons cibles (véhicules, Hélicoptères, bombardement, tirs…),
- Entraîner et comparer deux architectures d'apprentissage profond (CNN et CRNN) ,
- L'évaluation des performances des modèles ,
- La sélection du modèle optimal pour une implémentation sur carte embarquée (Raspberry Pi 5 ou Jetson Xavier),
- Fournir un prototype fonctionnel capable de détecter et de classifier des sons en temps réel pour un rover de surveillance.

1.6 Plan de travail

Notre projet est réalisé de manière progressive, en suivant cet ordre chronologique des tâches à accomplir :

- Tâche 1 : Recherche bibliographique et technologique
- Tâche 2 : Collecte et préparation des données
- Tâche 3 : Conception des architectures des modèles
- Tâche 4 : Compilation et entrainements des modèles
- Tâche 5 : Évaluation et choix d'un modèles
- Tâche 6 : Préparation pour implémentation embarquée
- Tâche 7 : Analyse des résultats et rédaction du rapport

Conclusion

Ce chapitre a permis de présenter le cadre général du projet ainsi que l'organisme d'accueil. Il a également précisé le contexte, la problématique et les objectifs fixés, offrant ainsi une vision claire du travail à réaliser pour pouvoir aborder l'état de l'art et l'étude de l'existant dans le chapitre suivant.

Chapitre 2

Étude de l’excitant et état de l’art

Partie I : Étude de l'existant – Approche technologique

Introduction

Cette première partie du deuxième chapitre présente les systèmes technologiques existants liés à la robotique mobile et à la surveillance acoustique. L'objectif est d'analyser les solutions déjà développées, leurs performances et leurs limites, afin d'être guider dans la conception du système proposé.

2.1 Robots mobiles et autonomes : Rover et Nomad

Les robots mobiles autonomes, tels que les Rovers et les Nomad, constituent la base de nombreux systèmes de recherche et d'exploration. Dans cette partie on va présenter leurs caractéristiques principales et leurs domaines d'application.

2.1.1 Robots Rover

Les **Rovers** sont des véhicules robotiques mobiles conçus pour se déplacer sur des terrains souvent accidentés ou non structurés dans le but d'explorer, d'inspecter ou d'effectuer des tâches dans des environnements dangereux ou inaccessibles. Ils sont généralement équipés de roues, de chenilles ou de jambes, et disposent souvent de capteurs embarqués, de caméras et de systèmes de communication pour transmettre les informations.

Ces robots trouvent de nombreuses applications dans différents domaines :

a. Exploration spatial

Ces véhicules terrestres sont conçus pour l'exploration planétaire. Ils sont généralement équipés d'instruments scientifiques, de bras robotiques, de caméras et de systèmes de navigation autonome, ce qui leur permet d'analyser le sol, les roches et l'atmosphère, de rechercher des signes de vie passée et d'envoyer des données précieuses vers la Terre.

Les figures ci-dessus présentes quelques exemples des Rover spatial :

Sojourner (1997) : Premier Rover martien couronné de succès, faisant partie de la mission Pathfinder. Il a testé la mobilité de base et envoyé des données et images vers la Terre.

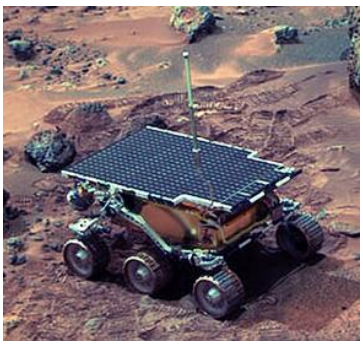

FIGURE 4 : Le Rover Sojourner **[2]**

Spirit (MER-A, 2004–2010) : A exploré le cratère Gusev et découvert des preuves d'activité passée de l'eau sur Mars avant de se retrouver bloqué dans un sol meuble.

FIGURE 5 : Le Rover Spirit **[3]**

Perseverance (2021) : Opère actuellement dans le cratère Jezero et se concentre sur l'astrobiologie en recherchant des signes de vie passée. Il collecte également des échantillons pour de futures missions de retour sur Terre.

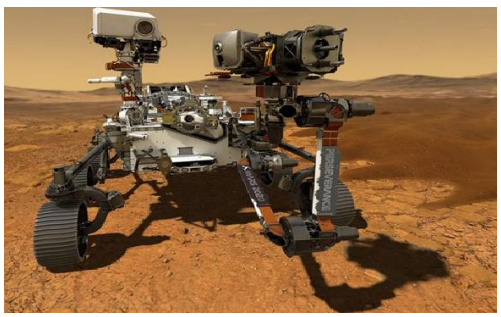

FIGURE 6 : Le Rover Perseverance **[4]**

b. Missions de sauvetage

Certains Rovers sont déployés dans des zones sinistrées (bâtiments effondrés, incendies, zones contaminées) pour localiser et assister des victimes comme le Rover **PackBot (iRobot)** représenter dans la figure ci-dessous.

FIGURE 7 : Le Rover PackBot (iRobot) **[5]**

c. Recherche environnementale et agriculture

D'autres Rovers sont employés pour collecter des données sur la qualité de l'air, du sol ou pour surveiller des écosystèmes sur de vastes zones .

On cite l'exemple du Rover **TerraSentia** :Ce petit rover agricole est conçu pour analyser la croissance des plantes, mesurer la biomasse, et cartographier la santé des cultures. Il utilise des caméras et des capteurs LiDAR pour collecter des données de terrain à grande échelle, favorisant la recherche en génétique végétale.

FIGURE 8 : Le Rover TerraSentia **[6]**

d. Applications militaires et sécuritaires

Certains Rovers terrestres sont conçues pour assister les forces armées dans diverses missions sans mettre en danger la vie humaine. Ces véhicules sont déployés pour des missions de reconnaissance, de détection d'explosifs ou d'inspection à distance de zones dangereuses.

On cite des exemples des Rovers à usage militaire :

TALON : Robot à chenilles polyvalent utilisé par l'armée américaine pour le déminage et la reconnaissance. Léger, rapide, il peut être équipé de caméras et de manipulateurs.

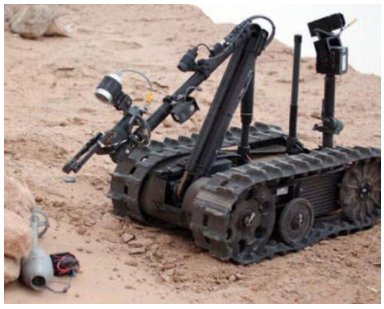

FIGURE 9 : Le Rover TALON **[7]**

Ripsaw : Véhicule sans pilote tout-terrain et à grande vitesse fabriqué par Howe & Howe Technologies. Il peut être équipé d'armes et de systèmes de surveillance et est utilisé dans des rôles de combat.

FIGURE 10 : Le Rover Ripsaw **[8]**

Crusher : Robot développé par Carnegie Mellon pour la DARPA. C'est un UGV autonome conçu pour les terrains difficiles et les charges lourdes, démontrant des capacités avancées de navigation et de mobilité grâce à l'IA.

FIGURE 11 : Le Rover Crusher **[9]**

2.1.2 Robots Nomad

Les robots **Nomad** sont des systèmes robotiques mobiles conçus pour se déplacer de manière autonome sur de grandes distances dans des environnements complexes, imprévisibles et hostiles. Contrairement aux Rovers traditionnels, qui opèrent souvent dans des zones limitées ou surveillé, les robots Nomad sont capables de maintenir leur autonomie sur le long terme, en s'adaptant aux conditions changeantes du terrain et aux obstacles rencontrés.

Ces robots combinent plusieurs technologies de pointes :

- **Différents Capteurs** (caméras, LIDAR, GPS, microphones, infrarouge) pour percevoir et analyser l'environnement,
- **Intelligence embarquée** pour la planification, la navigation et la prise de décision autonome,
- **Systèmes de locomotion robustes** (roues, chenilles ou pattes) adaptés aux terrains difficiles,
- **Communication longue portée** pour le suivi ou l'intervention humaine en cas de besoin.

Les robot Nomad, comme les robots Rover, sont utilisés dans des domaines d'exploitation similaires et relativement étroits, tels que :

a. Exploration spatial

Pour étudier des zones planétaires inaccessibles et hostile. On cite le robot Nomad **Zoë** qui a été utilisé dans le désert d'Atacama (Chili) pour simuler l'environnement de la planète Mars afin de tester des systèmes d'énergie solaire, de navigation autonome et de détection biologique.

FIGURE 12 : Le robot Nomad Zoë **[10]**

b. Recherche scientifique terrestre

Pour l'étude des environnements extrêmes tels que les déserts, les pôles et les volcans.

L'un des Nomad les plus utiliser dans se contexte est le **Nomad Robot** (Carnegie Mellon University, 1997), il a été conçu pour explorer d'une manière autonome les vastes déserts et détecter les météorites.

FIGURE 13 : Le robot Nomad **[11]**

c. Surveillance environnementale

Pour la surveillance de l'environnement tel que les océans, la collecte des données météorologiques et océanographiques, ou détection des pollutions marines. Ces plateformes fonctionnent souvent en totale autonomie pendant plusieurs mois.

On cite le **Wave Glider** utiliser pour collecter les données océaniques en temps réel pour qu'ils seront utiliser dans des applications scientifiques, commerciales et de défense.

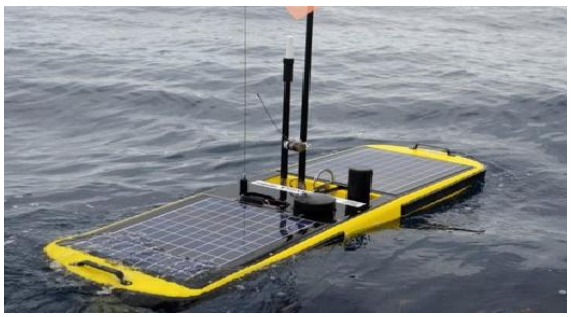

FIGURE 14 : Le robot Wave Glider **[12]**

d. Défense et sécurité

Pour la surveillance des zones frontières, la détection des explosives ou encore des missions de reconnaissance sur des terrains difficiles d'accès.

On donner comme exemple le robot **GuardBot** qui est un robot sphérique, fait pour la surveillance, la sécurité et la recherche. Il se déplace grâce à un pendule motorisé interne, lui permettant de traverser divers terrains (terre, sable, eau) et de rester stable.

FIGURE 15 : Le robot GuardBot **[13]**

2.2 Systèmes de surveillance acoustique existants

Contrairement aux systèmes robotiques étudiés précédemment, les solutions de surveillance acoustique se basent principalement sur la détection et la classification sonore indépendamment de la mobilité du support. Il existe plusieurs types de diapositives de surveillance acoustiques utilisées dans le domaine de sécurité, de la défense, de la surveillance environnementale et industrielle.

On présente ci-dessous quelques exemples de systèmes de surveillance militaire basés sur le son, couvrant différents domaines :

2.2.1 Systèmes de détection acoustique de tirs

Ces systèmes sont capables de détecter le son d'un tir et localiser précisément son origine. Ils reposent sur des réseaux de microphones, capables de capter les ondes sonores émises lors du tir puis de calculer la position de la source à l'aide de techniques de triangulation et de corrélation temporelle. Exemples de ses systèmes :

Boomerang (DARPA, 2004) : système développé pour les véhicules militaires américains, capable de détecter la direction et la distance d'un tir de projectile en temps réel, même en environnement bruyant.

FIGURE 16 : Boomerang **[14]**

ShotSpotter (SoundThinking Inc.) : C'est un réseau de capteurs installés dans plusieurs villes pour repérer et localiser les coups de feu en milieu urbain. Il utilise des algorithmes de filtrage avancés pour différencier les détonations d'armes à feu des bruits ambiants.

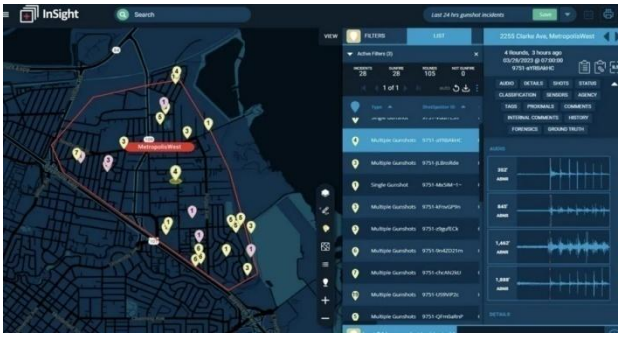

FIGURE 17 : Interface de l'application du ShotSpotter **[15]**

2.2.2 Systèmes de détection sous-marins (Sonar)

Les systèmes de détection acoustique sous-marins appeler aussi sonar, sont utilisés pour surveiller et détecter des objets ou des événements dans l'environnement marin en se basant sur la propagation du son dans l'eau. Ces systèmes reposent généralement sur des réseaux d'hydrophones capables de capter les ondes acoustiques émises par des sous-marins, des navires ou d'autres sources sous-marines. Il existe deux types de sonar :

- **Sonar passif** : utilisé pour détecter les bruits émis par les moteurs des sous-marins ou des navires sans émettre de signal actif, ce qui permet une surveillance discrète.
- **Sonar actif** : émet un signal acoustique et analyse l'écho réfléchi par les objets sous-marins, utilisé pour la cartographie et la détection de cibles spécifiques.

2.2.3 Systèmes de surveillance environnementale et industrielle

Les systèmes acoustiques sont aussi employés pour surveiller des zones naturelles ou industrielles sensibles. Ils permettent, par exemple, de détecter les mouvements sismiques, fuites de gaz ou pollutions sonores. Exemple de ses systèmes :

AudioMoth: utilisé dans la recherche environnementale pour enregistrer les sons d'animaux, détecter la présence d'espèces spécifiques ou surveiller des forêts tropicales.

FIGURE 18 : le diapositive AudioMoth **[16]**

2.2.4 Systèmes mobile de surveillance acoustique

Certaines plateformes robotiques récentes commencent à embarquer des capteurs acoustiques afin de compléter leur perception environnementale visuelle ou infrarouge. Cette tendance marque une convergence entre la robotique mobile et la surveillance acoustique.

On prend l'exemple de :

MicrodB : ce sont des drones autonomes intégrant un module de détection acoustique capables de localiser des bruits de drones ou de moteurs depuis la plateforme volante.

2.3 Analyse Critique de l'existant

L'étude des systèmes existants met en évidence que la détection acoustique a déjà trouvé de nombreuses applications dans des domaines variés tels que la défense, l'industrie ou encore la surveillance environnementale. Cependant, malgré cette maturité technologique, on remarque des limites lorsqu'il s'agit d'intégrer ces solutions dans des plateformes mobiles et autonomes adaptées à des environnements réels, complexes et dynamiques, on cite :

- Dépendance à la vision et au GPS pour la navigation autonome de la plus part des systèmes robotiques alors que l'intégration de capteurs acoustiques reste limitée.

- Les systèmes de détection acoustique existants sont souvent conçus pour des environnements contrôlés d'où la difficulté à reconnaître des sons en environnement bruité.
- Faible autonomie énergétique dans les dispositifs de surveillance acoustique.
- Absence d'adaptation à des situations dynamiques (sons changeants, milieux ouverts).

Ces constats ont conduit la société Avionav à proposer le développement d'un système mobile autonome, capable de combiner mobilité, perception acoustique et traitement embarqué, constituant ainsi l'objectif principal du projet NOMAD.

Conclusion

En résumé, cette partie de l'étude de l'existant montre que, malgré l'avancement dans le domaine des robotiques mobiles et la surveillance acoustique, il reste un besoin crucial de systèmes autonomes intégrant mobilité, perception acoustique et traitement embarqué.

Partie II : État de l'art – Approche scientifique

Introduction

Cette deuxième partie du deuxième chapitre présente les approches scientifiques pour la classification sonore, les types d'algorithmes existants dans la littérature et les bases de données utilisées dans la recherche**.** L'objectif est d'identifier les méthodes les plus adaptées au projet et de justifier le choix des modèles et de la base de données**.**

2.3 Approche de classification sonore

2.3.1 Approches classiques (Apprentissage automatique)

Les approches classiques de classification sonore reposent généralement sur l'extraction manuelle de caractéristiques acoustiques (feature engineering) à partir du signal audio (MFCC, spectrogramme, ZCR…) suivie d'un classificateur supervisé (SVM, Random Forest, KNN…)

2.3.1.1 MFCC+SVM

En combinant la capacité des MFCC à extraire des informations pertinentes sur le timbre et la dynamique temporelle du son avec la capacité des SVM à séparer efficacement les différentes classes on obtient un système de classification sonore à la fois fiable et performant.

Plusieurs travaux récents confirment l'efficacité de la combinaison MFCC + SVM pour la classification sonore. Par exemple, Jin *et al.* [R7] rapportent une précision d'environ **76,7 %** sur le dataset ESC-10, tandis que Sun et al. [R8] rapporte une précision comprise entre **81 %** et **91 %** pour la détection acoustique sous-marine, démontrant la robustesse mais aussi la variabilité des performances selon l'environnement.

2.3.1.2 Randon Forest

Le Random Forest est un classifieur par ensemble d'arbres de décision largement utilisé pour la classification sonore. Il est robuste au bruit et efficace pour des caractéristiques extraites telles que les MFCC, mais ne capture pas la dynamique temporelle des sons.

Dans des travaux récents, comme celle de Bahreini et al. [R9] le MFCC + RF atteint jusqu'à **92 %** de précision dans la classification des sons cardiaques, tandis que les travaux de

Mushtaq *et al*. [R9] rapporte une précision d'environ **72.7 %** sur le dataset ESC-10 avec RF en utilisant des caractéristiques classiques

2.3.2 Approches récentes (Apprentissage profond)

Les approches récentes pour la classification sonore reposent sur des modèles d'apprentissage profond (Deep Learning) qui permettent d'apprendre directement à partir de spectrogrammes ou de features extraites, réduisant la dépendance aux caractéristiques manuelles. Ces méthodes permettent une meilleure généralisation, surtout sur des bases de données volumineuses et variés.

2.3.2.1 CNN

Un CNN (Convolutional Neural Network) est un réseau de neurones qui analyse les spectrogrammes pour reconnaître automatiquement des sons ou des événements sonores.

Des recherches scientifiques montrent que les CNN surpassent largement les approches classiques pour des tâches multi-classes avec des données bruitées comme la recherche du Salamon et Bello [R11] qui ont montré qu'un CNN entraîné sur des Mel-spectrogrammes, avec augmentation des données, peut atteindre environ **73 %** d'exactitude sur les 10 classes du jeu de données UrbanSound8K.

2.3.2.2 RNN (LSTM+GRU)

Les RNN (Recurrent Neural Networks) sont conçus pour traiter les données séquentielles en utilisant une mémoire interne qui capture le contexte temporel. Les architectures LSTM (Long Short-Term Memory) et GRU(Gated Recurrent Unit) sont des variantes plus sophistiquées qui gèrent la mémoire à long terme à l'aide de "portes" de contrôle, le GRU étant une version simplifiée et plus rapide du LSTM qui fusionne ses portes d'entrée et d'oubli.

Dans ce cadre, Adavanne et al.[R12] ont utilisé un LSTM appliqué sur des Mel-spectrogrammes pour la détection d'événements sonores sur le jeu de données environnemental TUT-SED 2016, atteignant environ **72 %** de précision, montrant l'efficacité des RNN pour capturer les dépendances temporelles des sons.

2.3.2.3 CRNN

Un CRNN (Convolutional Recurrent Neural Network) est un réseau qui combine des CNN, pour extraire automatiquement des motifs locaux dans le spectrogramme, et des RNN, pour

modéliser les dépendances temporelles, utilisé pour la classification et la détection d'événements sonores.
Dans ce cadre, Cakir et al.[R13] ont appliqué un CRNN pour la détection d'événements sonores dans le cadre du DCASE 2017 Challenge, en utilisant le jeu de données TUT-SED. Leur modèle a atteint un F1-score d'environ **75–80 %**, dépassant les performances obtenues par des architectures CNN ou RNN utilisées seules.

2.3.2.4 Yamnet

YamNet (Yet Another Mobile Network for audio) est un modèle CNN pré-entraîné sur la base de donnée AudioSet du Google, capable de classer 521 classes sonores. Il prend en entrée des spectrogrammes audio (log-Mel) et utilise des convolutions pour extraire automatiquement des caractéristiques discriminantes. Il est conçu pour être efficace, léger et utilisable sur des appareils mobiles ou des systèmes à ressources limitées.

Une étude récente [R14] a montré que l'on pouvait utiliser YamNet-Trans, combinant YamNet pré-entraîné avec des MFCC et la Transformée de Fourier à court terme, pour classer des bruits urbains avec une précision de **94,21 %,** dépassant des modèles de références comme ResNet-50 et VGG-16.

2.3.2.5 Audio MAE

L'Audio-MAE (Masked Autoencoder for Audio) est un modèle auto-encodeur masqué il fonctionne sur le principe de l'apprentissage auto-supervisé en masquant de larges portions d'un spectrogramme audio et en entraînant un encodeur/décodeur à reconstruire les parties manquantes, apprenant ainsi une compréhension riche et contextuelle de la structure sonore sans nécessiter d'étiquettes de données.

Une étude récente [R15] a montré que Audio-MAE, un auto-encodeur masqué pré-entraîné sur AudioSet, permettait d'obtenir une précision de **94,1 %** sur le dataset ESC-50 après fine-tuning

2.3.2.6 AST

AST (Audio Spectrogram Transformer) est un modèle basé sur l'architecture Transformer appliquée aux spectrogrammes audio : le spectrogramme est découpé en patchs et traité par mécanismes d'attention (self-attention). Cette approche capture les dépendances globales dans le signal audio.

L'étude [R16] a montré que l'AST pré-entraîné sur AudioSet permettait d'obtenir une précision de 43 % mAP (précision moyenne) sur AudioSet, surpassant les modèles CNN classiques et démontrant l'efficacité de l'attention pour la classification multi-classes de sons complexes.

2.4 Comparaison des approches et justification du choix

Le tableau ci-dessous présente une comparaison entres les différents approches étudiées, en mettant l'accent sur la précision, la capacité à traiter la dimension temporelle, la complexité et l'adaptation à des systèmes embarqués :

TABLE 2 : Comparaison entre les différentes approches pour la classification sonore

Algorithme	Précision	Complexité	Capture temporelle	points forts	Limites
MFCC + SVM	Moyenne	Faible	Faible	Simple, rapide, embarquable, peu de données	Ne capture pas la séquence, dépend des features manuelles
MFCC + Random Forest	Moyenne	Faible	Faible	Robuste au bruit, facile à interpréter	Ne capture pas la dynamique temporelle, extraction limitée
CNN	Élevée	Moyenne	Partielle	Extraction automatique de motifs, bon compromis performance/légèreté	Ne capture pas la séquence complète, besoins modérés en données
RNN (LSTM/GRU)	Moyenne à élevée	Moyenne à élevée	Bonne	Capture séquences longues, adapté aux sons séquentiels	Plus complexe à entraîner, consommation élevée
CRNN	Élevée	Élevée	Très bonne	Combine motifs et séquence temporelle, idéal pour sons militaires polyphoniques	Complexité plus élevée, nécessite ressources importantes
YamNet	Élevée	Moyenne	Partielle	Pré-entraîné, léger, embarquable	Moins performant sur sons militaires spécifiques ou séquences longues
Audio-MAE	Très élevée	Très élevée	Partielle	Self-supervised, robuste, représentation riche	Très coûteux en calcul, difficile à embarquer
AST	Très élevée	Très élevée	Très bonne	Capture relations globales, multi-classes complexes	Très lourd, difficile à déployer sur cartes embarquées

En comparant ces différentes approches selon leurs avantages, limites et contraintes d'embarquement, nous choisiront de travailler de manière séparée sur deux modèles : le CNN,

pour l'extraction efficace des motifs à partir des spectrogrammes, et le CRNN**,** pour la capture de la dimension temporelle des sons militaires. Cette approche permet d'évaluer individuellement les performances de chaque modèle tout en conservant un compromis entre **précision, robustesse et faisabilité sur carte embarquée**.

2.5 Les bases de données sonores

Pour évaluer et entraîner les modèles de classification sonore, plusieurs bases de données publiques ont été utilisées dans la littérature :

TABLE 3 : Comparaison des principales bases de données sonores

Base de données	Nombre de classes	Taille	Type de sons	Environnement	Utilisation principale
ESC-50	50	2 000 échantillons	Sons environnementaux	Propre / Faiblement bruité	Benchmark général et évaluation de modèles
UrbanSound8K	10	8 732 échantillons	Sons urbains	Modérément bruité	Reconnaissance de sons urbains, robustesse au bruit
AudioSet	527	~2 000 000 clips	Sons variés (scènes, objets, animaux, machines…)	Réaliste / Très bruité	Pré-entraînement et apprentissage à grande échelle
MAD	7	8075 échantillons	Sons militaires et industriels (moteurs, armes, véhicules)	Bruité, terrain réel	Surveillance acoustique, défense,

2.6 Evaluation et choix de la base de donné

Pour entraîner et tester nos modèles de classification sonore, il faut une base représentative des sons militaires. Les bases existantes comme ESC-50, UrbanSound8K ou AudioSet ne couvrent pas ce domaine spécifique ni les événements typiques (tirs, explosions, moteurs, véhicules blindés).

La base de données **MAD (Military Audio Dataset)** [R 17] est donc retenue car elle contient exclusivement des sons militaires, avec une variété de sources et de bruits polyphoniques,

environ 8 000 échantillons répartis sur 7 classes et des enregistrements réalisés dans des conditions réalistes proches du terrain opérationnel.

2.7 Synthèse et solution adoptée

Dans le cadre de notre projet, nous allons choisir d'utiliser deux modèles distincts, CNN et CRNN, afin de comparer leurs performances et de sélectionner celui qui s'adapte le mieux aux contraintes d'un système embarqué. Les deux modèles seront entraînés et évalués sur la base de données MAD (Military Audio Dataset for Situational Awareness), représentative des sons militaires dans des environnements réalistes. Cette approche garantit que le modèle final retenu pour l'implémentation sur carte embarquée offre un compromis optimal entre **précision, robustesse et faisabilité opérationnelle**.

Conclusion

Dans ce chapitre on a présenté une analyse complète de l'existant d'un point de vue technologique et scientifique qui nous a permis d'identifier les solutions les plus performantes et de justifier les choix techniques qui seront développés dans le chapitre suivant, consacré à la conception et la réalisation du système proposé.

Chapitre 3

Conception et réalisation

Introduction

Ce chapitre présente la conception et la réalisation du système de détection et de classification sonore destiné à être embarqué sur un rover mobile autonome. Il décrit……...

3.1 Architecture globale et méthodologie

Notre système de détection et de classification automatique des sons suit une architecture modulaire inspirée des approches classiques de traitement audio. L'idée est de transformer les signaux audio bruts en représentations exploitables, puis d'entraîner des modèles de Deep Learning capables de classer ses différentes sources sonores (véhicules militaires, moteurs, bruit ambiant..).

La méthodologie suivie comprend les étapes suivantes :

- **Chargement et préparation des données audio** à partir de la base de données MAD choisit.
- **Prétraitement audio** incluant la normalisation temporelle, l'extraction de caractéristiques (MFCC, deltas, delta-deltas) pour capturer la dynamique du signal et l'application de techniques d'augmentation spectro-temporelle (SpecAugment).
- **Construction de deux modèles de Deep Learning :** un modèle **CNN** classique utilisant les MFCC comme images temps–fréquences et un modèle **CRNN**, combinant convolutions et réseaux récurrents BiLSTM.
- **Entraînement et validation :** l'entraînement est réalisé en utilisant des hyperparamètres ajustables (taux d'apprentissage, batch size, nombre d'époques...) pour optimiser la performance
- **Évaluation et comparaison :** Les modèles sont évalués sur des ensembles de données de validation et de test via des métriques comme l'accuracy, loss, matrice de confusion et tests d'inférence sur des fichiers audio réels. Cette étape permet de comparer les performances des différentes architectures et d'identifier le modèle le plus adapté à l'intégration embarquée sur le rover Nomad.
- **Déploiement sur carte embarquée (Jetson ou Raspberry Pi)**

Cette approche permet d'organiser le travail d'une manière clair et structuré, tout en facilitant la comparaison expérimentale entre les deux modèles de Deep Learning .

3.2 Environnement de travail

La clé pour la réussite d'un projet repose sur le choix d'un environnement du travail adéquat, combinant des ressources matérielles adaptées et une suite d'outils logiciels et de librairies qui seront détaillés dans la suite de cette section.

3.2.1 Environnement matériel

Tout au long du projet, et notamment pour les phases d'entraînement des modèles, nous avons utilisé un ordinateur portable dont les caractéristiques principales sont détaillées ci-dessous :

- **Système d'exploitation :** Windows 11
- **Processeur :** CPU i7
- **Mémoire RAM :**16 Go
- **Carte Graphique :** NVIDIA Geoforce RTX 3050 6 Go

3.2.2 Environnement logiciel

Le projet a été réalisé avec les outils suivants :

- **Langage de programmation** : Python 3.10
  - ➔ Il s'agit du langage principal utilisé pour le traitement numérique du signal et l'apprentissage profond, grâce à son écosystème riche et mature de bibliothèques spécialisées.

- **Framework :** TensorFlow 2.x (Keras API)
  - ➔ **TensorFlow 2.x** est un framework open-source de machine learning et deep learning, permettant de construire, entraîner et évaluer efficacement des modèles d'intelligence artificielle. **L'API Keras** intégrée offre une interface haut-niveau simplifiant la conception et l'expérimentation des réseaux de neurones.

- **Librairies :** Librosa, NumPy , Pandas, Matplotlib , Soundfile
  - ➔ **Librosa** pour le traitement audio et l'extraction MFCC des spectrogrammes et d'autres features acoustiques,
  - ➔ **Numpy et Pandas** pour la gestion des données, Numpy offrant des tableaux multidimensionnels et des fonctions mathématiques performantes, tandis que Pandas permet de manipuler et analyser facilement les données tabulaires.

➔ **Matplotlib** pour la visualisation des courbes d'apprentissage et des matrices de confusion.

- **Google Colab** pour l'entraînement GPU (Tesla T4 / L4)
  ➔ Il a été utilisé comme environnement de développement pour l'entraînement des modèles, puisque il offre un accès gratuit à des GPU (Tesla T4 / L4) ce qui permet d'accélérer les calculs.
- **L'optimisation mémoire** avec **mixed_precision**
  ➔ L'option **mixed_precision** permet de réduire l'utilisation de la mémoire GPU en combinant calculs en demi-précision et précision normale, sans perdre en précision du modèle.

3.3 Présentation de la base de données

Pour la réalisation de notre projet qui se base sur la détection et la classification sonore du véhicule NOMAD, on a choisie de travailler avec la base de données **MAD** (Military Audio Dataset for Situational Awareness) qui répondre a notre besoin : Ce dataset est récent et a été spécialement conçu pour les applications de détection **sonore militaire**, comme l'identification de véhicules, hélicoptères, tirs, ou activités humaines en extérieur.

3.3.1 Description générale de la base de données

La base de données MAD a été publié dans l'année **2024**[R 17], dans le cadre d'un travail visant à créer une base de données audio robuste pour l'entraînement des modèles capables d'opérer en situation réelle, similaires à celles rencontrées lors d'une mission opérationnelle où la détection d'événements sonores anormaux (comme des tirs, des explosions...) peut être cruciale pour la sécurité des troupes militaire. Cette base de données regroupe **7466** fichiers audio de **12** heures provenant de différentes sources, incluant :

- véhicules terrestres militaires et civils,
- moteurs divers,
- hélicoptères et avion de chasse,
- bruits environnementaux extérieurs (vent, activités humaines, ...).

3.3.2 Structure et classes de la base de données

La base de données MAD couvre plusieurs catégories représentatives des environnements militaires repartie en **7 classes** :

- **Classe 0 :** les moyens de communications militaire (communications radio ou vocales)
- **Classe 1 :** Tirs d'arme à feux,
- **Classe 2 :** Bruit de pas humains,
- **Classe 3 :** Tirs d'artillerie ou explosions,
- **Classe 4 :** Véhicules militaires,
- **Classe 5 :** Hélicoptères, drones grand tailles,
- **Classe 6 :** Avions de chasse.

La figure ci-dessous illustre les différentes classes du dataset MAD :

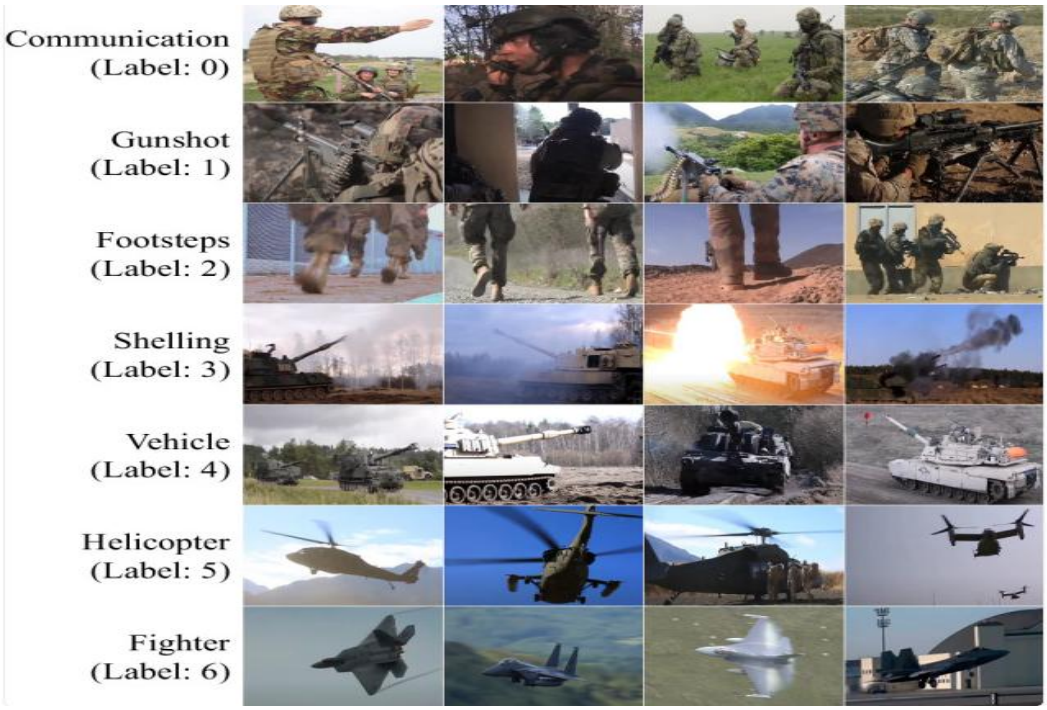

FIGURE 19 : les différentes classes de la base de données MAD [17]

3.3.3 Origine des données et processus de création

D'après, [R 17] les créateurs du dataset MAD l'ont construit en combinant plusieurs sources publiques en ligne principalement des vidéos. Voici les étapes suivie pour la création de la base données MAD:

- **Collecte et sélection des données (Data Selection) :** la collecte des données provenant des vidéos sur YouTube, en privilégiant des vidéos d'entraînement militaires

issues de différents pays et environnements réels. Les sons provenant de jeux ou de simulations ont été exclus pour garantir l'authenticité des données. Les auteurs ont utilisé un outil spécialisé **(yt_dlp)** pour le téléchargement des vidéos.

- **Segmentation et extraction des événements audio (Audio Event Segmentation) :**

Chaque événement sonore pertinent a été extrait des vidéos en utilisant l'outil **(FFmpeg)** puis segmenté avec une durée standardisée, afin de faciliter l'entraînement des modèles de deep learning.

- **Affinage des données (Data Refinement)** : afin d'affinées les données collecter les auteurs ont adopté les étapes suivantes :

- **Formatage :** Les vidéos MP4 ont été converties en fichiers audio .WAV avec un débit de 192 Kbps, un échantillonnage de 48 kHz et un seul canal.
- **Ré-échantillonnage:** Tous les fichiers audio ont été standardisés à un taux d'échantillonnage de 16 kHz avec la bibliothèque **(Librosa)** pour uniformiser les données et améliorer les performances des modèles.
- **Extraction :** Les événements audio ont été extraits selon leurs labels de segmentation avec(**Librosa) et (SoundFile)**, et sauvegardés individuellement.

- **Annotation des données (Data Labeling) :** Chaque clip audio a été manuellement annoté par des humains, en identifiant le son le plus dominant. Chaque segment audio a été vérifié par cinq annotateurs pour garantir sa qualité.
- **Configuration des données :** Diviser aléatoirement l'ensemble des données en jeux d'entraînement et de test avec l'outil (**scikit-learn)**.

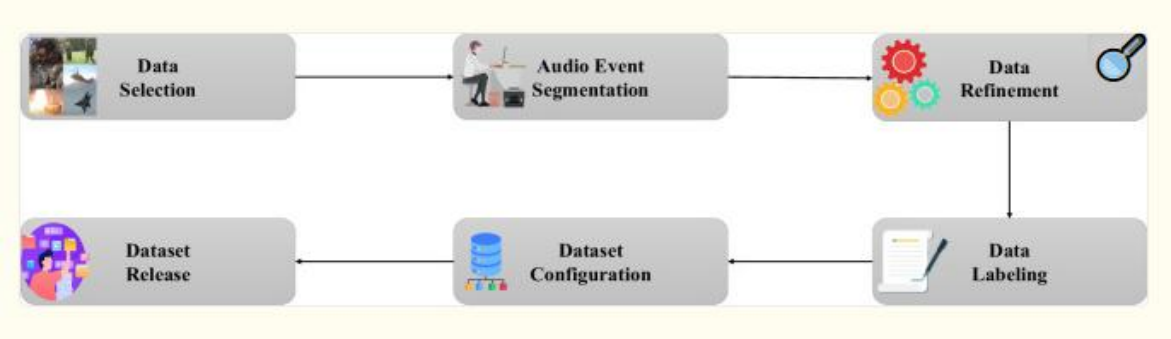

FIGURE 20 : Processus de création de la base de données MAD [17]

3.3.4 Préparation et format des fichiers

Tous les enregistrements audio provenant de différentes plateformes ont été converties en fichiers audio **WAV mono**, avec une fréquence d'échantillonnage de **16 kHz**, afin d'assurer une cohérence entre les enregistrements. Les échantillons ont été ensuite découpés à des enregistrements de **1-5 secondes**.

Ce format standardisé facilite l'intégration du dataset dans un pipeline de traitement sonore et permet une extraction fiable des caractéristiques MFCC.

3.3.5 Répartition des données

Dans l'article [17] le dataset MAD a été divisée en deux parties : **90** % pour l'entraînement et **10 %** pour le test, mais dans notre projet et afin de garantir une répartition plus équilibrée on a repartie la base de données en trois ensembles :

- **Ensemble d'entraînement (train)** : utilisé pour ajuster les paramètres des modèles CNN/CRNN.
- **Ensemble de validation (val)** : utilisé pour affiner les hyperparamètres .
- **Ensemble de test (test)** : permet d'évaluer objectivement les performances finales du modèle.

La base de données MAD comporte un total de **7 466 échantillons**. Ils sont répartis comme suite :

TABLE 4 : Répartition de la base de données MAD

Ensemble	**Nombre d'échantillons**	**Pourcentage**
Entraînement	**5 786**	**~77%**
Validation	**643**	**~10%**
Test	**1 037**	**~13%**

Les figures ci-dessous présentent la répartition des données du dataset MAD par classe, ainsi que le nombre total d'heures audio correspondant à chaque classe pour les ensembles d'entraînement, de validation et de test :

```
Tableau complet par classe (fichiers et heures audio) :
   train_files  train_hours  validation_files  validation_hours  test_files  \
0          696     0.773333                78          0.086667         207
1         1164     1.293333               129          0.143333         280
2          696     0.773333                77          0.085556         104
3          795     0.883333                88          0.097778         104
4          819     0.910000                91          0.101111         122
5          840     0.933333                94          0.104444          91
6          776     0.862222                86          0.095556         129

   test_hours
0    0.230000
1    0.311111
2    0.115556
3    0.115556
4    0.135556
5    0.101111
6    0.143333

Totaux globaux :
train_files          5786.000000
train_hours             6.428889
validation_files      643.000000
validation_hours        0.714444
test_files           1037.000000
test_hours              1.152222
```

FIGURE 21 : Répartition des données par classe et durée audio

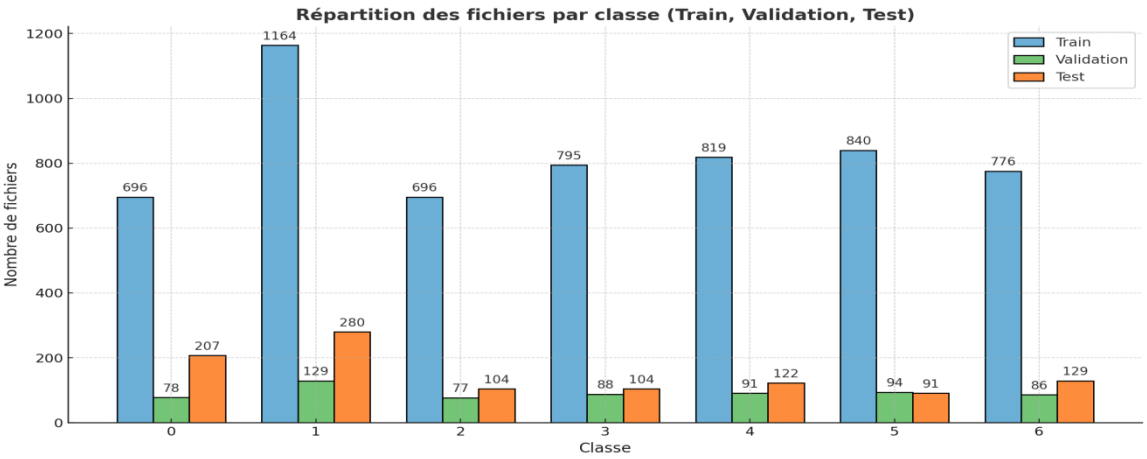

FIGURE 22 : Histogramme illustrant la répartition des données par classe

3.4 Prétraitement des données

Le prétraitement des données sert à préparer les formes d'onde audio brutes pour qu'elles puissent être utilisées efficacement par un modèle d'apprentissage profond.

La figure ci-dessous présente quelques exemples des signaux bruts du dataset MAD sous forme de waveforms, illustrant les variations d'amplitude au cours du temps.

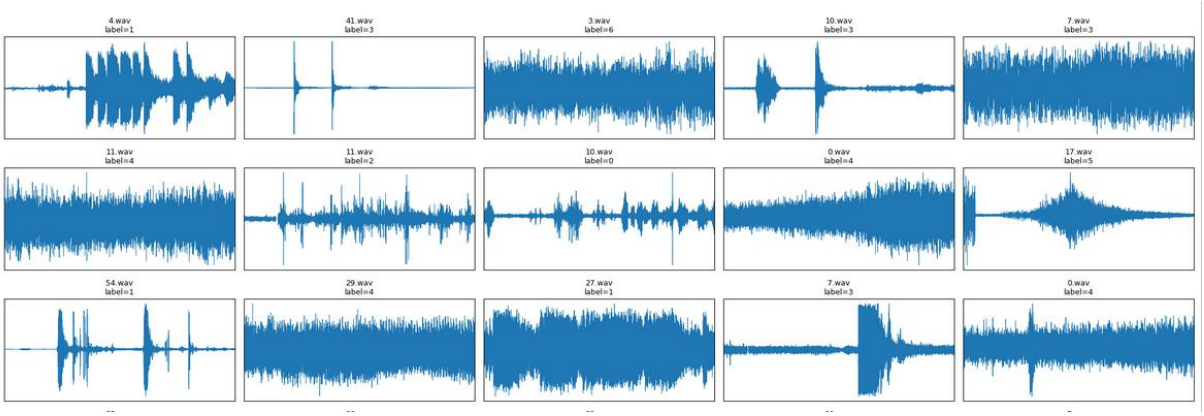

FIGURE 23 : Signaux bruts de la base de données MAD

3.4.1 Modélisation des données (MFCC, delta, delta-delta)

Dans cette étape Les signaux audio bruts sont transformés en représentations fréquentielles pour être exploitables par les modèles d'apprentissage profond. L'extraction des caractéristiques MFCC, Delta et Delta-Delta passe par plusieurs étapes :

Étape 1 (Découpage du signal et transformée de Fourier) : Le signal audio **x**[n] est segmenté en fenêtres de N=1024 échantillons avec un pas de déplacement H=512 échantillons, pour chaque segment xt[n], on applique une fenêtre de Hann w[n] puis la transformée de Fourier à court terme FFT

$$X_t(k) = \sum_{n=0}^{N-1} x_t[n]\, w[n]\, e^{-j2\pi kn/N}$$

Étape 2 (Conversion en échelle MEL) le spectre obtenu est projeté sur un banc de 64 filtres MEL, ce qui permet de reproduire la perception de la fréquence par l'oreille humaine.
Pour chaque filtre MEL (m) , l'énergie associée est :

$$E_m = \sum_{k=0}^{K-1} |X_t(k)|^2 H_m(k)$$

où $H_m(k)$ représente la réponse fréquentielle du filtre MEL.

Étape 3 (Log-MEL) : Le spectrogramme MEL est ensuite convertie en échelle logarithmique

$$\widetilde{E}_m = \log(E_m + \epsilon)$$

Étape 4 (Transformée en cosinus discrète (DCT)) : Une transformée en cosinus discrète est appliquée aux énergies log-MEL pour obtenir les **40 coefficients cepstraux MFCC** retenus dans notre configuration

$$C_n = \sum_{m=0}^{M-1} \widetilde{E}_m \cos\left[\frac{\pi n}{M}(m + 0.5)\right]$$

avec $n = 0, \ldots, 39$ (car 40 MFCC sont conservés).

Étape 5 (Normalisation des coefficients MFCC) : Afin de stabiliser l'apprentissage du modèle et de réduire les variations dues aux différences d'amplitude ou d'enregistrement on appliquer une normalisation par standardisation (z-score) des coefficients MFCC pour de garantir que chaque bande MFCC a une moyenne = 0 et un écart-type = 1

$$\mathrm{MFCC}_{\mathrm{norm}} = \frac{\mathrm{MFCC} - \mu}{\sigma}$$

Étape 6 (calcule du delta) : les coefficients Delta correspondent à la dérivée première des MFCC et reflètent la vitesse de variation des MFCC .

$$\Delta C_t = \frac{\sum_{n=1}^{N} n(C_{t+n} - C_{t-n})}{2\sum_{n=1}^{N} n^2}$$

Librosa utilise généralement $N = 2$.

Étape 7 (calcule du delta-delta) : Les coefficients Delta-Delta correspondent à la dérivée seconde des MFCC, ils capturent l'accélération ou les changements rapides dans le signal. Les coefficients delta-delta sont très utiles pour la reconnaissance de sons transitoires ou impulsifs.

Étape 8 (Construction du tenseur final) : Finalement, les trois matrices (MFCC, Delta et Delta-Delta) sont empilées pour former un tenseur tridimensionnel de dimension **(40 × T × 3)**, où : 40 = nombre de MFCC, T = nombre d'images temporelles , 3 = MFCC, Delta, Delta-Delta,

$$\text{FeatureTensor} = \begin{bmatrix} \text{MFCC} \\ \Delta\text{MFCC} \\ \Delta^2\text{MFCC} \end{bmatrix}$$

Ses tenseurs vont êtres les données d'entrer aux modèles CNN et CRNN .

TABLE 5 : Les étapes de modélisation des données

Type de caractéristiques	Signification	Etapes
MFCC	Caractéristiques spectrales principales	Extraction via FFT → MEL → log → DCT
Delta (MFCC')	Vitesse de variation des MFCC	Dérivée temporelle 1 degré
Delta-Delta (MFCC'')	Accélération des variations	Dérivée temporelle 2 dergré
Final	Tensor (MFCC + Delta + Delta-Delat²)	(40 × T × 3)

3.4.2 Augmentation des données

Afin de renforcer la robustesse du modèle et de compenser d'éventuels déséquilibres dans le dataset, on a appliquer une technique d'augmentation de données qui est **SpecAugment** (masquage temporel et fréquentiel) ou l'ajout de bruit.

3.5 Conception des modèles IA 27

3.5.1 Conception du model ????

3.5.1.1 architectures générales d'un CNN

3.5.1.2 architectures du modèle ???

3.5.2 Conception du model CRNN

3.5.2.1 architectures générales d'un CRNN

3.5.2.2 architectures adoptées pour l'entrainement

3.6 Entrainement des modèles/strategie d'entrainement et d'evaluation

Paramètres : epochs, learning rate, batch size, early stopping.

Métriques utilisées : précision, recall, F1-score.

Validation croisée ou split train/val/test.

Stratégie pour éviter le surapprentissage (regularization, dropout).

3.6.1 Hyperparamètre de l'entrainement

3.6.2 Entrainement (nombre des paramètre entrainer et paragraphe al entrainement)

3.5 Conception des modèles IA

3.5.1 Conception CNN

a/ Architecture général d'un modèle CNN

Le **CNN** est conçu pour traiter les spectrogrammes comme des images 2D :

- **Entrée** : spectrogramme ou MFCC normalisé.
- **Couches convolutionnelles** : extraction des motifs locaux (filtres 3x3 ou 5x5).
- **Couches de pooling** : réduction de dimension et mise en valeur des caractéristiques importantes.
- **Couches fully-connected** : combinaison des informations pour la classification finale.
- **Sortie** : softmax pour prédire la classe sonore.

b/ Architecture adopter pour l'entrainement

Avantages : rapide, moins coûteux en calcul, efficace pour les sons statiques ou répétitifs.

3.5.2 Conception CRNN

a/ Architecture général d'un modèle CRNN

Le **CRNN** combine CNN et RNN pour exploiter la dimension temporelle :

- **Entrée** : spectrogrammes ou MFCC normalisés (représentation 2D du signal audio).

- **Couches convolutionnelles** : extraction des motifs locaux dans le spectrogramme.
- **Couches récurrentes (LSTM ou GRU)** : modélisation des dépendances temporelles et capture de la dynamique sonore.
- **Couches fully-connected :** traitement des informations extraites pour préparer la classification.
- **Sortie :** couche softmax donnant la probabilité de chaque classe sonore (tirs, moteurs, alarmes, bruits mécaniques, bruit ambiant).

b/ Architecture adopter pour l'entrainement

On parle des couches utiliser(3 conv 3 maxpoling)tableau comparative entre resultata accur et nombre des couche

et pourquoi choisir fonction d'activation

Enfin on a utiliser une normalisation dans le rex cnn pour affiner les resultat de predection

Avantages : meilleure performance sur les sons complexes et dynamiques, capture la séquence temporelle des événements sonores.

1-Hyperparametre de l'entrainement (forme de tableau comparative entre cnn et crnn)

```
BATCH_SIZE = 32            lot de donner dans une seule iteration
EPOCHS = 150 d'aprzes notre exeperience preleminaire
LEARNING_RATE = 1e-3 0.001 (jomla mohamed)
```

(**nejmou paragraphe entrainement na7kiw feha ala phase de l'entrainement le tensor d'entrer les modele utiliser et le sortie**

. 4/ **Résultats expérimentaux et analyse**

Paragraphe simple :On va utiliser comme metrique pour l'evaluation des resultats :

accurancy (avec formule)et loss (formule) deduite de la matrice de confusion

2- comparaison des resultat de l'entrainement

Les coubre de l'accurency et loss des deux model

Et on interprete les resultat obteneu

Les matrice de confusion des 2 modele et on parle de chaque matrice seperement et on parles des proble de chaque classe (moint perfermant)

Matrice de confusion

Choix du modele retenus selon les resultat

6/inference : on parle de l'inference avec un fichier utiliser sur terrain de classe 1

```
File: C:\Users\Moham\Desktop\syrine\archive\MAD_dataset\test\077\4.wav
  1. class_1       0.9997
  2. class_3       0.0001
  3. class_5       0.0001
  4. class_4       0.0001
  5. class_2       0.0000
```

Enbarquement/implementataion /deployement sur carte

Conclusion

Ce chapitre a présenté la conception complète du système, incluant le cahier des charges fonctionnel et technique, l'architecture matérielle et logicielle, la présentation détaillée de la base de données MAD ainsi que la conception des modèles CNN et CRNN.

Chapitre 4

Réalisation et expérimentation

Webographie

[1] avionav.net/ [20/05/2025]

[2] en.wikipedia.org/wiki/Sojourner_(rover) [22/05/2025]

[3] science.nasa.gov/mission/mer-spirit/ [22/05/2025]

[4] www.jpl.nasa.gov/edu/resources/teachable-moment/meet-perseverance-nasas-newest-mars-rover/ [22/05/2025]

[5] robotsguide.com/robots/packbot [24/05/2025]

[6] https://blog.plantwise.org/2018/07/17/terrasentia-the-automated-crop-monitoring-robot/[24/05/2025]

[7] www.army-technology.com/projects/talon-tracked-military-robot/ [24/05/2025]

[8] www.military.com/off-duty/autos/textron-ripsaw-m3-big-screen-battlefield.html [24/05/2025]

[9] www.nrec.ri.cmu.edu/solutions/defense/crusher/ [24/05/2025]

[10] spectrum.ieee.org/cmu-zoe-robot-resumes-search-for-life-on-earth [24/05/2025]

[11] www.cs.cmu.edu/afs/cs/project/lri-3/www/nav97.html [24/05/2025]

[12] www.courrierinternational.com/article/2011/12/15/l-armee-des-robots-marins-debarque[24/05/2025]

[13] www.designboom.com/technology/guardbot-navigates-any-surface-turf-sand-snow-even-water-06-27-2018/[24/05/2025]

[14] www.upi.com/Defense-News/2014/06/17/Raytheons-gunshot-detection-system-being-deployed-by-utility-companies/3561403027139/ [1/06/2025]

[15] www.soundthinking.com/blog/disrupting-the-shooting-cycle-shotspotter-and-the-insight-app/ [1/06/2025]

[16] johnfkearney.com/using-the-audiomoth [3/06/2025]

[17] https://pmc.ncbi.nlm.nih.gov/articles/PMC11193796/[21/07/2025]

Références bibliographiques

[R1] NASA / Jet Propulsion Laboratory, “Mars Pathfinder – the start of modern Mars exploration,” NASA, Tech. Rep., Jul. 4, 1997.

[R2] QinetiQ North America, “TALON® Medium-Sized Tactical Robot,” QinetiQ, Product Brief, 2000-2020.

[R3] D. Wettergreen, D. Bapna, M. Maimone and G. Thomas, “Developing Nomad for Robotic Exploration of the Atacama Desert,” *Robotics and Autonomous Systems*, vol. 26, no. 3, pp. 127-148, Feb. 1999.

[R4] M. D. Wagner, D. Apostolopoulos, K. Shillcutt, B. Shamah, R. Simmons and W. L. “Red” Whittaker, “The Science Autonomy System of the Nomad Robot,” in *Proc. IEEE International Conference on Robotics & Automation (ICRA)*, Seoul, Korea, May 21-26, 2001.

[R5] “Remote Control Robot Breaks Rough Terrain Travel Record, Paves Path For Future Planetary Science Missions,” NASA, Aug. 6, 1997.

[R6] D. Wettergreen, M. Wagner, D. Jonak, V. Baskaran, M. Deans, S. Heys, D. Pane, T. Smith, J. Teza, D. Thompson, P. Tompkins et C. Williams, “Long-Distance Autonomous Survey and Mapping in the Robotic Investigation of Life in the Atacama Desert,” in *Proc. Field Robotics Center, Carnegie Mellon University*, Pittsburgh, PA, Tech. Rep., 2008.

[R7] X. Jin, Y. Song, and C. Chen, “Environmental sound classification using sparse coding and SVM,” in *Proc. IEEE Int. Conf. Computer Science and Information Technology (ICCSIT)*, Chengdu, China, 2010, pp. 151–155.

[R8] H. Sun, S. Zhou, and J. Li, “Underwater acoustic target recognition based on MFCC and SVM,” in *Proc. IEEE Int. Conf. Signal Processing (ICSP)*, Beijing, China, 2016, pp. 157–162.

[R9] M. Bahreini, R. Barati, A. Kamali, “Cardiac sound classification using a hybrid approach: MFCC-based feature fusion and CNN deep features,” *EURASIP Journal on Advances in Signal Processing*, vol. 2025:2, 2025.

[R10] Z. Mushtaq and S.-F. Su, “Efficient Classification of Environmental Sounds through Multiple Features Aggregation and Data Enhancement Techniques for Spectrogram Images,” *Symmetry*, vol. 12, no. 11, p. 1822, 2020.

[R11] J. Salamon and J. P. Bello, "Deep convolutional neural networks and data augmentation for environmental sound classification," IEEE Signal Processing Letters, vol. 24, no. 3, pp. 279–283, Mar. 2017.

[R12] S. Adavanne, P. Pertilä, et T. Virtanen, “Sound event detection using spatial features and convolutional recurrent neural network,” in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2017, pp. 771-775.

[R13] E. Çakır, G. Parascandolo, T. Heittola, H. Huttunen and T. Virtanen, “Convolutional Recurrent Neural Networks for Polyphonic Sound Event Detection,” *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 25, no. 6, pp. 1291-1303, June 2017.

[R14] L. F. Liu, Q. N. Xu, S. H. Mao, J. K. Mu, X. X. Zhao, W. H. Song, et P. Cheng, "YAMNet-based transfer learning for compact noise classification in urban and wireless systems," EURASIP Journal on Wireless Communications and Networking, vol. 2025, article 74, 2025. doi:10.1186/s13638-025-02483-8.

[R15] Z. Huang, et al., “Masked Autoencoders that Listen: Self-Supervised Audio Representation Learning,” NeurIPS, 2022.

[R 16] Y. Gong, L. Wang, R. Salakhutdinov, et al., “AST: Audio Spectrogram Transformer,” in *Proc. Int. Conf. Machine Learning (ICML)*, 2021.

[R 17] Kim, J.-W., Yoon, C. & Jung, H.-Y., « A Military Audio Dataset for Situational Awareness and Surveillance », *Scientific Data*, vol. 11, 668, 2024.

[R] E. Ackerman, “CMU’s Zoë Robot Resumes Search for Life on Earth,” *IEEE Spectrum*, 19 Jun. 2013.

[R] NASA, “Remote Control Robot Breaks Rough Terrain Travel Record, Paves Path for Future Planetary Science Missions,” Aug. 6, 1997. [Online]. Available: https://www.sciencedaily.com/releases/1997/08/970806054806.htm