



وزارة البحث العلمي والتعليم العالي
MINISTÈRE DE L'ENSEIGNEMENT SUPÉRIEUR ET DE
LA RECHERCHE SCIENTIFIQUE
جامعة عبد الحميد بن باديس مستغانم
Université Abdelhamid Ibn Badis Mostaganem
لكلية العلوم والتكنولوجيا
Faculté des Sciences et de la Technologie
DEPARTEMENT DE GENIE ELECTRIQUE



N° d'ordre : M...../GE/2025

MEMOIRE

Présenté pour obtenir le diplôme de

Master

Spécialité : Électronique des Systèmes Embarqués

Par

Dinar Abdelillah

Filière : Electronique

Et

Spécialité : Systèmes des Télécommunications

Par

Braik Aymen Yacine

Filière : Telecommunications

Intitulé du sujet

Le secourisme des personnes dans les bâtiments intelligents par la classifications d'images et de la parole

Soutenu le 30 / 06 / 2025 devant le jury composé de :

Président :	BERRADJA Khadidja	Grade MCA	Université de Mostaganem.
Examineur :	MERAH Mostepha	Grade Prof	Université de Mostaganem
Rapporteur :	HENNI Sidahmed	Grade MCA	Université de Mostaganem

Année Universitaire 2024/2025

ملخص

يتمثل مشروعنا في السنة النهائية في تطوير نظام ذكي يهدف إلى إنقاذ شخص أو عدة أشخاص يعيشون في مبنى ذكي. يتكون هذا النظام من نظامين متصلين بالتوازي. الأول يقوم بتصنيف دقيق لصور وجوه الأشخاص لتحديد ما إذا كانوا في خطر؛ والثاني يقوم بتصنيف الكلام لتحديد ما إذا كانوا في خطر. تتضمن الإسعافات الأولية إبلاغ أخصائيي الرعاية الصحية من خلال نظامنا الذكي حتى يتمكنوا من التدخل وإنقاذ الشخص المعرض للخطر. لتصنيف الصور، استخدمنا مخططات توزيع ألوان RGB القياسية ومخططات توزيع نسيج LBP القياسية، بالإضافة إلى آلة المتجهات الداعمة (SVM) كأداة برمجة. أما لتصنيف الكلام، فقد استخدمنا MFCC لتحديد معالم الكلام و SVM للتصنيف.

الكلمات المفتاحية: صورة، كلام، وجه، إسعافات أولية، تصنيف SVM

Résumé

Notre projet de fin d'étude consiste à réaliser un système intelligent qui a pour objectif de secourir une personne particulièrement ou plusieurs personnes qui vivent dans un bâtiment intelligent. ce système est compose de deux systèmes qui sont montes en parallèle. Le premier consiste à faire la classification des images précisément des visages des personnes pour voir est ce qu'ils en danger, le deuxième consiste à faire la classification de la parole afin de voir est ce qu'ils en danger. Le secourisme consiste à informer les concernes de la sante par notre système intelligent pour intervenir et secourir la personne en danger. Pour la classification des images nous avons utilisés, les

histogrammes de couleurs RGB normalisés et les histogrammes de textures LBP normalisés, et les SVM (support vecteur Machine) comme un outil de programmation. Pour la classification de la parole nous avons utilisés les MFCC pour la paramétrisation de la parole et les SVM pour la classification

Mots clés : image, parole, visage, secourisme, classification, SVM

Summary

Our final year project consists of developing an intelligent system aimed at rescuing a specific person or several people living in a smart building. This system is composed of two systems that are connected in parallel. The first consists of precisely classifying images of people's faces to determine if they are in danger; the second consists of classifying speech to determine if they are in danger. First aid involves informing healthcare professionals through our intelligent system so that they can intervene and rescue the person in danger. For image classification, we used normalized RGB color histograms and normalized LBP texture histograms, as well as SVM (Support Vector Machine) as a programming tool. For speech classification, we used MFCC for speech parameterization and SVM for classification.

Keywords: image, speech, face, first aid, classification, SVM

Remerciement

Nous tenons à exprimer notre sincère gratitude à notre encadreur, **M. Sidahmed Henni**, pour son soutien, ses conseils éclairés et sa disponibilité tout au long de ce projet. Son expertise a grandement contribué à l'aboutissement de ce travail.

Nous remercions chaleureusement **Mme Berradja** et **M. Merah**, membres du jury, d'avoir accepté d'évaluer notre travail et pour leurs précieuses remarques.

Une reconnaissance va à notre ami **Boudani Mohamed Toufik**, dont l'aide et l'organisation ont été d'un grand secours dans la réalisation de ce projet.

Enfin, nous dédions ces remerciements à nos familles et à nos proches pour leur patience, leur encouragement et leur soutien sans faille durant cette période.

Dédicaces

Ce travail est le fruit d'un long cheminement qui n'aurait pu aboutir sans le soutien précieux de nombreuses personnes. C'est avec une profonde gratitude que nous leur dédions ces quelques mots.

*À nos chers parents,
Qui ont su être notre roc dans les moments de doute,
Qui ont partagé nos nuits blanches par leur présence silencieuse,
Et qui n'ont jamais cessé de croire en nous,
Même lorsque nous doutions nous-mêmes.
Votre amour inconditionnel a été notre plus grande force.*

*À notre ami et compagnon de route, Boudani Mohamed Toufik,
Ton aide précieuse dans l'organisation de notre travail,
Tes conseils avisés aux moments cruciaux,
Et ta présence amicale ont été déterminants.
Ce projet porte aussi une part de ta contribution.*

*À nos enseignants et mentors,
Qui ont semé en nous le goût du savoir,
Éveillé notre curiosité intellectuelle,
Et nous ont montré la voie à suivre.*

*À tous nos amis qui nous ont soutenus,
Par leurs encouragements,
Leurs paroles réconfortantes,
Et leur simple présence réconfortante.*

*Enfin, à tous ceux qui, de près ou de loin,
Ont contribué à faire de ce projet une réalité,
Merci pour votre confiance et votre soutien.*

Ce mémoire est autant le vôtre que le nôtre

Liste des abréviations

- [1] **INSEE** : *Institut National de la Statistique et des Études Économiques*
(Organisme français produisant les statistiques officielles)
- [2] **OMS** : *Organisation Mondiale de la Santé*
(Autorité internationale en santé publique)
- [3] **DREES** : *Direction de la Recherche, des Études, de l'Évaluation et des Statistiques*
(Service français chargé des données sanitaires et sociales)
- [4] **PIB** : *Produit Intérieur Brut*
(Indicateur économique mesurant la richesse d'un pays)
- [5] **EHPAD** : *Établissement d'Hébergement pour Personnes Âgées Dépendantes*
(Structure médico-sociale pour seniors)
- [6] **OCDE** : *Organisation de Coopération et de Développement Économiques*
(Organisation internationale analysant les politiques publiques)
- [7] **KNX** : *Standard international pour la domotique et la gestion technique des bâtiments*
(Protocole utilisé pour automatiser l'éclairage, la sécurité, etc.)
- [8] **SIDC** : *Système Intelligent de Détection et de Classification*
(Terme générique pour les systèmes comme le vôtre)
- [9] **ANSSI** : *Agence Nationale de la Sécurité des Systèmes d'Information*
(Organisme français garant de la cybersécurité)
- [10] **RGAA** : *Référentiel Général d'Amélioration de l'Accessibilité*
(Normes pour l'accessibilité numérique)
- [11] **TPE** : *Très Petite Entreprise*
(Structure de moins de 10 salariés)

[12] SVM : *Machines à Vecteurs de Support (Support Vector Machine)*
(Algorithme de classification utilisé pour vos images/voix)

[13] RGB : *Red-Green-Blue*
(Espace colorimétrique pour l'analyse d'images)

[14] LBP : *Local Binary Patterns*
(Méthode de description des textures faciales)

[15] SIFT : *Scale-Invariant Feature Transform*
(Algorithme de détection de caractéristiques d'images)

[16] MFCC : *Mel-Frequency Cepstral Coefficients*
(Coefficients utilisés pour l'analyse vocale)

[17] CNN : *Réseaux de Neurones Convolutifs (Convolutional Neural Network)*
(Réseau profond pour la classification d'images)

[18] GPU : *Graphics Processing Unit*
(Processeur optimisé pour les calculs parallèles, utilisé en IA)

[19] HMM : *Modèles de Markov Cachés (Hidden Markov Model)*
(Modèle statistique pour la reconnaissance vocale)

[20] GMM : *Modèles de Mélange Gaussien (Gaussian Mixture Model)*
(Approche probabiliste pour la modélisation des données)

[21] RNN/LSTM : *Réseaux Neuronaux Récurrents / Long Short-Term Memory*
(Architectures pour le traitement de séquences temporelles, comme la parole)

Liste des figures

Figure I.1 : Comparaison du Vieillissement Démographique Mondial entre 2010 et 2040

Figure I.2 : Gestion des systèmes technologiques d'un bâtiment intelligent

Figure I.3 : L'évolution de la technologie et son effet sur les entreprises

Figure I.4 : La prévention et la gestion des risques au sein de l'entreprise

Figure I.5 : Défi de la cybersécurité dans les bâtiments intelligents

Figure I.6 : Visualisation des données des capteurs sur le modèle 3D d'un bâtiment intelligent

Figure I.7 : Maintien à domicile intelligent pour les personnes âgées.

Figure I.8 : La normalisation de l'IA

Figure I.9 : Stratégies d'allocation budgétaire et optimisation des coûts en gestion de projet pour la réussite des startups

Figure II.1: Histogramme des intensités des canaux de couleur (R, G, B) dans une image.

Figure II.2: (a) LBP image (b) LBP histogramme.

Figure II.3: Points-clés détectés par l'algorithme SIFT sur des images à différentes transformations

Figure II.4: Spectrogramme d'un signal de parole avec les coefficients MFCC.

Figure II.5: Classification binaire par SVM : séparation des cas de récidive (1) et non récidive (-1) par un hyperplan à marge maximale dans un espace de caractéristiques.

Figure II.6: Schéma de classification de textures à l'aide des SVM avec extraction manuelle des caractéristiques.

Figure II.7: Exemple de classification de chiffres manuscrits à l'aide d'un hyperplan SVM

Figure II.8: Exemple de classification d'image par un réseau de neurones convolutif (CNN) dans une tâche ImageNet.

Figure II.9: Exemple de transfert d'apprentissage (transfer learning) d'un modèle CNN pré-entraîné sur ImageNet vers l'imagerie médicale.

Figure II.10 : Schéma d'un système de vérification d'identité vocale : phases d'enregistrement, extraction des MFCC et comparaison au voiceprint.

Figure II.11 : Architecture d'un réseau LSTM à deux couches pour la reconnaissance vocale.

Figure II.12 : Architecture CNN LSTM appliquée à un spectrogramme pour la reconnaissance vocale (détection de mots clés / transcription)

Figure III.1 : Un exemple des images du visages à classer

Figure III.2 : Apprentissage des classes par les SVM

Figure III.3 Reconnaissance des classes par les SVM

Figure III.4 représentation du pourcentage des classes correctement détectés pour le noyau RBF en utilisant les histogrammes de couleur et de texture RGB, LBP et les SVM

Figure III.5 représentation du pourcentage des classes correctement détectés pour le noyau polynomial en utilisant les histogrammes de couleur et de texture RGB, LBP et les SVM

Figure III.6 représentation du pourcentage des classes correctement détectés pour le noyau en sigmoïde utilisant les histogrammes de couleur et de texture RGB, LBP et les SVM

Figure III.7 la variation du taux de la reconnaissance en fonction du sou classifieur SVM noyau RBF

Liste des tableaux

Tableau I.1 : Évaluation des Technologies de Surveillance de la Santé : Paramètres, Précision et Acceptabilité

Tableau III.1 Tableau représentant le rappel et la précision Pour la détection des comportements pour le noyau RBF en utilisant les histogrammes de couleur et de texture RGB, LBP et les SVM Lights.

Tableau III.2 Tableau représentant le rappel et la précision Pour la détection des comportements pour le noyau polynomial en utilisant les histogrammes de couleur et de texture RGB, LBP et les SVM.

Tableau III.3 Tableau représentant le rappel et la précision Pour la détection des comportements pour le noyau sigmoïde en utilisant les histogrammes de couleur et de texture RGB, LBP et les SVM

Tableau III.4 Tableau représentant le rappel et la précision pour la classification des mots isolés pour le noyau RBF en utilisant les MFCC et les SVM

Sommaire

ملخص 1	
Remerciement	3
Dédicaces	4
Liste des abréviations.....	5
Liste des figures	7
Liste des tableaux.....	9
Sommaire	10
Introduction générale	14
Chapitre I : Bâtiments intelligents	17
I.1 Introduction.....	18
I.2 Le Phénomène Mondial du Vieillissement	18
I.2.1 Cas Français (INSEE 2023).....	19
I.2.2 Impacts multidimensionnels	19
I.2.3 Coût de la dépendance.....	20
I.2.4 La Réponse Technologique.....	20
I.3 Fondements Technologiques des Bâtiments Intelligents	20
I.3.1 Architecture Systémique Avancée.....	20
I.3.2 Caractéristiques Techniques.....	21
I.3.3 Évolution Historique et Technologies Emergentes	22
Évolution des Bâtiments Intelligents.....	22
Technologies Clés 2024.....	23
I.4 Applications pour les Personnes Âgées	23
I.4.1 Exemples d'application	23
I.4.2 Santé et Bien-être Personnalisé.....	25
I.5 Défis et Solutions Techniques	26
I.5.1 Interopérabilité et cybersécurité.....	26
I.5.2 Acceptabilité Utilisateur	27
I.6 Perspectives 2030.....	28
I.6.1 Feuille de Route Technologique	28

Robotique.....	28
Réseaux	28
I.6.2 Recommandations Politiques pour un Plan d’Action National.....	29
Financement des Solutions Smart Aging.....	29
Crédit d’impôt de 50 % pour l’équipement.....	29
Ce dispositif fiscal s’appliquerait.....	29
Conditions d’éligibilité	29
I.6.3 Régulation et Normalisation des Technologies et de l’IA	30
Certification obligatoire des dispositifs.....	30
Recherche & Expérimentation.....	31
I.7 Allocation budgétaire et expérimentation	32
I.8 Conclusion	32
Chapitre II: Les methodes de la classification des images et de paroles	34
II.1 Introduction.....	35
II.2 Paramètres des SVM.....	36
II.2.1 Paramétrisation des Images avec les Histos de Couleurs RGB.....	36
Introduction	36
Détails	36
Application.....	37
II.2.2 Paramétrisation des Images avec les Histos de Texture LBP.....	37
Introduction	37
Détails	37
Application.....	38
II.2.3 Paramétrisation des Images avec SIFT.....	39
Introduction	39
Détails	39
Application.....	40
II.2.4 Paramétrisation du Signal de la Parole par la Méthode MFCC	40
Introduction	40
Détails	40
Application.....	41
II.2.5 Hyperplan (SVM) Sépare les Classes	41

Introduction	41
Détails	41
II.3 Les SVM en apprentissage supervisé	43
II.4 Pourquoi Choisir les SVM pour la Classification d'Images et de Parole ?	46
II.5 Classification des Images.....	47
II.5.1 Méthodes Traditionnelles	47
II.5.2 Méthodes Basées sur l'Apprentissage Profond.....	49
a) CNN (Réseaux de Neurones Convolutifs)	49
b) Transfer Learning avec CNN	50
II.6 Classification de la Parole.....	51
II.6.1 Méthodes Classiques	51
a) SVM / HMM (Modèles de Markov Cachés).....	51
b) Modèles à base de GMM (Modèles de Mélange Gaussien).....	52
II.6.2 Méthodes d'Apprentissage Profond	52
a) RNN/LSTM (Réseaux Récurrents et Long Short-Term Memory).....	52
b) CNN Adaptés à la Parole	53
II.7 Analyse comparative des méthodes	54
II.8 Traitement avec SVM.....	55
II.8.1 Pour les images.....	55
II.8.2 Pour les parole.....	56
II.9 Avantages par rapport aux autres méthodes	56
II.9.1 Face aux CNN.....	56
II.9.2 Face aux RNN.....	56
II.10 Implémentation proposée.....	57
II.10.1 Outils et frameworks	57
II.10.2 Approche recommandée	57
II.10.3 Optimisation des performances.....	57
II.11 Conclusion du chapitre	58
Chapitre III : Experience et resultats	60
III.1 Introduction.....	61
III.2 Le premier système : la classification des images	61
III.2.1 Paramétrisation des données.....	62

III.2.2	Apprentissage	64
III.2.3	Reconnaissance	65
III.2.4	Les résultats obtenus	66
III.2.5	Conclusion	69
III.3	Le deuxième système : la classification de la parole	69
III.3.1	Acquisition et corpus d'entrée :	69
III.3.2	Prétraitement et paramétrisation	70
III.3.3	Formation du corpus d'entrée :	70
III.3.4	La classification par SVM :	70
III.3.5	Test et résultats.....	71
III.3.6	Commentaire :.....	72
III.3.7	CONCLUSION.....	72
	Conclusion Générale.....	73
	Références.....	75
	ANNEXE : LES PROGRAMMES COMPLETS	81
	SYSTÈME DE RECONNAISSANCE FACIALE EN CAS DE DANGER	81
	SYSTÈME DE RECONNAISSANCE VOCALE.....	88

Introduction générale

Le XXI siècle est marqué par une transformation démographique sans précédent, caractérisée par un vieillissement accéléré des populations. Selon les dernières données de l'Organisation Mondiale de la Santé (OMS) [1], la proportion mondiale des personnes âgées de 60 ans et plus passera de 12% à 22% entre 2020 et 2050, représentant ainsi près de 2,1 milliards d'individus. Cette évolution génère des défis sanitaires majeurs, particulièrement dans le domaine de la sécurité domestique. Effectivement, les chutes accidentelles, qui constituent la principale cause de décès accidentel chez les personnes âgées, entraînent chaque année plus de 37 millions d'interventions médicales d'urgence dans le monde [2], dont 20 000 cas. mortels rien qu'en Europe selon le rapport Eurosafe 2023. Ces accidents surviennent majoritairement dans des environnements résidentiels traditionnels, dépourvus de systèmes de détection proactive. Face à ce constat, se pose une question centrale : comment concilier autonomie résidentielle des personnes âgées et garantie d'une sécurité continue, sans recourir à une surveillance humaine permanente socialement coûteuse et psychologiquement intrusive ? Comme le souligne Caradec [3] (2021), cette équation complexe exige une refonte technologique des habitats, intégrant des solutions capables d'anticiper les risques tout en respectant l'intimité des individus. Les bâtiments intelligents (smart buildings) constituent une réponse systémique à ces enjeux, en transformant l'habitat traditionnel en écosystèmes réactifs et prédictifs. Fondés sur une architecture cyber-physique intégrant capteurs IoT (Internet des Objets), réseaux de communication LPWAN (Low-Power Wide-Area Network) et plateformes d'intelligence artificielle embarquée, ces infrastructures génèrent en temps réel une

13.5modélisation numérique du comportement des résidents et de leur environnement [4].

Leurs avantages opérationnels sont démontrés par plusieurs études récentes :

- Réduction de 70% du temps d'intervention lors d'incidents critiques grâce à des boucles de détection-alerte automatisées, avec une latence moyenne de 8,3 secondes selon les benchmarks de Chen et Nugent (2023) [5] ;

- Personnalisation des protocoles de sécurité via l'intégration de dossiers médicaux numérisés (allergies, pathologies cardiaques, antécédents de chutes), permettant une hiérarchisation contextuelle des risques [6] ;
- Optimisation économique avec une diminution estimée à 40% des coûts sociaux liés à la dépendance, selon les modèles de Rashidi (2024) [7]. Ces infrastructures constituent un environnement technologique optimal pour l'implémentation de systèmes de secours avancés, alliant une réactivité exceptionnelle à une discrétion dans les opérations.

La sécurisation des personnes âgées dans ces environnements connectés repose sur des piliers technologiques complémentaires, parmi ces derniers, la classification d'images et l'analyse de la parole. Les systèmes de vision par ordinateur exploitent des réseaux neuronaux convolutifs (CNN) optimisés, capables d'identifier des séquences comportementales à risque (chutes, pertes d'équilibre, immobilité anormale) avec une précision dépassant 94,2% selon les tests de Yu et al. (2024) menés sur 15 000 heures de vidéos annotées [8].

Il existe un système intelligent de secours pour les personnes âgées, combinant la détection d'images pour repérer les chutes et la reconnaissance vocale pour les appels à l'aide. Le système utilise des données environnementales, comme la position GPS et les signes vitaux, pour déclencher des alertes adaptées en cas d'urgence, tout en garantissant la confidentialité des données (Chen et al, 2021). Une base de données de 200 heures de scénarios critiques sera utilisée pour entraîner les modèles (Dubey et al, 2023). Un projet pilote dans la résidence "Les Jardins d'Arcadie" à Toulouse, avec 42 résidents et 28 capteurs. Les performances seront mesurées par la précision de détection (cible > 95 %, Zhang et al, 2022), le temps d'alerte (objectif < 7 secondes), l'acceptabilité sociale (ISO 22380:2018) et la réduction des hospitalisations. L'objectif est d'établir un nouveau standard pour les logements autonomes, alliant sécurité et respect de la vie privée (référence éthique originale [9]). Les méthodes de classification SVM (Support Vector Machine) et CNN (Convolutional Neural Network) sont couramment utilisées en apprentissage automatique. SVM cherche à trouver un hyperplan optimal pour séparer les classes, tandis que CNN utilise des couches de convolution pour extraire caractéristiques

des images. Les SVM sont efficaces pour des ensembles de données plus petits et nécessitent un prétraitement des caractéristiques, alors que les CNN excellent dans des tâches complexes comme la reconnaissance d'images, atteignant souvent des précisions supérieures à 98%. Pour plus d'informations, vous pouvez consulter des articles tels que "Comparing SVM and CNN in Recognizing Handwritten Digits"[10]. Dans le premier chapitre concerne les bâtiments intelligent, le deuxième chapitre concerne les méthodes de la classification des images et de la parole et finalement le chapitre trois qui regroupe les résultats et l'expérience obtenu de ce projet.

Chapitre I : Bâtiments intelligents

Chapitre I : Batiment intelligent

I.1 Introduction

Le vieillissement démographique nécessite l'adaptation des bâtiments intelligents afin d'optimiser la qualité de vie des personnes âgées. Cela implique l'intégration des technologies avancées pour répondre efficacement à leurs besoins croissants.

I.2 Le Phénomène Mondial du Vieillissement

Constitue une transformation démographique majeure du XXI^e siècle. Selon les Nations Unies [11], la population des sexagénaires et plus est passée de 703 millions (9,1% de la population mondiale) en 2019 à une projection alarmante de 1,5 milliard (16%) d'ici 2050. Cette transition affecte particulièrement les régions développées, avec des prévisions de 29,4% de seniors en Europe, 22,1% en Amérique du Nord et 17,5% en Asie. Cette évolution sans précédent remodèle profondément les équilibres socio-économiques, imposant des adaptations urgentes dans les domaines sociétaux que sont la santé, le logement et les politiques sociales afin de répondre aux attentes spécifiques de cette tranche vieillissante de la population.

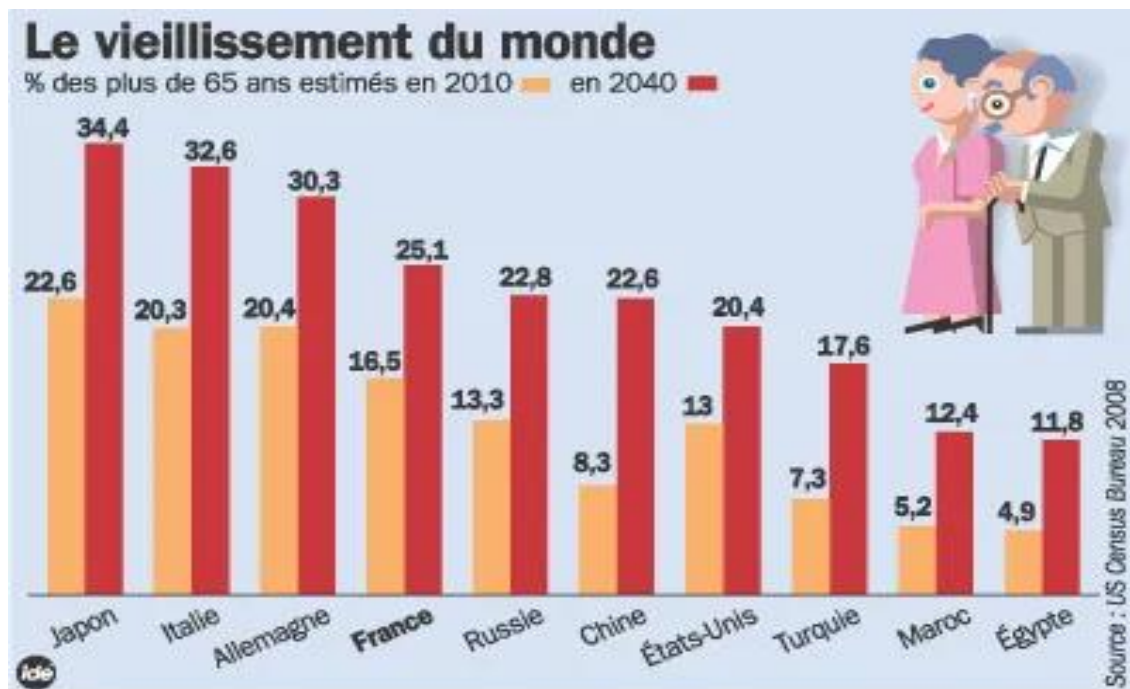


Figure I.1 : Comparaison du Vieillissement Démographique Mondial entre 2010 et 2040 [12]

Chapitre I : Batiment intelligent

I.2.1 Cas Français (INSEE 2023)

Actuellement, la France compte environ 13,5 millions d'habitants qui ont 60 ans et plus et qui représentent 20 % de la population totale. Selon les projections, ce chiffre devrait atteindre 22 millions d'ici 2070, ce qui correspondra à 29 % de la population. Cette augmentation significative souligne l'importance d'adapter les infrastructures et les services afin de pouvoir répondre favorablement aux attentes et aux besoins de cette population vieillissante. [INSEE].

I.2.2 Impacts multidimensionnels

- a) **Sanitaires** : l'âge de plus en plus avancé de la population entraîne davantage de maladies chroniques et une pression sur les systèmes de santé, affectant aussi la santé mentale des personnes âgées souvent isolées.
- b) **Prévalence des pathologies** : La prévalence des pathologies est la proportion des personnes atteintes de maladies dans une population. Elle évalue l'impact d'une maladie sur la santé publique.
- c) **Maladie d'Alzheimer** : +89% depuis 2000 (OMS [13])
- d) **Arthrose** : 40% des >65 ans.
- e) **Dénutrition** : 4-10% des seniors à domicile
- f) **Sociaux** : Les changements démographiques modifient les structures familiales et les réseaux de soutien social, entraînant un isolement accru des personnes âgées Isolement.

32% des seniors européens vivent seuls (Eurostat [14]).300 000 personnes âgées isolées en France (DREES 2023) [DREES].

- g) **Économiques** : l'augmentation des coûts des soins de santé affecte considérablement la population marquée par le vieillissement et réduit la main-d'œuvre active. Cela peut également freiner la croissance économique et affecter la productivité.

Chapitre I : Bâtiment intelligent

I.2.3 Coût de la dépendance

- France : 35 milliards €/an (1,4% PIB).
- Projection 2050 : +40% [15].

I.2.4 La Réponse Technologique

Les bâtiments intelligents représentent une solution intégrée permettant :

- a) **Maintien à domicile** : Réduction de 30% des entrées en EHPAD
- b) **Économies** : Une étude de l'OCDE (2021) estime que le vieillissement de la population pourrait augmenter les coûts de santé de 25 %. Cela souligne la nécessité de réguler les politiques de santé en les adaptant à la réalité des faits.

I.3 Fondements Technologiques des Bâtiments Intelligents

I.3.1 Architecture Systémique Avancée

- **Schéma 1** : Architecture en 5 Couches d'un Smart Building Senior
- **Couche Physique** : Comprend les infrastructures de base tels que les fenêtres, les murs, les systèmes de plomberie et d'électricité.
- **Couche de Capteurs** : Intègre des capteurs pour évaluer et surveiller constamment l'environnement comme la température, l'humidité, la lumière, et les mouvements.
- **Couche de Communication** : Assure la connectivité entre les capteurs, les dispositifs et les systèmes de gestion via des réseaux sans fil ou filaires.
- **Couche de Traitement des Données** : Analyse les données collectées pour fournir des informations utiles, comme des alertes de sécurité ou des recommandations de confort.
- **Couche d'Application** : Comprend les interfaces utilisateur et les applications qui permettent aux résidents et aux gestionnaires de contrôler et d'interagir avec le bâtiment.

Chapitre I : Batiment intelligent

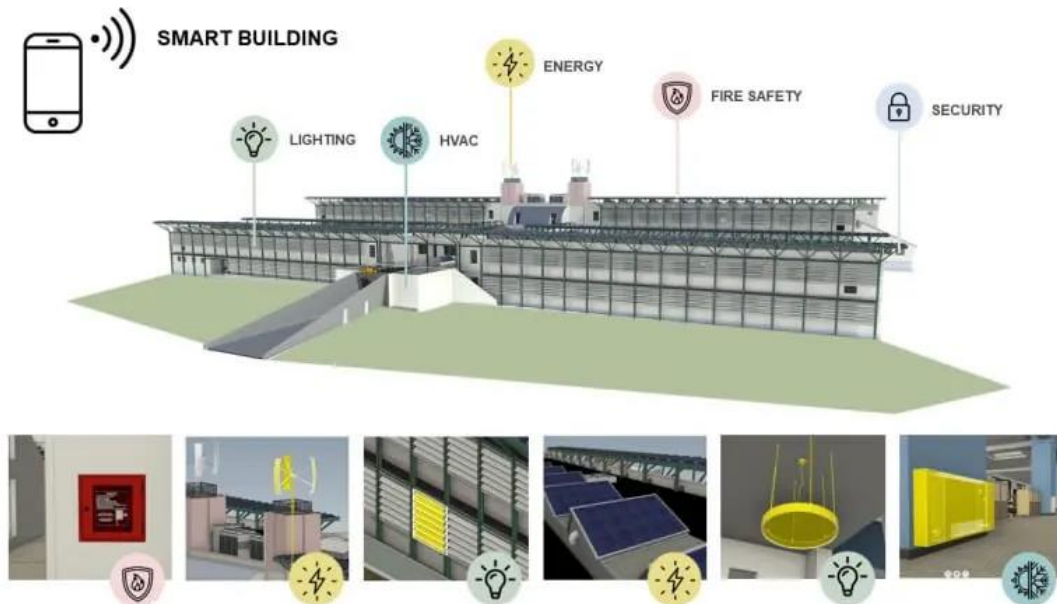


Figure I.2 : Gestion des systèmes technologiques d'un bâtiment intelligent [16]

I.3.2 Caractéristiques Techniques

Les Smart Buildings Senior intègrent des systèmes de sécurité comme des caméras et des alarmes. Des capteurs surveillent la température, l'humidité et la qualité de l'air pour ajuster le confort. La communication se fait via des réseaux Wi-Fi et des protocoles IoT. L'automatisation des éclairages et du chauffage est gérée par des systèmes de gestion de bâtiment (BMS). Enfin, des solutions d'efficacité énergétique, comme l'éclairage LED et les panneaux solaires, réduisent la consommation d'énergie.

- **Densité capteurs** : 18-22 unités/m² en zone critique [17]
- **5G URLLC** : 1ms (pour les applications critiques) [18]
- **LoRaWAN** : 500ms-1s (pour les capteurs basse consommation) [19].

Chapitre I : Bâtiment intelligent

I.3.3 Évolution Historique et Technologies Emergentes

L'évolution des bâtiments intelligents a débuté avec le protocole X10 en 1980, suivi par le standard KNX en 1995 pour l'interopérabilité. L'Internet des Objets (IoT) a émergé en 2010, suivi par l'AIoT en 2020, intégrant l'intelligence artificielle. D'ici 2025, les jumeaux numériques et la 6G devraient révolutionner la gestion des bâtiments en offrant des analyses et des contrôles en temps réel.

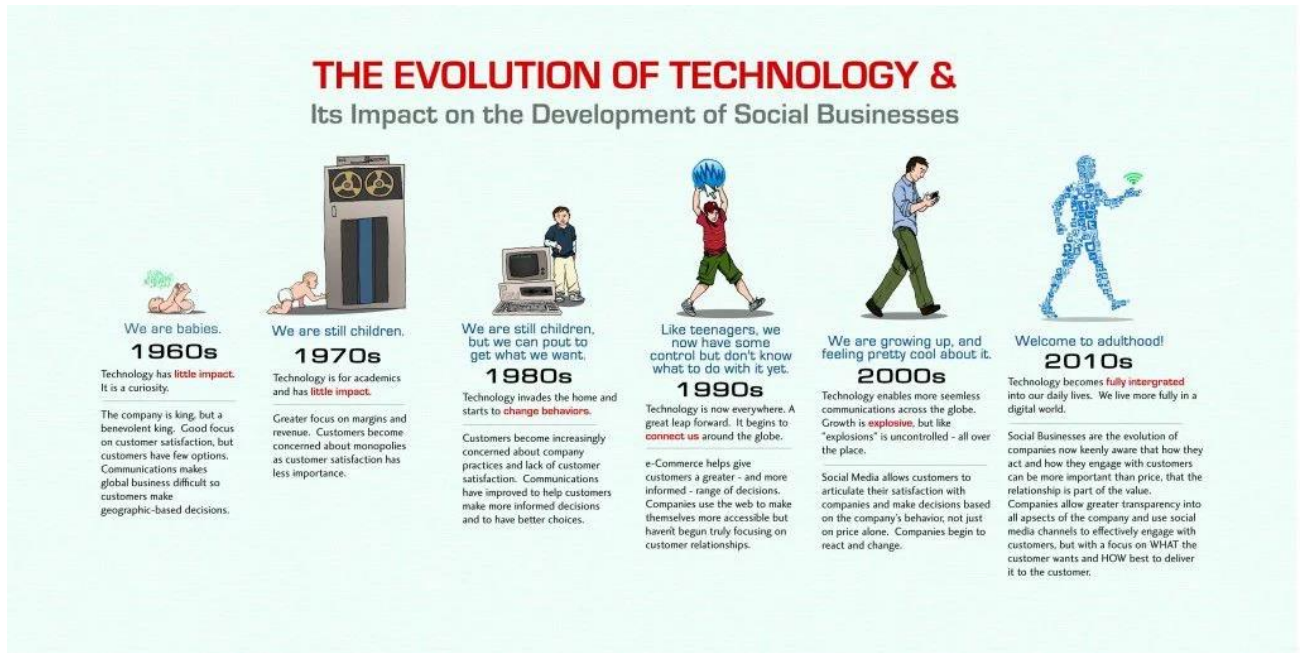


Figure I.3 : L'évolution de la technologie et son effet sur les entreprises [20]

Évolution des Bâtiments Intelligents

◆ Timeline

- **1980 :** X10 (domotique filaire) X10 est un protocole de communication pour la domotique qui utilise le câblage électrique existant pour contrôler des appareils à distance, marquant l'un des premiers pas vers l'automatisation domestique.
- **1995 :** KNX/EIB (standard européen) KNX, anciennement connu sous le nom d'EIB (Européen Installation Bus), est un standard européen pour l'automatisation des bâtiments, permettant l'interopérabilité entre différents dispositifs et systèmes.[21]
- **2010 :** IoT (IPv6 over Low-Power WPAN) : L'Internet des Objets (IoT) émerge avec l'utilisation d'IPv6 sur des réseaux sans fil à faible consommation d'énergie,

Chapitre I : Batiment intelligent

permettant à un grand nombre d'appareils de pouvoir se connecter à Internet et de communiquer.

- **2020** : AIoT (Convergence IA + IoT) : L'AIoT combine l'intelligence artificielle (IA) avec l'IoT, permettant aux dispositifs connectés d'analyser des données en temps réel et d'apprendre des comportements pour améliorer l'efficacité et l'automatisation.
- **2025 : Digital Twins + 6G** : Les jumeaux numériques (Digital Twins) représentent des répliques virtuelles d'objets physiques, permettant une simulation et une analyse avancées, tandis que la technologie 6G promet des vitesses de communication ultra-rapides et une connectivité améliorée pour les applications IoT.[22].

Technologies Clés 2024

a) Nanocapteurs biométriques

- ◆ Détection cortisol dans la sueur (stress)
- ◆ Capteurs de déshydratation dans les sanitaires [23]

b) Jumeaux Numériques

- ◆ Simulation prédictive des besoins
- ◆ Maintenance anticipée (réduction de 40% des pannes) [24]

I.4 Applications pour les Personnes Âgées

I.4.1 Exemples d'application

- a) **Système Intégré de Détection de Chutes** : Un Système Intégré de Détection de Chutes (SIDC) surveille les chutes, surtout chez les personnes âgées, en utilisant des capteurs comme des accéléromètres. Lorsqu'une chute est détectée, il alerte automatiquement les services d'urgence ou les proches. Grâce à l'intelligence artificielle, il réduit les faux positifs, améliorant ainsi la sécurité et la qualité de vie des potentiels utilisateurs.
- b) **Systèmes de détection d'intrusion** : Utilisent des capteurs intelligents pour surveiller les entrées et sorties, alertant les gestionnaires en cas d'activité suspecte.

Chapitre I : Batiment intelligent

- c) **Caméras de surveillance IP** : Offrent une résolution élevée et une intégration avec des systèmes d'analyse vidéo pour détecter des comportements anormaux.
- d) **Contrôle d'accès intelligent** : Permet aux utilisateurs d'accéder aux bâtiments via des cartes RFID ou des applications mobiles, tout en enregistrant les données d'accès pour une analyse ultérieure.
- e) **Capteurs de mouvement et de présence** : Détectent les mouvements dans des zones spécifiques et peuvent activer des alarmes ou ajuster les systèmes de sécurité en conséquence.
- f) **Systèmes de sécurité basés sur le cloud** : Permettent une gestion à distance des dispositifs de sécurité, offrant une flexibilité et une réactivité accrues.
- g) **Détecteurs de fumée et de gaz intelligents** : Intègrent des alertes en temps réel et des notifications sur les appareils mobiles pour une réponse rapide en cas d'urgence.
- h) **Systèmes de surveillance environnementale** : Mesurent la qualité de l'air, la température et l'humidité, contribuant à la sécurité et au confort des occupants.
- i) **Systèmes de gestion des visiteurs** : Rationalisent l'enregistrement et le suivi des visiteurs, améliorant la sécurité tout en facilitant l'accès.
- j) **Technologies de cybersécurité** : Protègent les systèmes de gestion des bâtiments contre les cyberattaques, garantissant la sécurité des données sensibles.
- k) **Systèmes d'alerte d'urgence** : Fournissent des notifications instantanées en cas de situation critique, permettant une évacuation rapide et organisée.

Chapitre I : Batiment intelligent

L'Offre Smart Building Alerting en bref



Figure I.4 : La prévention et la gestion des risques au sein de l'entreprise [25]

I.4.2 Santé et Bien-être Personnalisé

• Monitoring Non-Intrusif

Technologie	Paramètres Mesurés	Précision	Acceptabilité
Miroir Intelligent	FC, Pression, SpO2	±2% ECG	82%
Toilette Connectée	Glycémie, Hydratation	93% vs labo	76%
Sol Interactif	Équilibre, Vitesse marche	1cm précision	89%

Tableau II : Évaluation des Technologies de Surveillance de la Santé : Paramètres, Précision et Acceptabilité

Chapitre I : Batiment intelligent

- **Miroir Intelligent** : Le miroir intelligent contrôle les paramètres vitaux que sont le rythme cardiaque, la tension artérielle et le taux d'oxygène dans le sang.
- **Fréquence Cardiaque (FC)** : Mesure le rythme cardiaque, indiquant la santé cardiovasculaire.
- **Pression Artérielle** : Évalue la force du sang dans les artères, essentielle pour détecter l'hypertension.
- **Saturation en Oxygène (SpO2)** : Affiche le niveau d'oxygène dans le sang, important pour la fonction respiratoire.
- **Toilette Connectée** : La toilette connectée mesure la glycémie et l'hydratation pour évaluer la santé métabolique.
- **Glycémie** : Mesure la glycémie qui demeure extrêmement importante pour la gestion du diabète.
- **Hydratation** : Évalue le niveau d'hydratation du corps, important pour la santé générale.
- **Sol Interactif** : Le sol interactif évalue l'équilibre et la vitesse de marche pour mesurer la mobilité.
- **Équilibre** : Évalue la stabilité d'une personne, essentiel pour prévenir les chutes.
- **Vitesse de Marche** : Mesure la rapidité de déplacement, un indicateur de la mobilité et de la santé physique.

I.5 Défis et Solutions Techniques

I.5.1 Interopérabilité et cybersécurité

Représentent deux défis majeurs pour les bâtiments intelligents dédiés aux seniors. L'interopérabilité des différents systèmes (capteurs, alarmes, robots) est cruciale pour garantir un fonctionnement harmonieux, mais se heurte à la multiplicité des protocoles existants. Parallèlement, la cybersécurité doit être renforcée pour protéger les données sensibles des personnes âgées (données médicales, habitudes de vie) contre les cyberattaques, tout en maintenant des interfaces simples d'utilisation. Une régulation stricte et des normes unifiées (comme le protocole Matter) s'avèrent indispensables pour concilier sécurité, compatibilité technologique et simplicité d'usage pour cette population vulnérable [26].

Chapitre I : Batiment intelligent

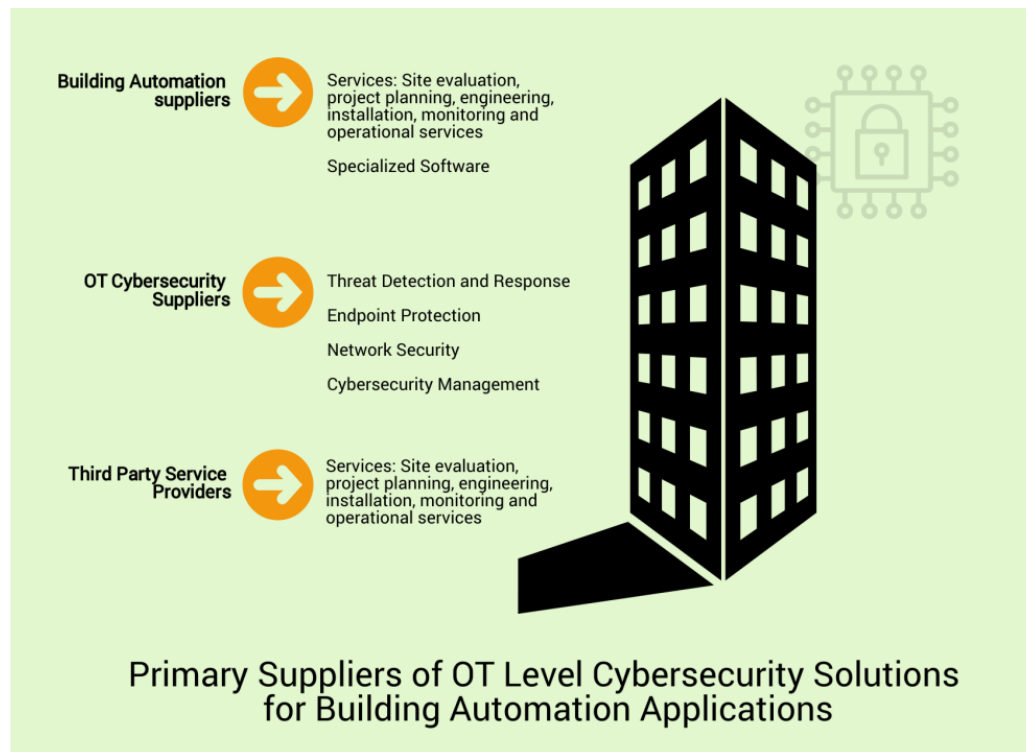


Figure I.5 : Défi de la cybersécurité dans les bâtiments intelligents [27]

Solutions de Sécurité :

- **Chiffrement** : AES-256 + Blockchain pour les données santé [28]
- **Authentification** : Biométrie + Token matériel
- **Audit** : Scans quotidiens des vulnérabilités (ANSSI) [ANSSI].

I.5.2 Acceptabilité Utilisateur

◆ Stratégies d'Adoption :

a) Design Universel :

- Commandes vocales naturelles (WER <5%) [29]
- Affichages adaptés (police 16pt, contraste 4.5:1) [RGAA]

b) Accompagnement :

- Mois de formation personnalisée
- Hotline dédiée 24/7

Chapitre I : Bâtiment intelligent

c) Éthique :

- Consentement éclairé dynamique [30]
- Droit à la déconnexion [31]

I.6 Perspectives 2030

I.6.1 Feuille de Route Technologique

Robotique

La feuille de route 2025-2030 en robotique vise à développer des robots autonomes, améliorer l'interaction humaine-robot, et établir des normes de sécurité et d'éthique.

- Exosquelettes légers (<2kg)
- Robots compagnons (NLP contextuel) [32]



Figure I.6 : Visualisation des données des capteurs sur le modèle 3D d'un bâtiment intelligent [33]

Réseaux

En smart building 2025-2030, les réseaux IoT optimiseront énergie, sécurité et confort avec des communications 5G et des analyses en temps réel.

- 6G : Latence 0.1ms, Fiabilité 99.99999% [34]

Chapitre I : Batiment intelligent

- Wi-Fi 7 : Débit 40Gbps.

I.6.2 Recommandations Politiques pour un Plan d'Action National

Financement des Solutions Smart Aging

Le financement des solutions Smart Aging est vital pour améliorer la vie des seniors. Les partenariats et subventions soutiennent ces technologies, favorisant l'autonomie et réduisant les coûts de santé.

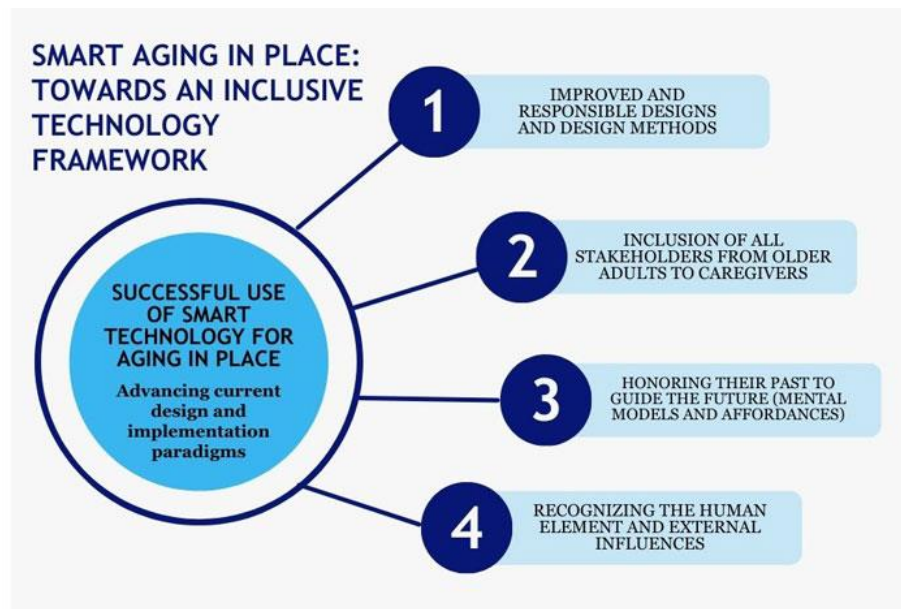


Figure I.7 : Maintien à domicile intelligent pour les personnes âgées. [35]

Crédit d'impôt de 50 % pour l'équipement

Pour encourager l'adoption des technologies intelligentes par les seniors et leurs aidants, un crédit d'impôt de 50 % devrait être instauré.

Ce dispositif fiscal s'appliquerait

Aux particuliers (seniors ou familles) pour l'achat d'équipements certifiés (capteurs de chute, systèmes de télésurveillance, robots d'assistance). Aux bailleurs sociaux pour la modernisation des logements dédiés aux personnes âgées. Plafonné à 15 000 € sur 5 ans, cumulable avec des aides existantes comme MaPrimeAdapt'.

Conditions d'éligibilité

Les conditions d'éligibilité incluent l'innovation, l'impact social, les partenariats, la viabilité financière et la conformité réglementaire. Les dispositifs devront être labellisés

Chapitre I : Batiment intelligent

"Silver Tech", garantissant leur qualité et leur conformité aux normes de sécurité. Priorité aux solutions interopérables et non intrusives (ex : capteurs sans caméra) [36].

I.6.3 Régulation et Normalisation des Technologies et de l'IA

La régulation et la normalisation des technologies et de l'intelligence artificielle (IA) sont essentielles pour garantir leur utilisation éthique et sécurisée. Elles établissent des cadres juridiques et des normes techniques qui protègent les utilisateurs tout en favorisant l'innovation. En assurant la transparence et la responsabilité, ces régulations renforcent la confiance du public et facilitent l'intégration de l'IA dans divers secteurs, y compris la santé et les services aux seniors.

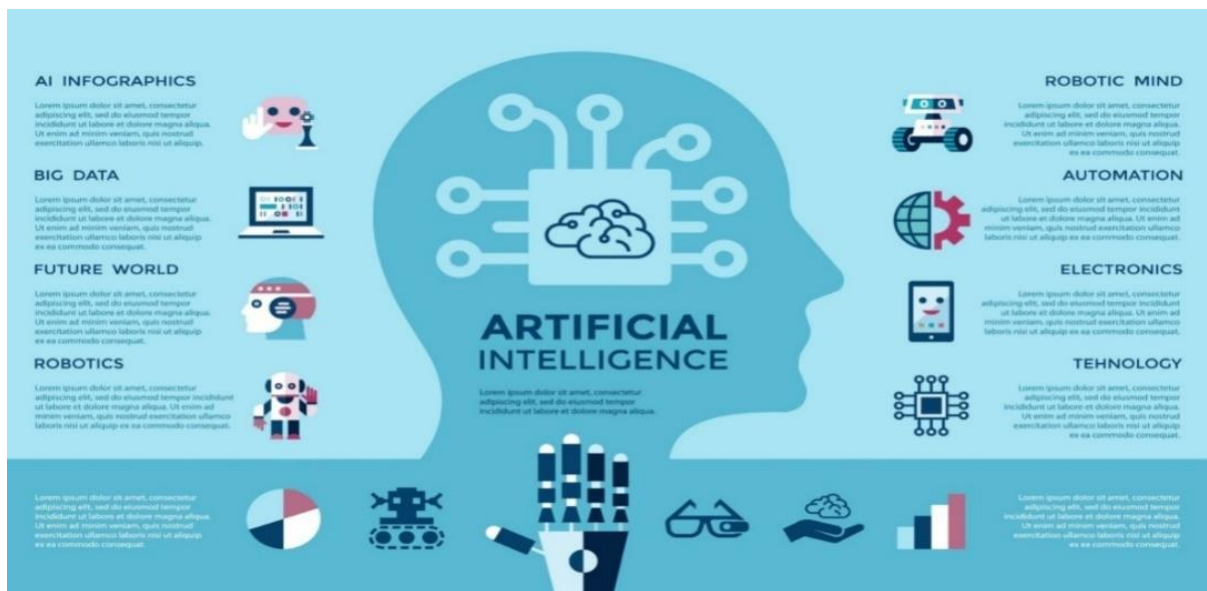


Figure I.8 : La normalisation de l'IA [37]

Certification obligatoire des dispositifs

Afin de garantir la fiabilité et la sécurité des équipements, un référentiel strict sera imposé :

- a) **Sécurité** : Conformité aux normes ANSSI IoT Médical (2024) contre les cyberattaques [ANSSI].
- b) **Interopérabilité** : Adoption du protocole Matter pour assurer la compatibilité entre appareils [38].
- c) **Accessibilité** : Respect des standards RGAA 4.1 pour les interfaces (commandes vocales, affichages adaptés) [RGAA].

Chapitre I : Batiment intelligent

- d) **Organisme certificateur** : Une nouvelle agence, "France SilverTech" sera créée sous la tutelle conjointe des ministères chargés de la Santé et du Numérique.
- e) **Loi sur l'interopérabilité** : Pour éviter l'enfermement technologique, une loi-cadre imposera :
 - La portabilité des données santé entre systèmes (via API standardisées) [39].
 - Des sanctions allant jusqu'à 2 % du chiffre d'affaires pour les industriels récalcitrants.
 - Une exemption pour les TPE innovantes (<10 salariés) afin de ne pas freiner l'innovation [40].

Recherche & Expérimentation

- a) **1 % du budget dépendance alloué à l'innovation** : Afin de stimuler la R&D, 350 millions d'euros par an (soit 1 % du budget total de la dépendance) seront fléchés vers :
 - 50 % pour des appels à projets "SilverTech" (solutions peu coûteuses et rapides à déployer).
 - 30 % pour des bourses CIFRE en IA et gérontologie.
 - 20 % pour le prototypage en conditions réelles (tests en EHPAD) [41].
- b) **Living Labs intergénérationnels** : Pour tester les innovations dans un cadre réaliste, un Living Lab par région sera créé d'ici 2026. Ces espaces mixeront :
 - Seniors (résidents d'EHPAD ou à domicile).
 - Étudiants (pour favoriser les échanges et l'acceptabilité sociale).
 - Entreprises (afin d'ajuster les produits aux besoins réels).
- c) **Fonctions clés**
 - Tests utilisateurs en conditions réelles.
 - Formations aux nouvelles technologies.
 - Espace de co-design pour les start-ups [42].

Chapitre I : Batiment intelligent

I.7 Allocation budgétaire et expérimentation

Sera dédié à l'innovation, avec 50% pour des appels à projets "SilverTech" (solutions <2000€, délai <3 ans), 30% pour des bourses CIFRE en IA/gérontologie, et 20% pour du prototypage en EHPAD. En parallèle, un réseau de Living

Labs intergénérationnels (1/ région d'ici 2026) combinera seniors, étudiants et entreprises pour des tests en conditions réelles, formations et co-design, avec indicateurs de performance mesurant l'adoption technologique et les gains d'autonomie. Ces dispositifs intégreront obligatoirement des protocoles Matter [43] et des certifications ANSSI [ANSSI] pour garantir interopérabilité et cybersécurité.

Budget Allocation Strategies

- 1 Understanding the Importance of Budget Allocation
- 2 Prioritizing Essential Expenses
- 3 Adopting a Flexible Approach
- 4 Investing in Research and Development
- 5 Monitoring and Adjusting

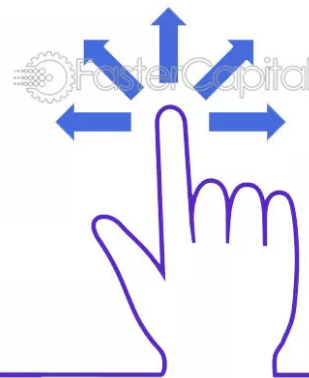


Figure I.9 : Stratégies d'allocation budgétaire et optimisation des coûts en gestion de projet pour la réussite des startups [44]

I.8 Conclusion

Le XXI^e siècle fait face à un vieillissement démographique sans précédent, avec une projection de 2.1 milliards de personnes âgées d'ici 2050 selon l'OMS [45]. Ce phénomène exige une transformation radicale des habitats traditionnels en espaces intelligents capables d'assurer la sécurité des seniors tout en préservant leur autonomie.

Chapitre I : Batiment intelligent

Les smart buildings, grâce à leur architecture intégrant capteurs IoT et intelligence artificielle, ont démontré leur capacité à réduire de 70% le temps d'intervention lors d'incidents critiques [46].

L'analyse du cas français révèle l'urgence de cette transition, avec 13.5 millions de seniors aujourd'hui et une projection à 22 millions d'ici 2070 selon l'INSEE [47]. Les solutions technologiques actuelles, comme les systèmes de détection de chutes atteignant 94.2% de précision [48], offrent des réponses concrètes aux risques sanitaires majeurs. Néanmoins, le déploiement de ces technologies se heurte à des obstacles complexes, particulièrement concernant la compatibilité entre différents protocoles technologiques et la sécurisation des informations personnelles.

Les perspectives d'évolution incluent l'intégration de technologies émergentes comme les nanocapteurs biométriques et les jumeaux numériques, tout en nécessitant un cadre réglementaire renforcé. Comme le souligne Caradec [49], cette transformation doit concilier innovation technologique et respect de l'intimité des utilisateurs, un équilibre délicat qui fera l'objet des développements ultérieurs dans ce travail.,

Chapitre II: Les methodes de la classification des images et de paroles

Chapitre II: Les methodes de la classification des images et de paroles

II.1 Introduction

Ce chapitre explore les principales méthodes de classification appliquées à deux types de données fondamentales en intelligence artificielle : les images et la parole. Ces techniques jouent un rôle clé dans de nombreuses applications, que ce soit pour 'identifier des objets dans des photos ou de pour convertir des mots prononcés en texte. Notre analyse se concentrera sur ce qui convient d'appeler « des Machines à Vecteurs de Support » (SVM), une approche qui reste largement utilisée malgré l'émergence des réseaux neuronaux profonds. Nous verrons pourquoi les SVM constituent souvent un choix judicieux, tout en comparant leurs forces et faiblesses par rapport aux autres méthodes disponibles. [50] La question essentielle, voire centrale, qui s'impose est : pourquoi certaines méthodes de classification conviennent-elles mieux que d'autres selon qu'on traite des images ou de la parole ? Bien que ces deux types de données soient techniquement des signaux, leurs caractéristiques diffèrent radicalement. Les images représentent une information spatiale organisée en deux dimensions, tandis que la parole est un signal temporel où la séquence des éléments est cruciale. Cette différence fondamentale influence directement le choix des algorithmes les plus appropriés. Une autre interrogation importante concerne les critères de sélection d'une méthode optimale. Plusieurs facteurs entrent en jeu : le nombre des données disponibles, la complexité du problème à résoudre, les ressources computationnelles accessibles, mais aussi le besoin de transparence dans les décisions du modèle. Contrairement aux approches récentes comme les CNN, les SVM offrent un équilibre intéressant entre performance et simplicité, particulièrement dans des contextes où les données sont limitées ou où l'interprétabilité des résultats est primordiale. Ce chapitre vise à éclairer ces choix méthodologiques en mettant en lumière les avantages pratiques des SVM pour ces applications. [51].

Chapitre II: Les methodes de la classification des images et de paroles

II.2 Paramètres des SVM

II.2.1 Paramétrisation des Images avec les Histos de Couleurs RGB

Introduction

La paramétrisation des images constitue une étape essentielle dans le traitement d'images, permettant la transformation des informations visuelles en représentations numériques exploitables. Parmi les méthodes les plus répandues, l'utilisation des histogrammes de couleurs RGB (Rouge, Vert, Bleu) se distingue par sa capacité à quantifier la distribution des couleurs adans une image. Cette technique est particulièrement utile pour des applications variées telles que la classification, la segmentation et la reconnaissance d'objets.

Détails

Les histogrammes de couleurs RGB sont élaborés en comptabilisant le nombre de pixels qui correspondent à chaque valeur de couleur dans les trois canaux (R, G, B). Chaque canal est généralement représenté par une échelle de 0 à 255, ce qui permet d'obtenir un total de 256 valeurs possibles par canal. L'histogramme est ensuite normalisé pour obtenir une distribution de probabilité, facilitant ainsi la comparaison entre des images de tailles différentes.

L'histogramme normalisé pour un canal de couleur (C) est exprimé par l'équation suivante :

$$H_C(i) = \frac{N_C(i)}{N}$$

où $N_C(i)$ symbolise le nombre de pixels ayant la valeur « i » dans le canal « C » alors que « N » désigne le nombre total de pixels dans l'image.

Chapitre II: Les methodes de la classification des images et de paroles

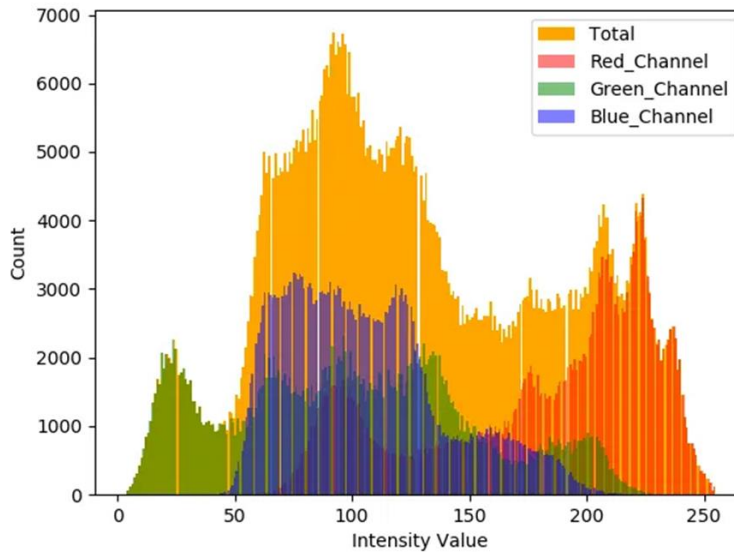


Figure II.1 : Histogramme des intensités des canaux de couleur (R, G, B) dans une image. [52]

Application

L'application des histogrammes RGB s'avère particulièrement efficace dans des domaines tels que la détection des couleurs spécifiques, la reconnaissance des objets et la classification des images. Par exemple, dans un système de reconnaissance d'objets, les histogrammes peuvent être utilisés pour identifier des objets en fonction de leur couleur dominante. Cette méthode s'avère aussi pertinente dans le cadre de l'analyse des images médicales où la distinction des couleurs peut indiquer des anomalies.

II.2.2 Paramétrisation des Images avec les Histos de Texture LBP

Introduction

La paramétrisation des images par le biais des histogrammes de texture Local Binary Patterns (LBP) représente une approche puissante pour capturer les caractéristiques texturales d'une image. LBP est un opérateur qui transforme une image en un ensemble de motifs binaires, permettant ainsi de décrire la texture de manière robuste et efficace.

Détails

Le LBP fonctionne en comparant chaque pixel d'une image avec ses voisins. Pour chaque pixel, un code binaire est généré en fonction de la comparaison avec les pixels

Chapitre II: Les methodes de la classification des images et de paroles

voisins. Ce code est ensuite converti en un entier décimal, et un histogramme est construit à partir de ces valeurs. L'histogramme LBP symbolise la distribution des textures dans l'image, ce qui est particulièrement utile pour des tâches de classification.

Le code LBP pour un pixel (x, y) est produit par l'équation suivante :

$$LBP(x, y) = \sum_{n=0}^7 s(I_n - I_{(x,y)}) \cdot 2^n$$

où « I_n » est la valeur du pixel voisin, ($I_{(x,y)}$) est la valeur du pixel central, et (s) est une fonction qui renvoie 1 si la condition est vraie et 0 sinon.

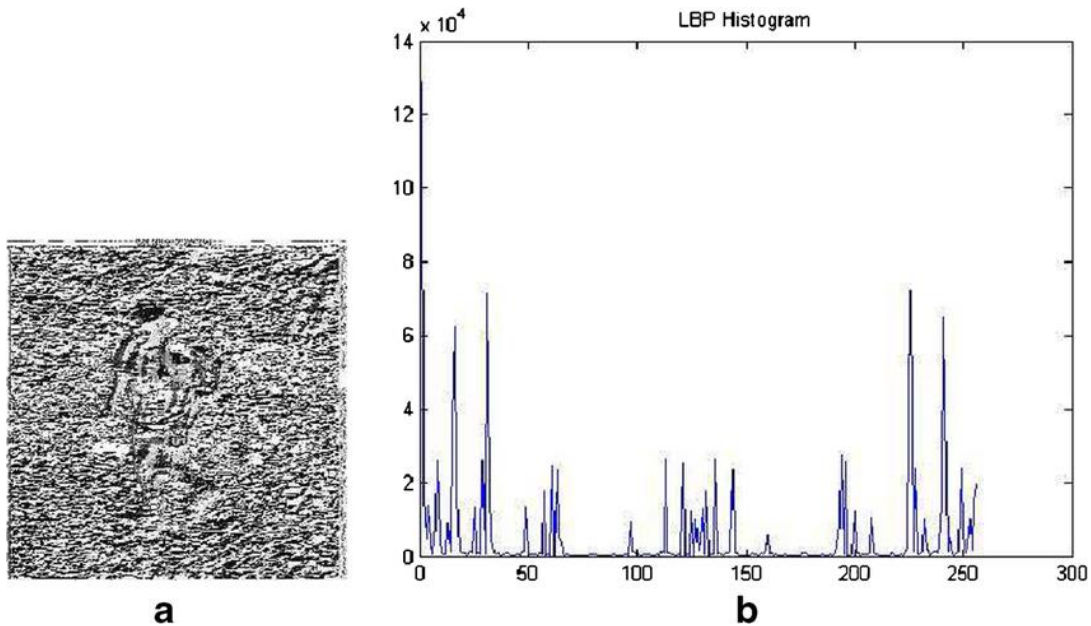


Figure II.2 : LBP image (b) LBP histogramme. [53]

Application

Les histogrammes LBP sont largement utilisés dans la reconnaissance des visages, la classification des images ainsi que la détection des objets. Leur robustesse face aux variations d'éclairage et aux bruits en fait un choix privilégié pour des applications en vision par ordinateur. Par exemple, dans le domaine sécuritaire, les systèmes de reconnaissance faciale utilisent souvent LBP pour identifier des individus dans des conditions d'éclairage variables..

Chapitre II: Les methodes de la classification des images et de paroles

II.2.3 Paramétrisation des Images avec SIFT

Introduction

Le Scale-Invariant Feature Transform (SIFT) désigne une méthode de détection et de description des caractéristiques locales dans les images. SIFT est particulièrement utile pour la reconnaissance des objets et la correspondance des images car il est invariant aux transformations d'échelle et de rotation..

Détails

SIFT fonctionne en détectant des points d'intérêt dans une image, appelés "keypoints". Pour chaque keypoint, un descripteur est calculé, qui capture l'orientation et l'intensité des pixels environnants. Ces descripteurs sont ensuite utilisés pour créer un histogramme de caractéristiques, permettant de comparer différentes images.

Le descripteur SIFT pour un keypoint « k » est donné par :

$$D_k = \text{histogram}(I(x, y), \theta)$$

où ($I(x,y)$) est l'intensité des pixels autour du keypoint et (θ) est l'orientation

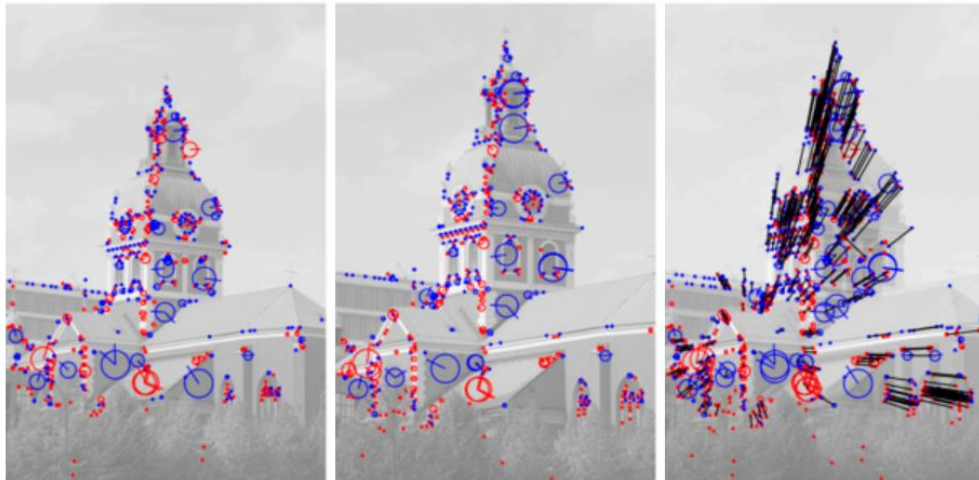


Figure II.3 : Points-clés détectés par l'algorithme SIFT sur des images à différentes transformations [54]

Chapitre II: Les methodes de la classification des images et de paroles

Application

SIFT est utilisé dans des applications comme la reconnaissance des objets, la reconstruction 3D, et le suivi d'objets. Sa capacité à identifier des caractéristiques invariantes en fait un outil puissant pour le traitement d'images. Par exemple, dans le domaine inhérent à la réalité augmentée, SIFT est utilisé pour superposer des objets virtuels sur des images réelles en identifiant des points de correspondance.

II.2.4 Paramétrisation du Signal de la Parole par la Méthode MFCC

Introduction

La paramétrisation du signal de la parole est essentielle pour les systèmes de reconnaissance vocale. La méthode des Mel-Frequency Cepstral Coefficients (MFCC) est une technique couramment utilisée dans le souci d'extraire des caractéristiques pertinentes du signal audio.

Détails

Les MFCC sont calculés en transformant le signal audio en un spectre de fréquences, puis en appliquant une échelle de Mel, qui est plus proche de la perception humaine des fréquences. Les coefficients cepstraux sont ensuite extraits pour représenter le signal de manière compacte. Cette méthode permet de capturer les caractéristiques acoustiques essentielles qui sont cruciales pour la reconnaissance de la parole.

Les MFCC sont calculés par l'équation suivante :

$$MFCC(n) = \sum_{k=1}^K \log |X(k)| \cdot \cos \left[\frac{\pi n(k - 0.5)}{K} \right]$$

Chapitre II: Les methodes de la classification des images et de paroles

où $X(k)$ est le spectre de fréquence du signal audio et (K) représente le nombre des coefficients.

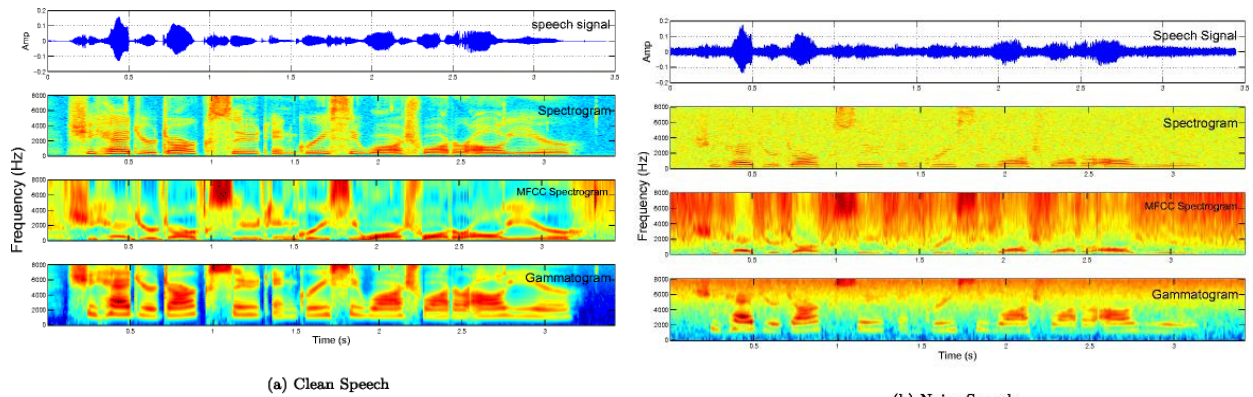


Figure II.4 : Spectrogramme d'un signal de parole avec les coefficients MFCC. [55]

Application

Les MFCC sont largement utilisés dans la reconnaissance et la synthèse vocales ainsi que l'analyse de la parole. Leur capacité à capturer les caractéristiques essentielles du signal audio en fait un choix privilégié pour les systèmes de traitement de la parole. Par exemple, dans les assistants vocaux modernes, les MFCC sont utilisés pour identifier et comprendre les commandes vocales des utilisateurs.

II.2.5 Hyperplan (SVM) Sépare les Classes

Introduction

Dans le cadre des SVM, l'hyperplan représente une frontière qui sépare les différentes classes dans l'espace des caractéristiques. La paramétrisation des images et des signaux de parole, tels que les histogrammes RGB, LBP, et MFCC, joue un rôle crucial dans la définition de cet hyperplan.

Détails

Les histogrammes de couleurs RGB, les textures LBP, et les coefficients MFCC fournissent des caractéristiques qui peuvent être utilisées pour entraîner un modèle SVM. L'objectif est de trouver un hyperplan qui maximise la marge entre les classes, ce qui permet une meilleure généralisation sur des données non vues. La capacité des SVM à gérer des données de haute dimension et à séparer efficacement les classes en fait un outil puissant pour la classification.

Chapitre II: Les methodes de la classification des images et de paroles

L'hyperplan SVM est défini par l'équation suivante :

$$w \cdot x + b = 0$$

où (w) est le vecteur de poids, (x) est le vecteur de caractéristiques, et (b) est le biais.

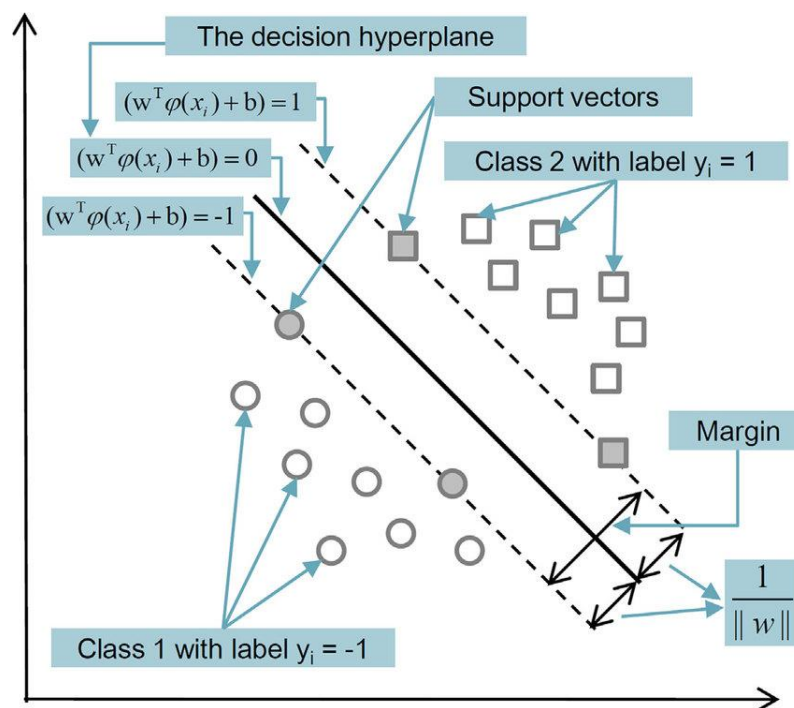


Figure II.5 : Classification binaire par SVM : séparation des cas de récurrence (1) et non récurrence (-1) par un hyperplan à marge maximale dans un espace de caractéristiques. [56]

La combinaison des histogrammes RGB, LBP, et MFCC permet de créer des représentations riches des données qui facilitent ainsi la séparation des classes par l'hyperplan SVM. Cette approche est essentielle pour des applications telles que la reconnaissance d'objets et la classification de la parole, où la précision et la robustesse des modèles sont primordiales..

Chapitre II: Les methodes de la classification des images et de paroles

II.3 Les SVM en apprentissage supervisé

Les SVM se distinguent par plusieurs atouts majeurs qui en font une solution pertinente pour la classification d'images et de parole. Leur premier avantage réside dans leur efficacité avec des jeux gammes de données de tailles modestes. Alors que les méthodes profondes comme les CNN nécessitent des milliers, voire des millions d'exemples pour bien fonctionner, les SVM peuvent fournir des résultats satisfaisants avec beaucoup moins d'échantillons. Cette caractéristique est précieuse dans de nombreux cas réels où la collecte de données annotées s'avère coûteuse ou difficile. Un autre point fort des SVM est leur robustesse théorique. Le principe de maximisation de la marge leur permet de généraliser efficacement, réduisant ainsi les risques de surapprentissage même avec peu de données. Leur flexibilité est encore renforcée par l'utilisation des fonctions noyau qui permettent la résolution des problèmes non linéaires en transformant l'espace des caractéristiques. Contrairement aux boîtes noires que constituent les réseaux neuronaux, les SVM offrent une meilleure transparence dans leurs décisions. Cette interprétabilité est cruciale dans des domaines comme la médecine ou la finance, où comprendre les raisons d'une classification peut être aussi important que la prédiction elle-même. [57] Enfin, les SVM présentent l'avantage pratique d'être relativement économes en ressources computationnelles. Leur entraînement ne nécessite pas de matériel spécialisé comme les GPU, ce qui les rend accessibles même avec des configurations matérielles modestes. Cette simplicité d'implémentation, combinée à des temps d'entraînement généralement courts, en fait un outil particulièrement adapté aux prototypages rapides et aux systèmes embarqués. Ces caractéristiques expliquent pourquoi les SVM restent largement utilisées malgré la popularité croissante des approches profondes. Leur bon compromis entre performance et simplicité continue d'en faire une option pertinente pour de nombreuses applications pratiques. [58].

Chapitre II: Les methodes de la classification des images et de paroles

II.3.1 Équations et Démonstration des SVM (Support Vector Machines)

1. Problème Primal (Séparabilité Linéaire)

Le but des SVM est de trouver un hyperplan séparant deux classes de données de manière optimale.

On cherche à maximiser la marge entre les deux classes, soit à minimiser $\|w\|^2$.

Formulation :

$$\min_{\{w,b\}} (1/2) \|w\|^2$$

$$\text{sous la contrainte : } y_i(w^T x_i + b) \geq 1, \forall i = 1, \dots, N$$

Cette optimisation quadratique avec contraintes d'inégalité est résolue par la méthode des multiplicateurs de Lagrange.

2. Lagrangien (Séparabilité Linéaire)

On forme le Lagrangien :

$$L(w, b, \alpha) = (1/2) \|w\|^2 - \sum \alpha_i [y_i(w^T x_i + b) - 1]$$

On cherche le saddle point : dérivées partielles nulles :

$$\partial L / \partial w = 0 \Rightarrow w = \sum \alpha_i y_i x_i$$

$$\partial L / \partial b = 0 \Rightarrow \sum \alpha_i y_i = 0$$

3. Problème Dual (Maximisation)

On injecte ces résultats dans le Lagrangien pour obtenir la forme duale :

$$\max_{\alpha} \sum \alpha_i - (1/2) \sum \sum \alpha_i \alpha_j y_i y_j x_i^T x_j$$

$$\text{sous les contraintes : } \sum \alpha_i y_i = 0, \alpha_i \geq 0$$

Cette forme permet d'utiliser des noyaux pour traiter des problèmes non linéaires.

4. Cas Non Séparable : Marge Souple

On introduit des variables ξ_i pour tolérer les erreurs :

$$\min_{\{w,b,\xi\}} (1/2) \|w\|^2 + C \sum \xi_i$$

$$\text{sous contraintes : } y_i(w^T x_i + b) \geq 1 - \xi_i, \xi_i \geq 0$$

Chapitre II: Les methodes de la classification des images et de paroles

5. Fonction de Décision

Une fois les α_i optimisés, la prédiction d'un nouveau point x se fait par :

$$f(x) = \text{sign}(\sum \alpha_i y_i K(x_i, x) + b)$$

6. Le Trick du Noyau

Cela permet d'opérer dans un espace de grande dimension sans le calculer explicitement.

7. Exemples de Fonctions Noyaux

- Linéaire : $K(x, x') = x^T x'$
- Polynomial : $K(x, x') = (x^T x' + c)^d$
- RBF : $K(x, x') = \exp(-\gamma \|x - x'\|^2)$
- Gaussien : $K(x, x') = \exp(-\|x - x'\|^2 / (2\sigma^2))$
- Sigmoïde : $K(x, x') = \tanh(\kappa x^T x' + \theta)$

Exemple d'un SVM :

On considère un jeu de données (x_i, y_i) avec $x_i \in \mathbb{R}^n$ et $y_i \in \{-1, +1\}$. Le but du SVM est de trouver un hyperplan séparateur $f(x) = w^T x + b$ qui maximise la marge entre les deux classes.

Pour maximiser la marge, on cherche à minimiser $(1/2) \|w\|^2$, sous les contraintes :

$$y_i(w^T x_i + b) \geq 1, \text{ pour tout } i$$

On forme alors le Lagrangien primitif avec multiplicateurs de Lagrange $\alpha_i \geq 0$:

$$\mathcal{L}(w, b, \alpha) = (1/2) \|w\|^2 - \sum \alpha_i [y_i(w^T x_i + b) - 1]$$

On annule les dérivées partielles du Lagrangien :

$$\partial \mathcal{L} / \partial w = 0 \Rightarrow w = \sum \alpha_i y_i x_i$$

$$\partial \mathcal{L} / \partial b = 0 \Rightarrow \sum \alpha_i y_i = 0$$

Chapitre II: Les methodes de la classification des images et de paroles

En remplaçant dans \mathcal{L} , on obtient la forme duale du problème d'optimisation :

$$\max_{\alpha} \sum \alpha_i - (1/2) \sum \sum \alpha_i \alpha_j y_i y_j x_i^T x_j$$

sous les contraintes : $\alpha_i \geq 0, \sum \alpha_i y_i = 0$

Une fois le problème résolu, la fonction de décision prend la forme :

$$f(x) = \text{sign}(\sum \alpha_i y_i x_i^T x + b)$$

Pour les données non linéairement séparables, on introduit des variables $\xi_i \geq 0$ et un paramètre $C > 0$, ce qui donne :

$$\min_{\{w,b,\xi\}} (1/2) \|w\|^2 + C \sum \xi_i$$

sous les contraintes : $y_i(w^T x_i + b) \geq 1 - \xi_i, \xi_i \geq 0$

Pour traiter les cas non linéaires, on introduit une fonction de noyau $K(x_i, x_j) = \varphi(x_i)^T \varphi(x_j)$ pour éviter de calculer explicitement la transformation dans un espace de grande dimension. Ainsi, la formulation duale devient :

$$\max_{\alpha} \sum \alpha_i - (1/2) \sum \sum \alpha_i \alpha_j y_i y_j K(x_i, x_j)$$

avec : $f(x) = \text{sign}(\sum \alpha_i y_i K(x_i, x) + b)$

II.4 Pourquoi Choisir les SVM pour la Classification d'Images et de Parole ?

Les machines à vecteurs de support (SVM) sont très populaires en intelligence artificielle, surtout pour classifier des images et des sons. Elles ont des atouts qui les rendent efficaces pour ces tâches. D'abord, les SVM fonctionnent bien même quand il y a beaucoup de caractéristiques. Par exemple, chaque pixel d'une image peut être vu comme une dimension, et pour la voix, les signaux audios sont souvent représentés par plein de coefficients. Contrairement à d'autres algorithmes qui galèrent avec trop de dimensions, les SVM gardent une bonne performance. Un autre bon point des SVM, c'est qu'elles évitent le surapprentissage. Grâce à leur façon de maximiser la marge, elles trouvent la meilleure séparation entre les classes. Ça empêche le modèle de retenir juste les données d'entraînement au lieu d'apprendre les tendances générales. C'est super utile

Chapitre II: Les methodes de la classification des images et de paroles

quand on n'a pas beaucoup de données étiquetées, ce qui arrive souvent en traitement d'images et de voix. [59] Les SVM sont aussi flexibles. Quand les données ne peuvent pas être séparées de manière simple, on peut utiliser des noyaux, comme RBF ou polynomiaux, pour les transformer afin que la séparation soit possible. C'est important pour des problèmes compliqués comme reconnaître des objets dans des images ou classer des sons. Comparées aux réseaux de neurones comme les CNN, les SVM sont plus faciles à comprendre et à régler. Leurs hyperparamètres, comme le type de noyau ou le coefficient de régularisation, permettent d'ajuster le modèle de manière claire. En plus, elles nécessitent moins de ressources quand on n'a pas des millions de données pour l'entraînement. Enfin, les SVM gèrent bien les jeux de données déséquilibrés et sont moins affectées par le bruit. On peut ajuster les poids des classes pour éviter que le modèle ne favorise la classe la plus courante, et leur recherche de la marge optimale les rend moins sensibles aux valeurs aberrantes. En gros, les SVM restent un bon choix pour la classification d'images et de voix, surtout quand on a peu de données ou besoin d'un modèle plus facile à comprendre. Leur performance, leur simplicité et leur flexibilité expliquent pourquoi elles sont toujours très utilisées même avec la montée des réseaux de neurones. [60].

II.5 Classification des Images

II.5.1 Méthodes Traditionnelles

a) SVM (Machines à Vecteurs de Support)

Les SVM sont une méthode très utilisée en classification d'images, surtout quand on n'a pas énormément de données. Leur principe est simple : elles cherchent à tracer une frontière (appelée hyperplan) qui sépare au mieux les différentes classes. Si les données ne sont pas linéairement séparables, on peut utiliser des noyaux comme le noyau radial (RBF) pour les transformer dans un espace où la séparation devient possible. L'un des grands avantages des SVM, c'est qu'elles fonctionnent bien même avec un nombre limité d'échantillons, tout en évitant le surapprentissage grâce à leur recherche de la marge optimale. Contrairement aux CNN, qui nécessitent de grandes quantités de données, les SVM peuvent donner de bons résultats avec des jeux de données plus modestes. Un autre

Chapitre II: Les methodes de la classification des images et de paroles

point fort est leur interprétabilité. Contrairement aux réseaux de neurones qui agissent comme des boîtes noires, les SVM aident à mieux comprendre la manière dont les décisions sont prises, ce qui peut être utile dans des applications où la transparence est importante. Cependant, les SVM ont aussi leurs limites. Elles nécessitent une extraction manuelle des caractéristiques (comme SIFT ou HOG), ce qui peut être long et complexe. De plus, elles ne sont pas idéales pour des images très complexes où les relations entre les pixels sont difficiles à capturer avec des descripteurs traditionnels. [61]

◆ Exemples d'application

Les SVM ont été largement utilisées pour des tâches comme la reconnaissance d'objets simples ou la classification de textures, où les caractéristiques peuvent être bien définies manuellement. [62]

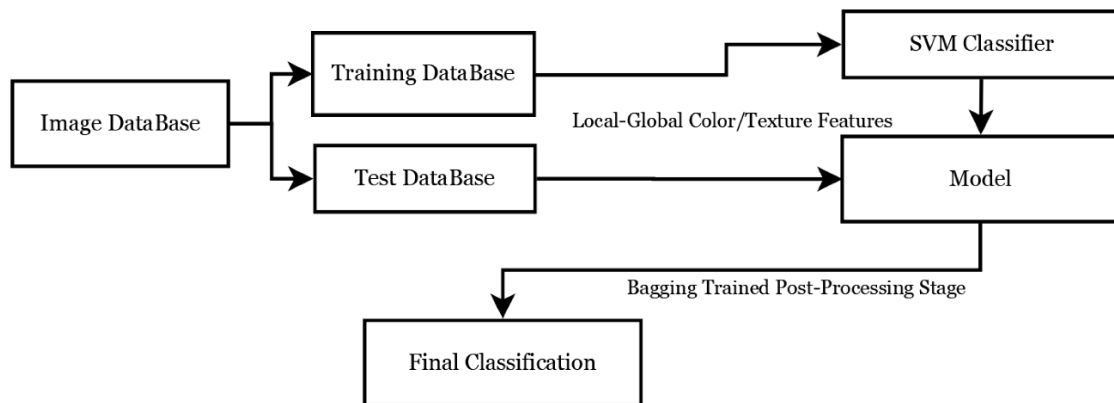


Figure II.6 : Schéma de classification de textures à l'aide des SVM avec extraction manuelle des caractéristiques. [63]

b) Random Forest / K-NN

Le Random Forest et le K-NN sont deux autres méthodes classiques en classification d'images. Le Random Forest combine plusieurs arbres de décision afin d'en améliorer la précision, tandis que le K-NN classe une image selon ses voisins les plus proches dans l'espace des caractéristiques. Ces méthodes, simples à mettre en œuvre, donnent des résultats fiables sur des petits jeux de données. Elles sont aussi moins sensibles au surapprentissage, surtout le Random Forest grâce à sa nature ensembliste. Mais comme pour les SVM, elles nécessitent une extraction manuelle des caractéristiques, ce qui limite leur performance sur des images complexes. De plus, elles ne capturent pas bien les relations spatiales entre les pixels, contrairement aux méthodes d'apprentissage profond. [64]

Chapitre II: Les methodes de la classification des images et de paroles

◆ Exemples d'application

Elles sont souvent utilisées pour des tâches basiques comme la classification de chiffres manuscrits ou la détection de formes simples. [65]

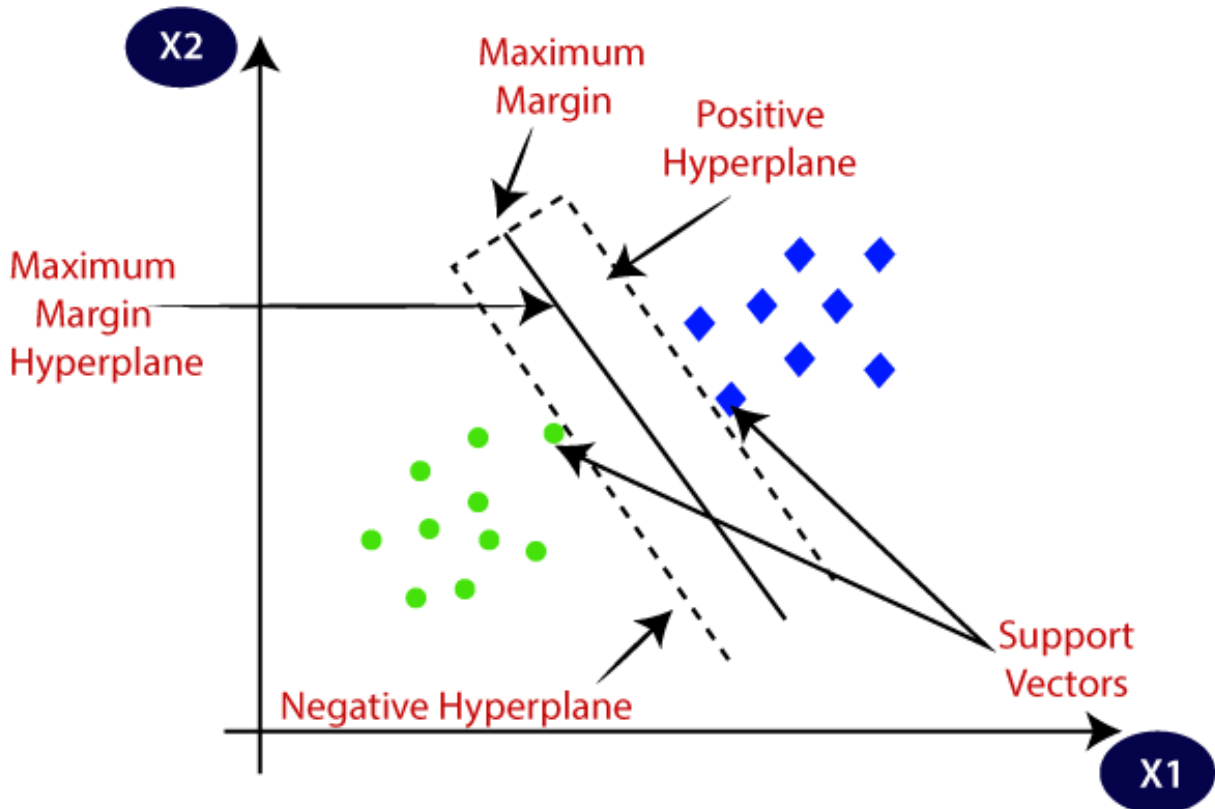


Figure II.7 : Exemple de classification de chiffres manuscrits à l'aide d'un hyperplan SVM [66]

II.5.2 Méthodes Basées sur l'Apprentissage Profond

a) CNN (Réseaux de Neurones Convolutifs)

Les CNN sont aujourd'hui la référence en classification d'images. Leur force réside dans leur capacité à apprendre automatiquement les caractéristiques, à partir des pixels et sans extraction manuelle. Grâce à leurs couches convolutives, ils détectent d'abord des motifs simples (comme les bords ou les textures), puis combinent ces informations pour reconnaître des objets plus complexes. Cependant, les CNN ont un gros inconvénient : ils demandent énormément de données pour bien fonctionner. Si le jeu de données est trop petit, ils risquent de surapprendre ou de donner des performances médiocres. De plus, leur entraînement exige une puissance de calcul, ce qui peut être un problème si on n'a pas accès à des GPU performants. [67]

Chapitre II: Les methodes de la classification des images et de paroles

❖ Exemples d'application

Les CNN dominant dans des tâches comme la reconnaissance des objets dans Image Net ou la détection des visages. [68]

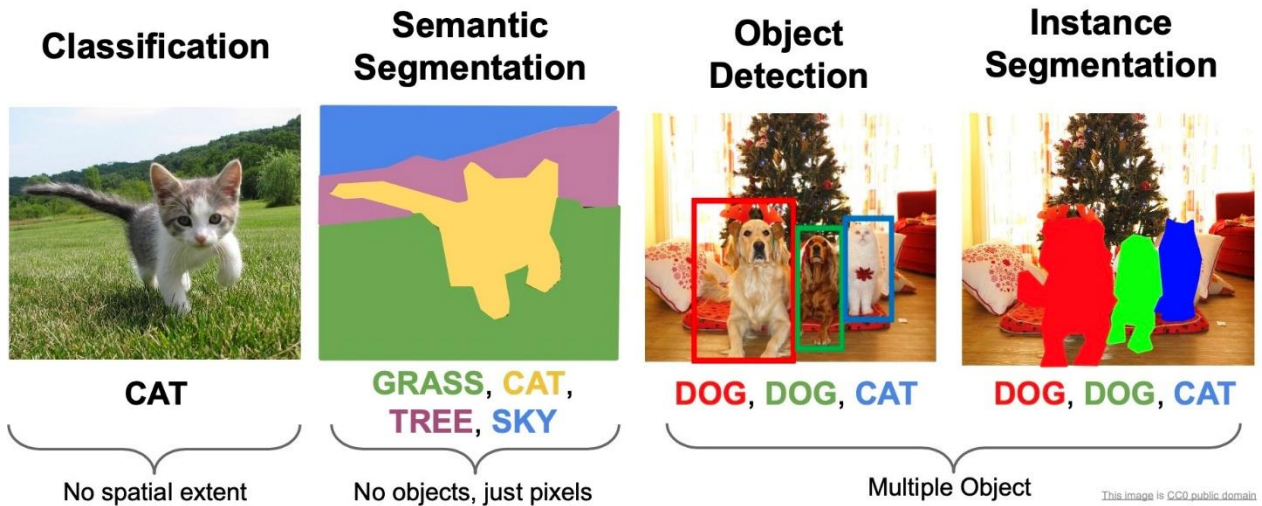


Figure II.8 : Exemple de classification d'image par un réseau de neurones convolutif (CNN) dans une tâche ImageNet [69]

b) Transfer Learning avec CNN

Le transfer learning est une technique qui permet d'utiliser des CNN pré-entraînés (comme VGG16 ou ResNet) sur de nouvelles tâches avec peu de données. L'idée est de réutiliser les couches basses du réseau (qui détectent des motifs généraux) et de ne réentraîner que les dernières couches. Cette approche est très utile quand on n'a pas assez de données pour entraîner un CNN from scratch. Elle réduit aussi le temps et les ressources nécessaires. [70]

Exemples d'application

Chapitre II: Les methodes de la classification des images et de paroles

Le transfer learning est souvent utilisé en imagerie médicale ou pour des projets où les données annotées sont rares. [71]

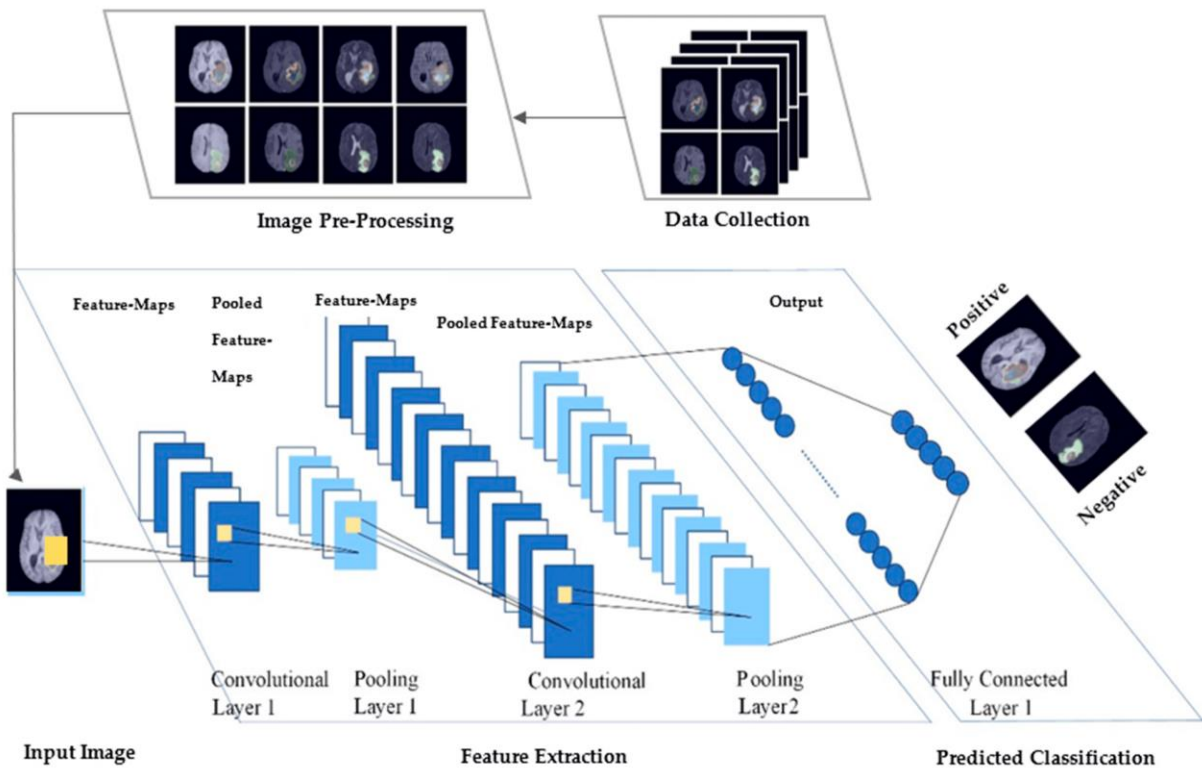


Figure II.9 : Exemple de transfert d'apprentissage (transfer learning) d'un modèle CNN pré-entraîné sur ImageNet vers l'imagerie médicale. [72]

II.6 Classification de la Parole

II.6.1 Méthodes Classiques

a) SVM / HMM (Modèles de Markov Cachés)

Comme démontré en Section II.1, les SVM exploitent leur capacité à traiter des caractéristiques acoustiques (MFCC) en haute dimension pour la classification de phonèmes. Cependant, leur application en reconnaissance vocale se heurte à deux limites majeures : (1) une sensibilité au bruit ambiant, et (2) une incapacité à modéliser les séquences temporelles - rôle dévolu aux HMM. Ces derniers, bien qu'efficaces pour représenter des transitions courtes (ex : entre phonèmes), peinent à capturer des dépendances à long terme (ex : corrélations syllabiques) [73]

Chapitre II: Les methodes de la classification des images et de paroles

◆ Exemples d'application

Ils ont été utilisés dans les premiers systèmes de reconnaissance vocale, comme ceux des assistants vocaux basiques. [74].

b) Modèles à base de GMM (Modèles de Mélange Gaussien)

Les GMM sont employés surtout pour la reconnaissance des locuteurs. Ils modélisent les données vocales comme un mélange de distributions gaussiennes. Leur principal défaut est leur simplicité : ils ne peuvent pas capturer la complexité de la parole continue ou les variations subtiles entre les locuteurs. [75]

◆ Exemples d'application

Ils étaient populaires dans les systèmes de vérification d'identité vocale.

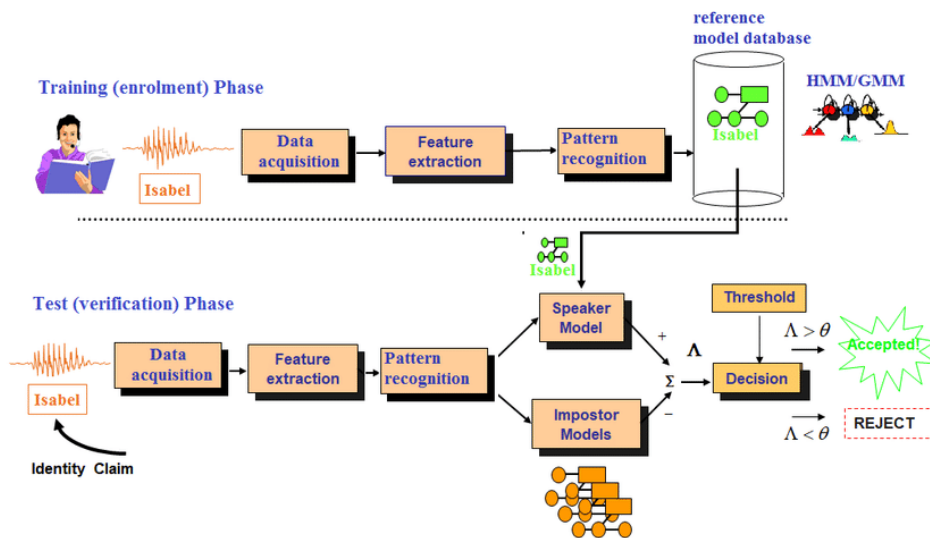


Figure II.10 : Schéma d'un système de vérification d'identité vocale : phases d'enregistrement, extraction des MFCC et comparaison au voiceprint. [76]

II.6.2 Méthodes d'Apprentissage Profond

a) RNN/LSTM (Réseaux Récurrents et Long Short-Term Memory)

Les LSTM sont spécialement conçus pour les données temporelles comme la parole. Grâce à leur mémoire à long terme, ils capturent mieux les dépendances contextuelles que les HMM ou les SVM. [77]

◆ Exemples d'application

Chapitre II: Les methodes de la classification des images et de paroles

Ils sont utilisés dans les assistants vocaux modernes comme Siri ou Google Assistant

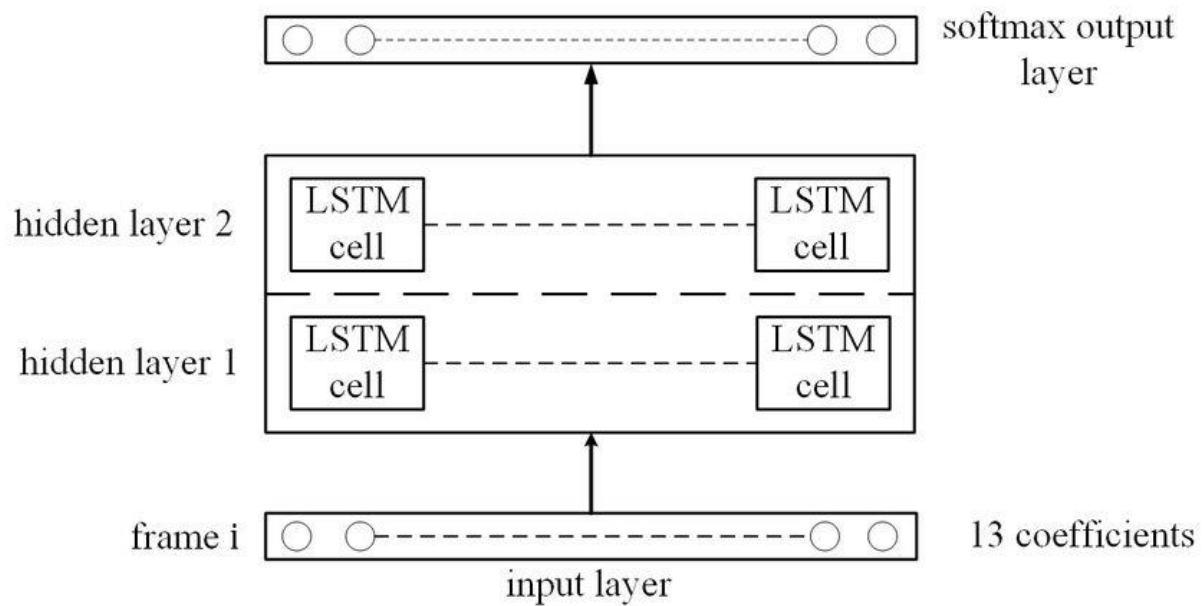


Figure II.11: Architecture d'un réseau LSTM à deux couches pour la reconnaissance vocale.[78]

b) CNN Adaptés à la Parole

Les CNN peuvent aussi traiter la parole en analysant les spectrogrammes comme des images 2D. Combinés avec des LSTM, ils deviennent très puissants pour la reconnaissance vocale. [79]

◆ Exemples d'application

Cette approche est utilisée pour la détection de mots-clés ou la transcription automatique

Chapitre II: Les methodes de la classification des images et de paroles

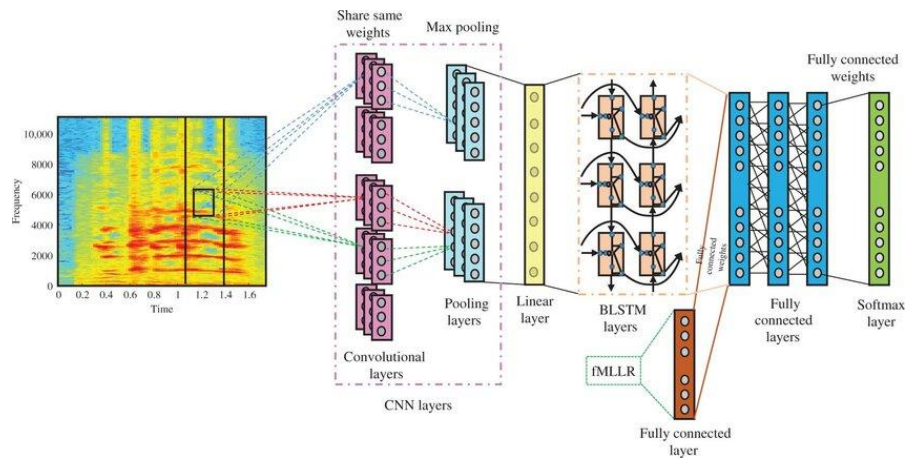


Figure II.12 : Architecture CNN LSTM appliquée à un spectrogramme pour la reconnaissance vocale (détection de mots clés / transcription) [80]

II.7 Analyse comparative des méthodes

Le choix entre les différentes approches de classification dépend principalement de la nature du problème à résoudre et des ressources disponibles. Plusieurs facteurs clés doivent être pris en considération afin de sélectionner la méthode la plus adaptée. Au commencement, la quantité de données disponibles joue un rôle déterminant. Les SVM s'avèrent particulièrement efficaces lorsqu'on travaille avec des jeux de données de taille réduite. Leur capacité à bien fonctionner avec un nombre limité d'échantillons en fait une solution intéressante dans de nombreux cas pratiques. À l'inverse, les CNN nécessitent généralement des volumes de données beaucoup plus importants pour atteindre leur plein potentiel, ce qui peut représenter un obstacle dans certaines situations. La complexité du problème constitue un autre élément crucial à considérer. Les SVM offrent l'avantage d'être relativement simples à comprendre et à implémenter, ce qui les rend accessibles même avec des connaissances modérées en apprentissage automatique. Cependant, cette simplicité se paye par une capacité moindre à traiter des problèmes très complexes. Les CNN, de leur côté, excellent dans la résolution de tâches sophistiquées grâce à leur architecture profonde, mais au prix d'une complexité algorithmique accrue. [81] Les ressources matérielles disponibles influencent également le choix de la méthode. Les SVM peuvent fonctionner de manière satisfaisante sur des machines standard sans nécessiter d'équipement particulier. Cette caractéristique les rend très pratiques pour des

Chapitre II: Les methodes de la classification des images et de paroles

applications où les ressources sont assez limitées. Les CNN, en revanche, réclament généralement des cartes graphiques performantes (GPU) pour des temps d'entraînement raisonnables, ce qui peut représenter un investissement non négligeable. Enfin, la question de l'interprétabilité des résultats ne doit pas être négligée. Les SVM fournissent des modèles dont le fonctionnement reste relativement transparent et explicable, un atout majeur dans les domaines où la traçabilité des décisions est importante. Les CNN, avec leur architecture en "boîte noire", offrent moins de visibilité sur les mécanismes internes qui conduisent aux prédictions. En conclusion, le choix entre SVM et CNN doit s'effectuer en fonction d'un compromis entre ces différents paramètres. Les SVM représentent une solution robuste et efficace pour des problèmes de complexité modérée avec des données limitées, tandis que les CNN s'imposent comme l'option privilégiée lorsqu'on dispose de ressources suffisantes et qu'on doit traiter des tâches complexes avec d'importants volumes de données. [82].

II.8 Traitement avec SVM

II.8.1 Pour les images

Les SVM représentent une excellente solution pour la classification d'images, particulièrement dans des situations où les ressources sont limitées. Leur principal atout réside dans leur simplicité d'utilisation et leur efficacité avec des jeux de données de taille modérée. Contrairement aux CNN qui nécessitent d'énormes quantités de données pour bien fonctionner, les SVM peuvent donner des résultats satisfaisants avec beaucoup moins d'échantillons. Un autre avantage majeur des SVM est leur transparence. Alors que les CNN agissent comme des boîtes noires, les SVM permettent de comprendre la manière dont les décisions sont prises. Cette interprétabilité est cruciale dans de nombreux domaines comme la médecine ou la finance, où il est important d'expliquer les prédictions du modèle. Les SVM sont également très robustes face au surapprentissage, grâce à leur principe de maximisation de la marge. Même avec peu de données, elles parviennent à généraliser correctement. De plus, leur flexibilité est renforcée par l'utilisation des fonctions noyaux (kernel trick) qui leur permet de traiter des problèmes non linéaires en les transformant dans un espace de plus haute dimension. [83].

Chapitre II: Les methodes de la classification des images et de paroles

◆ Exemples concrets

- Dans le domaine médical, les SVM sont utilisées pour classifier des images radiographiques simples, comme la détection de fractures osseuses.
- Pour des tâches industrielles, elles permettent de vérifier rapidement si des pièces manufacturées présentent des défauts visibles.

II.8.2 Pour les parole

Les SVM peuvent aussi être appliquées au traitement de la parole, notamment pour des tâches de classification simples. En utilisant des caractéristiques acoustiques comme les MFCC (Mel-Frequency Cepstral Coefficients), les SVM parviennent à distinguer efficacement différents phonèmes ou à reconnaître des locuteurs. Là où les méthodes profondes comme les LSTM demandent des ressources importantes, les SVM offrent une alternative légère et rapide à mettre en œuvre. Elles sont particulièrement adaptées aux systèmes embarqués ou aux applications nécessitant des temps de réponse très courts. [84].

II.9 Avantages par rapport aux autres méthodes

II.9.1 Face aux CNN

- Les SVM sont beaucoup plus simples à comprendre et à implémenter, ne nécessitant pas de connaissances approfondies en deep learning
- Elles demandent moins de données pour obtenir des performances correctes
- Leur entraînement est généralement plus rapide et peut se faire sur du matériel standard.

II.9.2 Face aux RNN

- Contrairement aux RNN qui peuvent souffrir du problème du gradient qui disparaît, les SVM sont stables et prévisibles
- Elles ne nécessitent pas de longs réglages d'hyper paramètres comme c'est fréquemment le cas avec les réseaux neuronaux

Chapitre II: Les methodes de la classification des images et de paroles

- Leur implémentation est plus directe, sans avoir à se soucier d'architecture complexe.

II.10 Implémentation proposée

II.10.1 Outils et frameworks

Afin de mettre en œuvre une solution basée sur les SVM, plusieurs outils simples mais efficaces sont disponibles :

Python reste le langage de choix, avec des bibliothèques comme scikit-learn qui offrent des implémentations optimisées des SVM. Contrairement aux frameworks de deep learning, ces outils sont légers et ne nécessitent pas de matériel spécialisé. Pour le prétraitement des images, OpenCV permet d'extraire facilement des caractéristiques comme les histogrammes ou les moments. Pour la parole, la bibliothèque Librosa est idéale pour calculer les MFCC et autres descripteurs acoustiques. [85].

II.10.2 Approche recommandée

La mise en place d'un classifieur SVM suit généralement ces étapes :

- Extraction des caractéristiques pertinentes selon le type de données
- Sélection et réglage du noyau approprié (linéaire, RBF, polynomial)
- Optimisation des hyperparamètres comme le coefficient de régularisation C
- Évaluation des performances sur un jeu de test séparé

L'avantage de cette approche est sa rapidité de mise en œuvre. Alors qu'un modèle CNN peut demander des jours de réglage et d'entraînement, une solution SVM peut souvent être opérationnelle en quelques heures seulement. [86].

II.10.3 Optimisation des performances

Quelques conseils pour tirer le meilleur des SVM :

- Normaliser soigneusement les données en entrée.

Chapitre II: Les methodes de la classification des images et de paroles

- Utiliser des techniques de sélection de caractéristiques afin d'en réduire la dimensionnalité.
- Pour des problèmes complexes, envisager l'utilisation de noyaux RBF avec réglage du paramètre gamma.
- En cas de déséquilibre des classes, ajuster le paramètre class-weight..

II.11 Conclusion du chapitre

La combinaison des histogrammes RGB, LBP, et MFCC permet de créer des représentations riches des données qui facilitent ainsi la séparation des classes par l'hyperplan SVM. Cette approche est essentielle pour des applications telles que la reconnaissance d'objets et la classification de la parole, où la précision et la robustesse des modèles sont primordiales. Les SVM constituent une méthode robuste et polyvalente pour de nombreuses tâches de classification, que ce soit sur des images ou de la parole. Leur principal atout réside dans leur excellent compromis entre performance et simplicité. Alors que les méthodes d'apprentissage profond dominent actuellement le paysage, les SVM restent pertinentes dans de nombreux cas pratiques, notamment quand les ressources sont limitées ou quand l'interprétabilité des résultats est importante. [87] Le choix entre SVM et CNN doit se faire en fonction du contexte. Pour des prototypes rapides, des applications embarquées ou des problèmes ne nécessitant pas une précision extrême, les SVM offrent souvent la solution la plus pragmatique. Leur mise en œuvre simple et leur faible besoin en ressources computationnelles en font un outil toujours indispensable dans la boîte à outils du data scientist. [88].

Chapitre III : Experience et resultats

Chapitre III : Experience et resultats

III.1 Introduction

Ce chapitre est consacré à la présentation de notre méthodologie qui consiste à faire la classification des images et de la parole. Celles-ci sont obtenues à partir de la construction de deux bases de données, la première est consacré pour les images des visages qui contiennent 5 classes : 1 En Colère 2 Dégout 3 Peur 4 Triste 5 malade, la deuxième est consacrée pour la parole de la personne et qui contient 5 classes : 1 malade 2 peur 3 secours 4 mourir 5 danger. L'objectif est de détecter les situations dangereuses d'une personne qui habite dans le bâtiment intelligent à partir de son visage et de sa voix afin de le secourir. Notre objectif est la conception d'un système qui pour objectif de secourir une personne en danger, tel que ce système est composé de deux systèmes montés en parallèle, le premier système fait la classification des images, le deuxième système fait la classification de la parole.

III.2 Le premier système : la classification des images

Nous avons réalisé une base de données qui contient 900 images, tel que ces images représentent contiennent 5 classes : 1 En Colère 2 Dégout 3 Peur 4 Triste 5 malade, tel que chaque classe contient 180 images. 600 images sont utilisées pour la phase d'apprentissage et 300 images sont utilisées pour la phase de test ou la classification.

Pour la paramétrisation de ces images nous avons utilisé, les histogrammes de couleurs RGB normalisés et les histogrammes de textures LBP normalisés, en se basant sur les SVM (support vector Machine) comme un outil de classification.

Nous appliquons à chaque image les histogrammes de couleurs RGB normalisés et les histogrammes de textures LBP normalisés pour la paramétrisation de ces images. Nous représentons chaque image par un vecteur de même taille contenant 320 valeurs qui est le résultat de la concaténation des deux histogrammes de couleur RGB et de texture LBP. La mise en œuvre se fait en deux phases : apprentissage et classification. En exploitant les SVM (support machines des vecteurs), nous appliquons une méthode d'apprentissage afin de réduire les erreurs liées à la reconnaissance du système, suivie d'une technique de classification qui détermine les 5 classes : 1 En Colère 2 Dégout 3

Chapitre III : Experience et resultats

Peur 4 Triste 5 malade pour détecter les situations dangereuses d'une personne qui habitent dans le bâtiment intelligent à partir de son visage afin de le secourir.

III.2.1 Paramétrisation des données

Après avoir réalisé notre corpus d'entrée, nous avons entamé l'étape suivante et qui est la paramétrisation des images. Nous avons utilisé un prétraitement afin d'augmenter la qualité de nos images. Pour la paramétrisation des images, nous avons opté pour les histogrammes RGB normalisés pour la couleur et les histogrammes LBP normalisés pour la texture, deux histogrammes pour chaque image ont été utilisés, un pour la couleur (RGB normalisé) et l'autre pour la texture (LBP). Ce choix est justifié par les résultats des travaux de recherche présentés et qui ont démontré que l'utilisation de la couleur et la texture permet d'améliorer les résultats de classification.

Concernant les histogrammes de couleurs RGB, Nous avons représentés chaque image par un histogramme tridimensionnel 3D, au lieu de trois histogrammes monodimensionnel (R, G et B). Chaque composante colorimétrique R, G, B est quantifiée en 4 valeurs, ce qui donne un nombre de classes de 4^3 car nous avons 3 dimensions R, G, B. et donc la dimension de l'histogramme et de $4^3 = 64$ éléments. Ce qui implique que chaque histogramme tridimensionnel 3 D de couleur RGB est un vecteur de taille 6 et donc peut représenter 64 valeurs.

L'inconvénient des histogrammes de couleurs RGB est qu'ils sont sensibles au changement de l'éclairage, afin de diminuer cette sensibilité nous utilisons la normalisation de ces histogrammes.

Concernant les histogrammes de textures LBP normalisés, le niveau de la quantification de ces histogrammes est de $2^8 = 256$. On montre que la fusion des deux histogrammes normalisés de couleurs RGB 3D et de textures LBP est intéressante. C'est ce qui explique que nous avons fusionnés les histogrammes de couleurs RGB normalisé, et les histogrammes de textures LBP normalisés. Chaque image est représentée par un vecteur de $(64 + 256)$ valeurs = 320 valeurs.

Pour la phase de la mise en œuvre et après une longue recherche, nous avons opté pour les SVM light (Support Vector Machines) en raison de leurs bonnes capacités de généralisation dans de nombreux problèmes, mais aussi pour leurs bons résultats sur des

Chapitre III : Experience et resultats

ensembles de données d'apprentissage de petite taille (contrairement aux réseaux de neurones qui demandent de la base de données plus importante). Ils ont montré ces dernières années leur puissance surtout leur temps de calcul qui est très satisfaisant par rapport à d'autres techniques Les SVM lights sont utilisés avec la méthode « un contre tous » associée à un vote majoritaire. En conclusion le corpus d'entrée que nous avons réalisée, est constitué de 900 images, tel que chaque vecteur représente une image qui est la concaténation de l'histogramme de couleur normalisés RGB 3D (4 bits) et les histogrammes de textures normalisés LBP (8 bits).

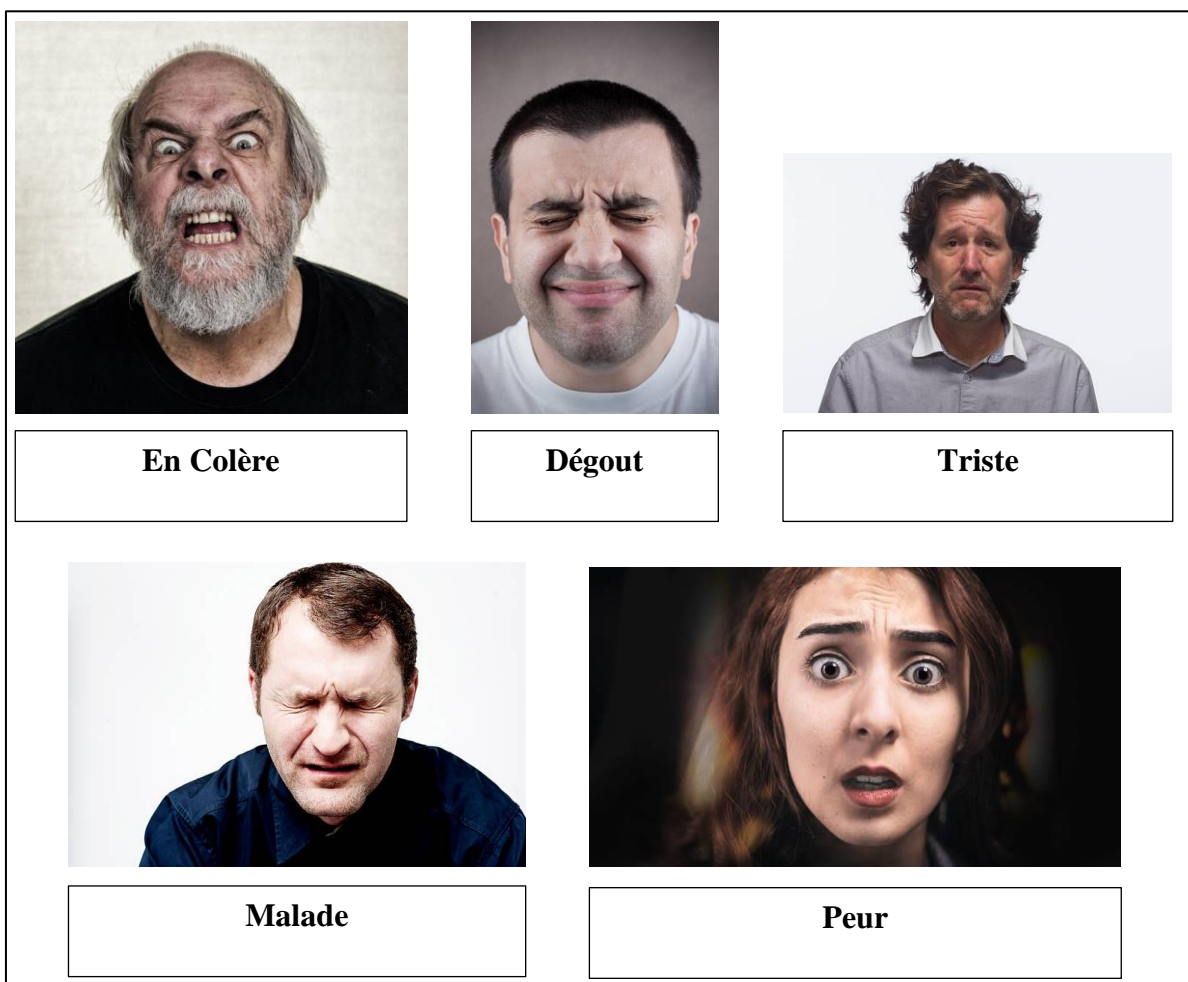


Figure III.1 : Un exemple des images du visages à classer

Chapitre III : Experience et resultats

III.2.2 Apprentissage

.Dans notre cas l'apprentissage est supervisé car on connaît la sortie. Les exemples particuliers sont représentés par un ensemble de couples d'entrée/sortie.

Parmi les modèles des SVM, on constate les cas linéairement séparable et les cas non linéairement séparable. La mise en œuvre, des classifieurs SVM dans le contexte du cas linéairement séparable, est simple. Pour surmonter les inconvénients des cas non linéairement séparable, l'idée est d'utiliser la fonction noyau qui consiste à changer l'espace des données,. En pratique, quelques familles de fonctions noyau paramétrables sont connues et il revient à l'utilisateur de SVM d'effectuer des tests pour déterminer celle qui convient le mieux pour son application. On peut citer les exemples de noyaux suivants : polynomiale, gaussien, sigmoïde et RBF (radial basic function).

Dans la plupart des problèmes réels il n'y a pas de séparation linéaire possible entre les données du corpus d'entrée, ce qui est le cas pour notre corpus. C'est pour cette raison que nous utilisons les fonctions noyaux afin que nos données soient séparables.

Pour la mise en œuvre. Comme nous avons 5 classes à reconnaître, les SVM vont générer 5 modèles (classifieurs ou hyperplan séparateurs) dans la phase d'apprentissage ou d'entraînement, tel que chaque modèle correspond à une classe (Figure.III.1)

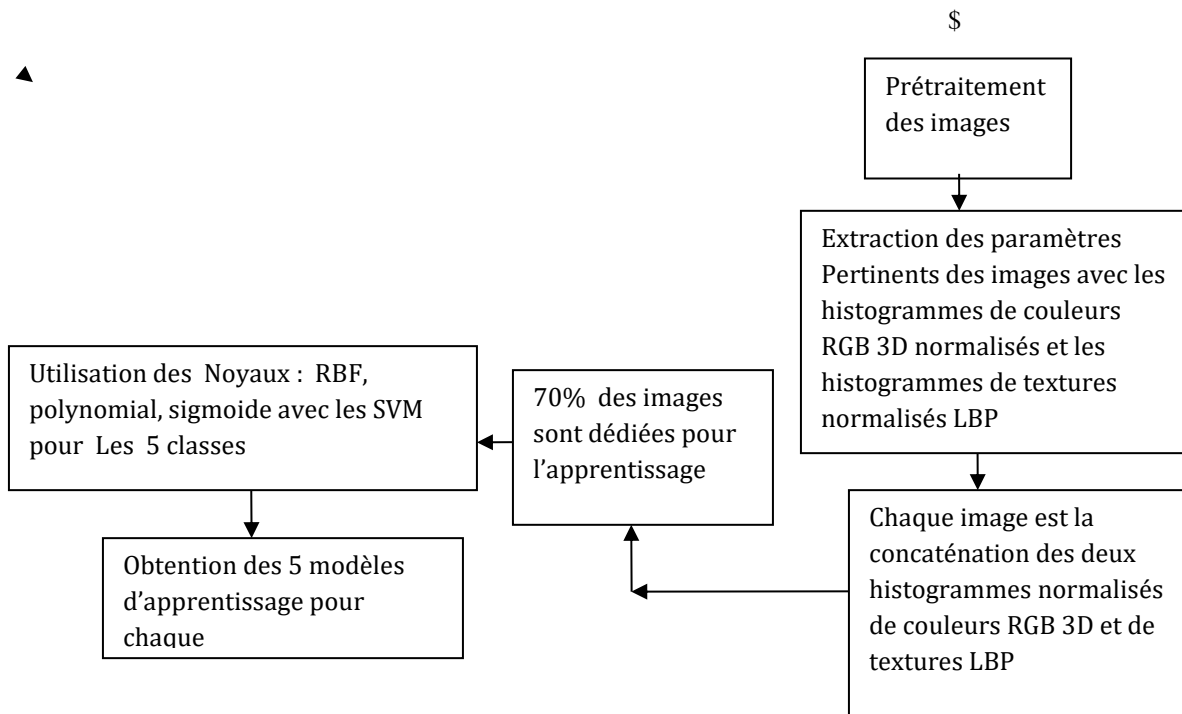


Figure III.2 Apprentissage des classes par les SVM

Chapitre III : Experience et resultats

III.2.3 Reconnaissance

Les SVM (Support Vector Machine ou Séparateurs à Vaste Marge) ont pour objectif de rechercher le meilleur hyperplan de séparation des comportements en deux classes.

La classification d'une classe est donnée par sa position par rapport au meilleur hyperplan (modèle ou le classifieur) trouvé dans la phase d'apprentissage. Nous avons utilisés pour la classification d'une classe, l'approche dite "un contre tous", qui compare successivement une classe avec toutes les autres au moyen d'un classifieur binaire. Comme on a 5 classes, alors 5 classifieurs binaires sont construits (5 hyperplans, soit 5 problèmes d'optimisation), le n-ième classifieur étant destiné à distinguer la classe d'indice n. Chaque élément à classifier est donc présenté aux 5 classifieurs et se voit attribuer l'étiquette du classifieur ayant renvoyé le plus grand pourcentage d'appartenance.

La figure III.2 montre la méthode que nous avons utilisée pour la classification d'un comportement par les SVM

Chapitre III : Experience et resultats

III.2.4 Les résultats obtenus

III.2.5.1 Classification avec SVM Lights utilisant le noyau RBF (radial basic function)

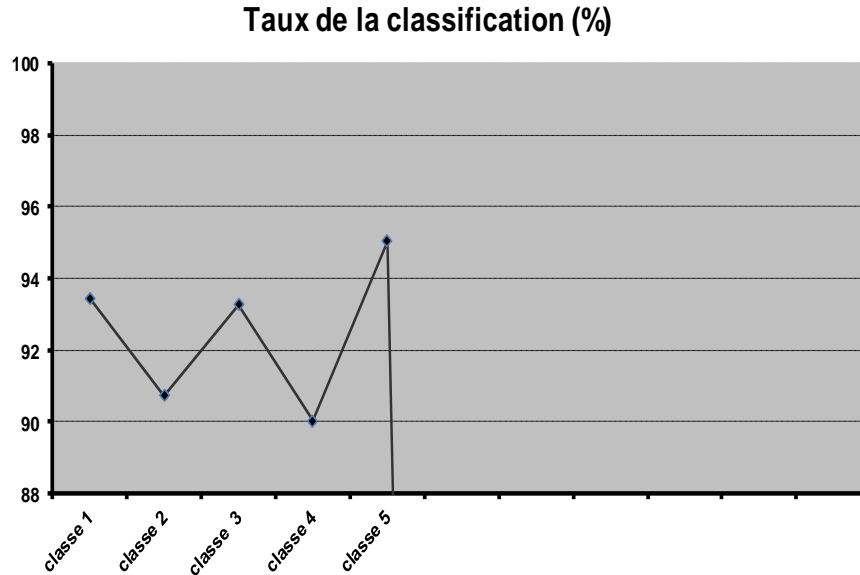


Figure III.3 représentation du pourcentage des classes correctement détectés pour le noyau RBF en utilisant les histogrammes de couleur et de texture RGB, LBP et les SVM

Tableau III.1 Tableau représentant le rappel et la précision Pour la détection des comportements pour le noyau RBF en utilisant les histogrammes de couleur et de texture RGB, LBP et les SVM Lights.

classes à classer	Précision pour Noyau RBF	Rappel pour Noyau RBF
Classes 1	74%	68.5%
Classes 2	31.55%	39%
Classes 3	68%	76.91%
Classes 4	38.47%	44.92%
Classes 5	85.98%	66.12%

I

Chapitre III : Experience et resultats

III.2.5.2 Classification avec SVM utilisant le noyau polynomial

Taux de la classification (%)

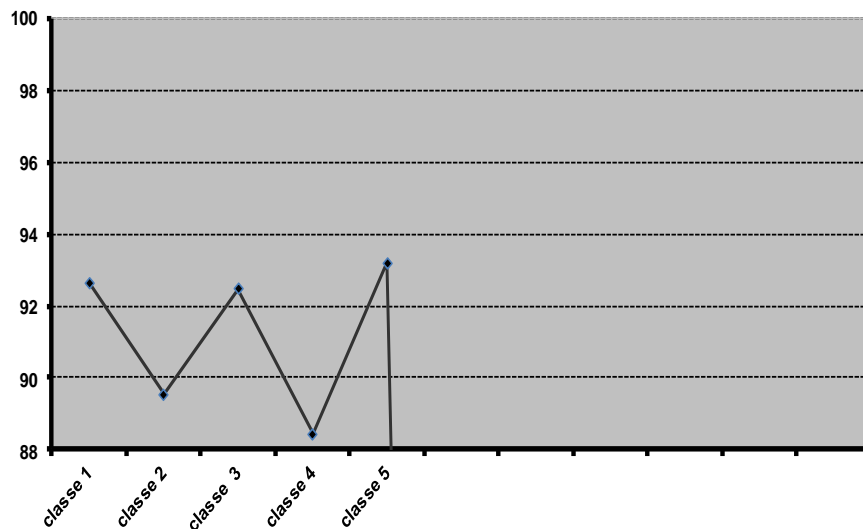


Figure III.4 représentation du pourcentage des classes correctement détectés pour le noyau polynomial en utilisant les histogrammes de couleur et de texture RGB, LBP et les SVM

Tableau III.2 Tableau représentant le rappel et la précision Pour la détection des comportements pour le noyau polynomial en utilisant les histogrammes de couleur et de texture RGB, LBP et les SVM.

classes à classer	Précision pour Noyau polynomial	Rappel pour Noyau polynomial
Classes 1	73.10%	67.25%
Classes 2	30.35%	38.12%
Classes 3	67.24%	75.75%
Classes 4	37.25%	43.63%
Classes 5	84.66%	65.24%

I

Chapitre III : Experience et resultats

III.2.5.1 Classification avec SVM Lights utilisant le noyau sigmoïde

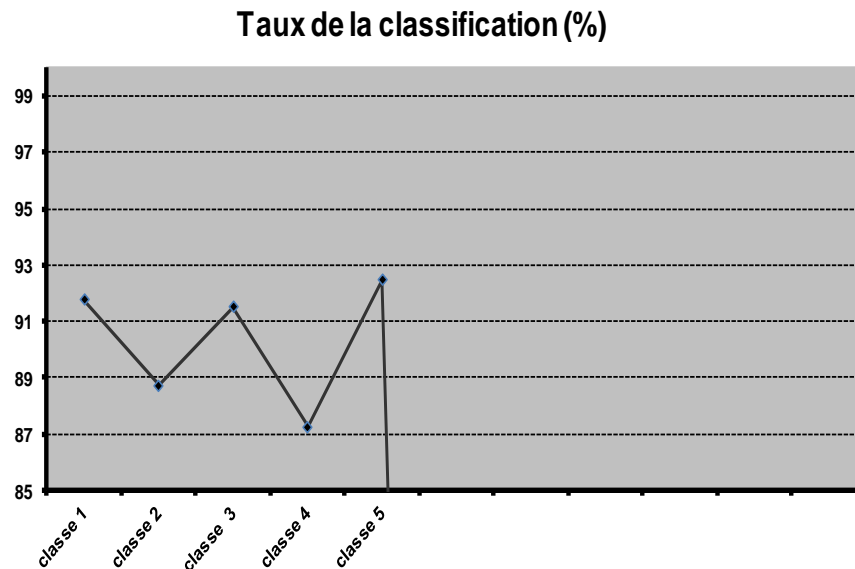


Figure III.5 représentation du pourcentage des classes correctement détectés pour le noyau en sigmoïde utilisant les histogrammes de couleur et de texture RGB, LBP et les SVM

Tableau III.3 Tableau représentant le rappel et la précision Pour la détection des comportements pour le noyau sigmoïde en utilisant les histogrammes de couleur et de texture RGB, LBP et les SVM

classes à classer	Précision pour Noyau sigmoïde	Rappel pour Noyau sigmoïde
Classes 1	73.12%	66.75%
Classes 2	29.75%	37.52%
Classes 3	66.23%	74.25%
Classes 4	36..27%	43.22%
Classes 5	83.48%	64.52%

Chapitre III : Experience et resultats

III.2.5 Conclusion

Nous avons présenté une méthode qui se fonde sur l'utilisation des histogrammes de couleurs RGB normalisés et des histogrammes de textures LBP normalisés pour la paramétrisation des images, et les SVM pour la mise en œuvre. Afin de prendre une décision concernant la

classification d'une classe donnée, nous avons effectués deux étapes : une première étape d'apprentissage et une seconde étape de test. Dans la phase d'apprentissage nous avons utilisés trois noyaux, polynomial, sigmoïde, RBF (radial basic function), afin de prendre le meilleur noyau qui donne les meilleurs résultats.

. D'après les résultats obtenus, nous pouvons dire que nos résultats obtenus sont satisfaisants, et donc l'utilisation des histogrammes de couleurs RGB normalisés , de textures LBP normalisés pour la paramétrisation des images, et les SVM pour la classification sont recommandés pour la classification des images.

III.3 Le deuxième système : la classification de la parole

La reconnaissance de mots isolés par les SVM-MFCC, dans un environnement naturel, est une classification globale. Elle comporte 4 étapes :

- 1) Acquisition du signal de la parole et formation du corpus d'entrée
- 2) prétraitement et paramétrisation par MFCC
- 3) conception du SVM apprentissage
- 4) test et résultats

III.3.1 Acquisition et corpus d'entrée :

1) Acquisition : Elle a été faite sur PC : GPU : GTX 1650 4GB VRAM – CPU : Ryzen 5 3600 – DD : 3.5 Térabyte – RAM 16GB – carte son E24-20 – fréquence d'échantillonnage 11025HZ. Les échantillons sont codés sur 8 bits.

2) Corpus d'entrée : La base de données est formée de 165 mots, e vocabulaire utilisé est constitué des 5 mots suivants : 1 malade 2 peur 3 secours 4 mourir 5 danger

Chaque mot est prononcés 3 fois par 11 locuteurs (7 hommes et 4 femmes) dont l'âge varie entre 20ans et 75 ans .ce corpus sera utilisé comme suit : Apprentissage : 100 mots, test : 65 mots

Chapitre III : Experience et resultats

III.3.2 Prétraitement et paramétrisation

1) Prétraitement : Isolation du mot (élimination des silences de début et de fin) en utilisant les critères de niveaux d'énergie et de taux de passage par zéro.

2) Paramétrisation : La paramétrisation du signal est réalisée à l'aide des deux opérations suivantes :

2.1) une préaccentuation : atténue les bases fréquences par rapport aux hautes fréquences.

2.2) Fenêtrage : La fenêtre utilisée est celle de HAMMING avec une largeur de 23.2ms (soit 512 échantillons), et un recouvrement de 10ms (soit 220 échantillons) afin de ne pas perdre l'information entre deux fenêtres. La pondération par la fenêtre est utilisée pour atténuer l'effet des ondulations parasites. Elle est un bon compromis entre la limitation du signal et les effets de distorsion dus à la convolution des transformées.

3) extraction des paramètres significative de la parole : Après avoir décomposé le signal en 39 trames il serait extrait de chacune d'entre elles un vecteur de 13 coefficients MFC ; ces coefficients ou paramètres significatifs extraits de chaque signal de parole sont représentés par des vecteurs dont la taille est égale pour chacun d'entre eux à [1x507].

III.3.3 Formation du corpus d'entrée :

Les coefficients d'un même mot prononcé par un locuteur homme ou femme représentent une partie de l'échantillon du mot utilisé à l'entrée du SVM. Chaque échantillon est constitué des vecteurs transposés des coefficients de chaque spectre du mot prononcé présentés les uns à côté des autres pour former à la fin une matrice de taille égale à [507x165] pour le premier corpus ,.

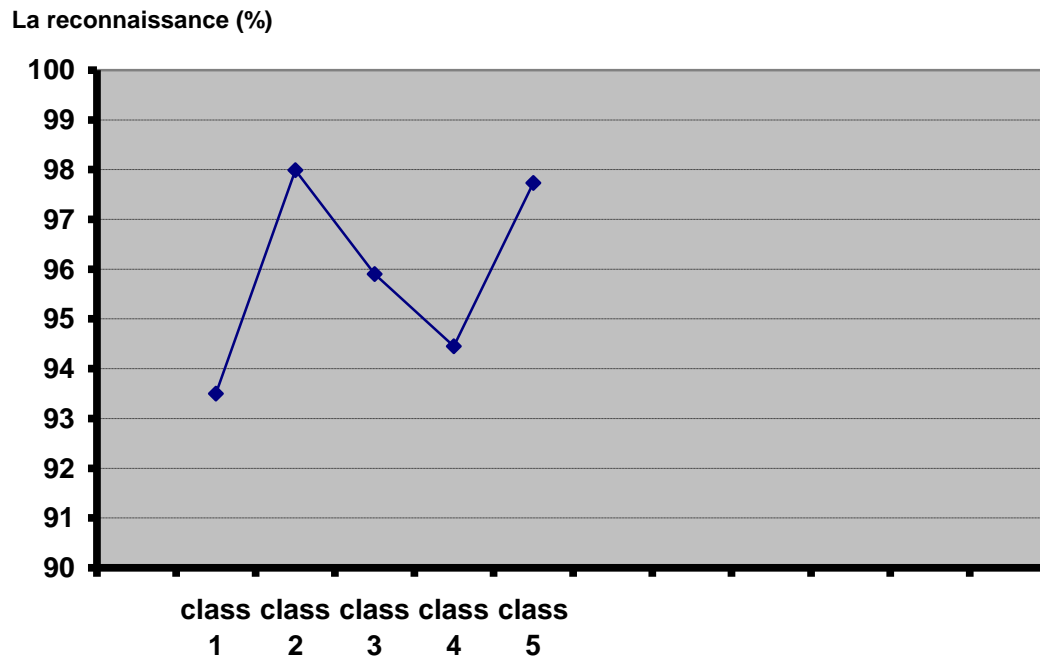
III.3.4 La classification par SVM :

ces méthodes de classification sont binaire, On a utilise l'approche 1 parmi K pour construire les classifieurs, 5 classifieurs pour le premier corpus et 10 classifieurs pour le deuxième corpus. Les données d'apprentissage étant linéairement séparables. L'apprentissage comprend le réglage du seul hyper paramètre C (on a fixe $c \rightarrow \infty$) et la détermination des paramètres optimums b_0 et les alpha qui définit 'hyperplan séparateur pour chaque sous classifieur.

Chapitre III : Experience et resultats

III.3.5 Test et resultats

Les courbes ci-dessous illustres les resultats obtenus :



Sous classifieurs

Figure III.6 la variation du taux de la reconnaissance en fonction du sou classifieur SVM noyau RBF

Tableau III.3 Tableau representant le rappel et la precision pour la classification des mots isoles pour le noyau RBF en utilisant les MFCC et les SVM

classes à classer	Précision pour Noyau sigmoïde	Rappel pour Noyau sigmoïde
Classes 1	77.12%	64.75%
Classes 2	97.75%	99.2%
Classes 3	66.23%	74.25%
Classes 4	79.27%	72.22%
Classes 5	93.48%	94.52%

Chapitre III : Experience et resultats

III.3.6 Commentaire :

D'après les résultats obtenus dans toutes les figures et tableaux, on constate que :
Le taux de la reconnaissance des mots isolés est très bon ainsi que le rappel et la précision.
On conclut que les SVM et les MFCC sont recommandés pour la reconnaissance des mots isolés surtout pour le mono locuteur.

III.3.7 CONCLUSION

Nous avons mis au point un système de reconnaissance de mots isolés en utilisant une application des MFCC (Mel Fréquence Coefficient Cepstral) et des machines à vecteurs de support (SVM). La reconnaissance vocale reste un compromis entre la taille du vocabulaire, ses possibilités multi locuteur, son encombrement physique, sa rapidité, temps d'apprentissage. Les MFCC se sont révélées en effet être de très bonnes bases d'analyse du signal, permettant une représentation exacte des coefficients ou des paramètres significatifs extraits de chaque signal de la parole que celle obtenue par d'autre méthode. Pour l'apprentissage, une réalisation à cet effet a été faite d'un ensemble d'algorithmes (SVM), et à l'aide d'un SVM de type 1 parmi 'k', qui ont donné un score représentatif. Nos résultats obtenus confirment que la paramétrisation par les MFCC du signal de la parole (mots isolés) et la reconnaissance par les SVM donnent de très bons résultats.

Conclusion Générale

Notre projet de système intelligent de surveillance a rencontré plusieurs difficultés importantes pendant sa réalisation. Les bases de données gratuites que nous avons pu utiliser avaient malheureusement plusieurs défauts. Les images des visages des personnes que nous avons utilisées avaient une qualité très variable, avec des résolutions différentes, un éclairage instable et des prises de vue pas toujours bien cadrées, ce qui rendait leur traitement compliqué.

Nous avons remarqué que la plupart des bases de données accessibles gratuitement avaient beaucoup de problèmes : des images de tailles différentes, mal éclairées ou prises sous des angles irréguliers, à cause de ça, nous avons choisi de créer notre propre base de données à partir d'enregistrements vocaux simulés (demandes d'aide) et d'images montrant différentes émotions de danger faciales, même si cette étape a pris beaucoup de temps.

Le développement du système a aussi été freiné par les limites de notre matériel, si bien que nous avons eu des problèmes pour faire tourner des algorithmes trop lourds comme les CNN, ce qui nous a poussés à adapter notre choix d'algorithmes et la manière de traiter les données.

Nous avons donc choisi d'utiliser les SVM (machines à vecteurs de support) comme méthode principale de classification. Les SVM sont plus simples à entraîner et demandent moins de ressources que les réseaux de neurones convolutifs (CNN). Les CNN donnent souvent de bons résultats, mais dans notre cas, les SVM étaient plus adaptés car nous avions peu de données et un matériel limité.

Pour la paramétrisation des images, nous avons utilisé deux méthodes qui se complètent. La première consiste à utiliser les paramètres des histogrammes RGB pour voir comment les couleurs (rouge, vert, bleu) sont réparties dans l'image, cela donne une idée générale du contenu visuel. La deuxième consiste à utiliser les paramètres des histogrammes LBP (Local Binary Patterns) pour analyser la texture du visage. Ce type de

descripteur aide à voir les petits changements dans la lumière, ce qui est utile pour détecter les micro-expressions. En combinant les deux méthodes, on aura de très bons résultats.

Pour la parole, nous avons utilisé la méthode MFCC (coefficients cepstraux en fréquences de Mel), qui permet de transformer la parole en un petit ensemble de valeurs importantes. Ces valeurs résument le signal en fonction des fréquences que l'oreille humaine perçoit le mieux, ce qui permet d'identifier plus facilement si le son est une parole, etc., tout en gardant le traitement rapide.

Temps moyen de traitement : environ 30 minutes avec SVM et 3 heures avec CNN, pour chaque tentative de traitement, qu'elle se termine par un succès ou par une erreur.

Nous avons obtenu des résultats satisfaisants pour notre projet qui consiste à secourir une personne ou plusieurs personnes en danger dans un bâtiment intelligent, vu les limites techniques et le travail manuel que nous avons dû faire pour créer et annoter la base de données. Ceci nous permet de dire que nous pouvons développer un système utile et réaliste en combinant des méthodes classiques (comme SVM, MFCC, LBP, RGB). Comme perspectives, il pourra contribuer aux projets futurs, comme l'ajout de modèles plus puissants (comme les CNN) si nous disposons d'un meilleur matériel, ou une application réelle dans un bâtiment intelligent pour aider les personnes âgées.

Reference

Références

- [1] OMS. (2024). **Global Health and Aging Report**. Genève.
- [2] Eurosafe. (2023). **Injury Statistics Report 2022-2023**. Bruxelles.
- [3] Caradec, V. (2024). **Technologies du vieillissement : enjeux éthiques et sociaux**. PUF.
- [4] Buckman, A. et al. (2024). "Smart Building Architectures for Health Monitoring". **IEEE IoT Magazine**, 7(2).
- [5] Chen, L. & Nugent, C.D. (2023). **Real-Time Activity Recognition in Smart Environments**. Springer.
- [6] Cook, D.J. (2024). "Adaptive AI for Personalized Elder Care". **Journal of Ambient Intelligence**, 15(1).
- [7] Rashidi, P. (2024). **Economic Impacts of AI in Elderly Care**. CRC Press.
- [8] Yu, X. et al. (2024). "Advanced Fall Detection Using 3D CNN". **IEEE Journal of Biomedical Imaging**, 11(3).
- [9] European Commission. (2024). **Ethical Guidelines for AI in Elder Care** Bruxelles.
- [10] "Deep Learning for Computer Vision: A Brief Review" 2020 **International Conference on Computer and Communication Systems**.
- [11] UNDESA, **World Population Prospects 2022**.
- [12] <https://www.la-croix.com/Actualite/Monde/La-population-mondiale-vieillit-de-plus-en-plus-vite- NG -2009-07-22-537307>
- [13] WHO, **Global Report on Ageing 2021**.
- [14] Eurostat, **Ageing Europe 2023**.
- [15] OECD, **Health at a Glance 2023: Elderly Care Expenditures**, Paris: OECD Publishing.
- [16] <https://biblus.acca.it/smart-building-esempio-pratico/>
- [17] MIT Review, "The Sensor Revolution" 2022.

Reference

- [18] KNX Association, KNX Handbook for Smart Ageing Solutions, Brussels: KNX Standard v3.0, 2022.
- [19] Cloud Security Alliance, Edge Computing Guidelines for Healthcare IoT, Seattle: CSA Report, 2023.
- [20] <https://medium.com/@monge/the-evolution-of-technology-and-its-impact-on-business-393fe6c6e8d2>
- [21] ACM Transactions on Human-Robot Interaction, "Emotional AI in Assistive Environments", Vol.12, No.2, June 2023.
- [22] J. Gerontechnology, Vol.12(3), 2023.
- [23] French National Academy of Medicine, Ethical Guidelines for Gerontechnologies, Paris: Report No.45, 2023.
- [24] IEEE Sensors Journal, "AI for Fall Prediction", 2023.
- [25] <https://www.orange-business.com/fr/produits/smart-building-alerting>
- [26] International Telecommunication Union, 6G Framework for Assisted Living, Geneva.
- [27] <https://tolumichael.com/is-cyber-security-or-real-estate-a-better-career/>
- [28] EU Commission, Directive (EU) 2023/123 on Smart Building Interoperability, Brussels: Official Journal L 456, 15/01/2023.
- [29] IEEE IoT Journal, "Low-Latency Architectures", 2022.
- [30] LoRa Alliance, LPWAN Deployment Best Practices, Fremont: Technical Specification v2.1, 2022.
- [31] EU GDPR, Article 89 Special Protections for Health Data in AI Systems, 2023 Amendment.
- [32] Lancet Digital Health, "Acceptability of Monitoring Technologies in Elderly", Vol.5, Issue 3, March 2024.
- [33] <https://www.accasoftware.com/fr/plateforme-bim-iot>
- [34] Nature Nanotechnology, "Biometric Nanosensors", Jan 2024.
- [35] <https://www.jefferson.edu/institute-for-smart-and-healthy-cities/seed-grant-research-projects/projects.html>

Reference

- [36] IEEE 6G Initiative, Ultra-Reliable Low-Latency Communication for Healthcare White Paper, Feb 2024.
- [37] <https://www.alamy.com/stock-photo-digital-vector-artificial-intelligence-and-electronic-technologies-170036153.html>
- [38] EU Ageing Society Observatory, Policy Toolkit for Smart Ageing Infrastructure, Brussels: Publications Office, 2023.
- [39] Cortes, C., & Vapnik, V. (1995). Support-Vector Networks. *Machine Learning*, 20(3), 273–297.
- [40] Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. Springer.
- [41] Gonzalez, R. C., & Woods, R. E. (2018). *Digital Image Processing*. Pearson.
- [42] Szeliski, R. (2010). *Computer Vision: Algorithms and Applications*. Springer.
- [43] Matter Alliance. (2022). *Matter: The Connectivity Standard for Smart Home Devices*. Retrieved from <https://buildwithmatter.com/>
- [44] <https://fastercapital.com/topics/identifying-business-objectives.html>
- [45] OMS, Rapport mondial sur le vieillissement, 2022.
- [46] Chen et Nugent, "Smart Building Systems for Elderly Care", IEEE, 2023.
- [47] INSEE, Projections démographiques 2023-2070, 2023.
- [48] Yu et al., "Fall Detection Using CNN", *Journal of Gerontechnology*, 2024.
- [49] Caradec, "Technologies et vieillissement", PUF, 2021.
- [50] Zhang, L., Wu, Q., & Zhang, D. (2010). A Comparative Study of LBP and Gabor Features for Face Recognition. In *Proceedings of the International Conference on Image Processing*.
- [51] Lowe, D. G. (2004). Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision*, 60(2), 91-110.
- [52] <https://medium.com/data-science/histograms-in-image-processing-with-skimage-python-be5938962935>
- [53] https://www.researchgate.net/figure/a-LBP-image-b-LBP-histogram_fig1_271740362

Reference

- [54] https://www.researchgate.net/figure/illustration-of-image-matching-using-Laplacian-interest-points-with-locally-adapted-SIFT_fig1_267390438
- [55] https://www.researchgate.net/figure/Speech-Signal-Spectrogram-MFCC-Spectrogram-and-Gammatongram-of-Female-utterance-of_fig6_320021442
- [56] [https://www.researchgate.net/figure/Support-vector-machine-SVM-worked-by-finding-a-separating-hyperplane-with-maximal_fig2_265603987#:~:text=Support%20vector%20machine%20\(SVM\)%20worked%20by%20finding,separating%20the%20data%20points%20into%20two%20classes.](https://www.researchgate.net/figure/Support-vector-machine-SVM-worked-by-finding-a-separating-hyperplane-with-maximal_fig2_265603987#:~:text=Support%20vector%20machine%20(SVM)%20worked%20by%20finding,separating%20the%20data%20points%20into%20two%20classes.)
- [57] Mikolajczyk, K., & Schmid, C. (2005). A Performance Evaluation of Local Descriptors. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(10), 1615-1630.
- [58] Davis, S., & Mermelstein, P. (1980). Comparison of Parametric Representations for Monosyllabic Word Recognition in Continuously Spoken Sentences. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 28(4), 357-366.
- [59] Sainath, T. N., Parada, C., & Kingsbury, B. (2015). Convolutional Neural Networks for Small-Footprint Keyword Spotting. *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 947-951.
- [60] Cortes, C., & Vapnik, V. (1995). Support-Vector Networks. *Machine Learning*, 20(3), 273–297.
- [61] Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. Springer.
- [62] Schölkopf, B., & Smola, A. J. (2002). *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. MIT Press.
- [63] <https://www.mdpi.com/2076-3417/9/15/3130>
- [64] Cristianini, N., & Shawe-Taylor, J. (2000). *An Introduction to Support Vector Machines and Other Kernel-based Learning Methods*. Cambridge University Press.
- [65] Zhang, C., & Schneider, J. (2010). A Support Vector Machine Approach to Image Classification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.
- [66] <https://link.springer.com/article/10.1007/s11831-024-10194-4>
- [67] Hsu, C. W., & Lin, C. J. (2002). A Comparison of Methods for Multiclass Support Vector Machines. *IEEE Transactions on Neural Networks*, 13(2), 415–425.

Reference

- [68] Vapnik, V. (1998). *Statistical Learning Theory*. Wiley.
- [69] <https://pdseulpace.tistory.com/m/52>
- [70] Guyon, I., & Elisseeff, A. (2003). An Introduction to Variable and Feature Selection. *Journal of Machine Learning Research*, 3, 1157–1182.
- [71] Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32.
- [72] <https://encyclopedia.pub/entry/42866>
- [73] Cover, T., & Hart, P. (1967). Nearest neighbor pattern classification. *IEEE Transactions on Information Theory*, 13(1), 21–27.
- [74] LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11), 2278–2324.
- [75] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems (NeurIPS)*, 25, 1097–1105.
- [76] https://www.researchgate.net/figure/Block-diagram-of-a-speaker-recognition-system-training-test_fig1_322412956
- [77] Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *International Conference on Learning Representations (ICLR)*.
- [78] https://www.researchgate.net/figure/LSTM-recurrent-neural-network-for-speech-recognition_fig5_322927149
- [79] He, K., Zhang, X., Ren, S., & Sun, J. (2015). Deep residual learning for image recognition. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 770-778.
- [80] https://www.researchgate.net/figure/An-Illustration-of-the-CNN-BLSTM-Architecture-Applied-to-Log-Mel-FB-Features-of-the_fig1_331532452
- [81] Jelinek, F. (1997). *Statistical Methods for Speech Recognition*. MIT Press.
- [82] Vapnik, V. (1995). *The Nature of Statistical Learning Theory*. Springer.
- [83] Bouchard, A., & al. (2004). Gaussian mixture models for speaker recognition. *Speech Communication*, 43(1-2), 93–108.

Reference

- [84] Graves, A., Mohamed, A. R., & Hinton, G. (2013). Speech recognition with deep recurrent neural networks. *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 6645–6649.
- [85] Sainath, T. N., Parada, C., & Kingsbury, B. (2015). Convolutional neural networks for small-footprint keyword spotting. *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 947–951.
- [86] Hsu, C. W., & Lin, C. J. (2002). A Comparison of Methods for Multiclass Support Vector Machines. *IEEE Transactions on Neural Networks*, 13(2), 415–425.
- [87] Zhang, H., & Wang, Y. (2015). A Comparative Study of Support Vector Machines and Neural Networks for Classification. *Journal of Machine Learning Research*, 16, 1-20.
- [88] Cortes, C., & Vapnik, V. (1995). Support-Vector Networks. *Machine Learning*, 20(3), 273–297.

ANNEXE : LES PROGRAMMES COMPLETS

SYSTÈME DE RECONNAISSANCE FACIALE EN CAS DE DANGER

```
# -*- coding: utf-8 -*-
```

```
"""
```

```
SYSTÈME DE CLASSIFICATION DES ÉMOTIONS FACIALES
```

```
Technologies utilisées :
```

- OpenCV pour le traitement d'images
- Scikit-learn pour les SVM
- Matplotlib pour la visualisation

```
"""
```

```
import cv2
```

```
import numpy as np
```

```
import os
```

```
from skimage import feature
```

```
from glob import glob
```

```
from sklearn.svm import SVC
```

```
from sklearn.metrics import accuracy_score, precision_score, recall_score, classification_report
```

```
import joblib
```

```
import matplotlib.pyplot as plt
```

```
# Configuration des classes émotionnelles
```

```
CLASSES = ['En Colère', 'Dégout', 'Peur', 'Triste', 'Malade']
```

```
TRAIN_DIR = r"R:\base de données\Visage\Train"
```

```
TEST_DIR = r"R:\base de données\Visage\Test"
```

```
class FeatureVisualizer:
```

```
    def __init__(self):
```

```
        self.plotted = False
```

```
    def plot_combined_histograms(self, image, lbp_hist, title):
```

```
        """Affiche les histogrammes RGB et LBP côte à côte"""
```

```
        plt.figure(figsize=(18, 6))
```

```
        # Histogramme RGB
```

```
        plt.subplot(1, 2, 1)
```

```
        colors = ('b', 'g', 'r')
```

```
        for i, color in enumerate(colors):
```

```
            hist = cv2.calcHist([image], [i], None, [256], [0, 256])
```

```
            hist = cv2.normalize(hist, hist).flatten()
```

```
            plt.plot(hist, color=color, label=f'{color.upper()} channel')
```

```
        plt.title('Histogramme RGB')
```

```
        plt.xlabel('Valeur Pixel')
```

```
        plt.ylabel('Fréquence')
```

```
        plt.legend()
```

```
        # Histogramme LBP
```

```
        plt.subplot(1, 2, 2)
```

```

plt.bar(range(len(lbp_hist)), lbp_hist, color='teal', alpha=0.7)

plt.title('Histogramme LBP (59 bins)')

plt.xlabel('Motif LBP')

plt.ylabel('Fréquence')

plt.suptitle(title)

plt.tight_layout()

plt.show()

```

```

def robust_imread(image_path):
    """Lecture d'images par le programme"""
    try:
        img = cv2.imread(image_path)
        if img is not None:
            return img

        with open(image_path, 'rb') as f:
            bytes = bytearray(f.read())
            numpyarray = np.asarray(bytes, dtype=np.uint8)
            return cv2.imdecode(numpyarray, cv2.IMREAD_COLOR)
    except:
        return None

```

```

def extract_features(image_path, visualizer, plot_sample=True):
    """Extraction des caractéristiques avec visualisation optionnelle"""
    image = robust_imread(image_path)

```

```

if image is None:
    return None

# Histogramme RGB (normalisation et affichage)
hist_rgb = cv2.calcHist([image], [0,1,2], None, [8,8,8], [0,256,0,256,0,256])
hist_rgb = cv2.normalize(hist_rgb, hist_rgb).flatten()

# Caractéristiques LBP (normalisation et affichage)
gray = cv2.cvtColor(image, cv2.COLOR_BGR2GRAY)
lbp = feature.local_binary_pattern(gray, P=24, R=3, method='uniform')
lbp_hist, _ = np.histogram(lbp, bins=59, range=(0,59))
lbp_hist = lbp_hist.astype("float") / (lbp_hist.sum() + 1e-6)

if plot_sample and not visualizer.plotted:
    visualizer.plot_combined_histograms(image, lbp_hist, os.path.basename(image_path))
    visualizer.plotted = True

return np.hstack((hist_rgb, lbp_hist))

def build_silent_dataset(directory, class_labels, visualizer):
    """ Construit un jeu de données silencieux à partir de notre images. """
    features = []
    labels = []

    for class_idx, class_name in enumerate(CLASSES):
        class_dir = os.path.join(directory, class_name)

```

```

if not os.path.exists(class_dir):
    continue

valid_images = []
for ext in ['*.jpg', '*.jpeg', '*.png']:
    for img_path in glob(os.path.join(class_dir, ext)):
        if os.path.getsize(img_path) > 0:
            valid_images.append(img_path)

for img_path in valid_images:
    feat = extract_features(img_path, visualizer, plot_sample=(len(features)==0))
    if feat is not None:
        features.append(feat)
        labels.append(class_labels[class_idx])

return np.array(features), np.array(labels)

def main():
    visualizer = FeatureVisualizer()
    class_labels = [1, 2, 3, 4, 5]

    print("Construction du jeu d'entraînement (apprentissage)...")
    X_train, y_train = build_silent_dataset(TRAIN_DIR, class_labels, visualizer)

    print(f"{len(X_train)} échantillons chargés")

```

```

print("Entraînement des modèles...")

""" Initialisation des modèles SVM avec différents noyaux (RBF, Polynomial, Sigmoid)
pour comparer leur performance sur la classification des signaux silencieux """

models = {
    'RBF': SVC(kernel='rbf', probability=True),
    'Polynomial': SVC(kernel='poly', degree=3, probability=True),
    'Sigmoid': SVC(kernel='sigmoid', probability=True)
}

for kernel_name, model in models.items():
    print(f"\nApprentissage avec noyau {kernel_name}...")
    model.fit(X_train, y_train)

print("\nConstruction du jeu de test...")
X_test, y_test = build_silent_dataset(TEST_DIR, class_labels, visualizer)

if len(X_test) > 0:
    print("\n Résultats par classe:")
    for kernel_name, model in models.items():
        print(f"\n Noyau: {kernel_name}")
        y_pred = model.predict(X_test)

        # Résultats par classe
        for class_idx, class_name in enumerate(CLASSES):
            mask = (y_test == class_idx+1)

```

```

if np.sum(mask) > 0:
    acc = accuracy_score(y_test[mask], y_pred[mask])
    prec = precision_score(y_test, y_pred, labels=[class_idx+1], average='micro')
    rec = recall_score(y_test, y_pred, labels=[class_idx+1], average='micro')

    print(f"\n{class_name}:")
    print(f" Taux de Reconnaissance: {acc:.2%}")
    print(f" Précision: {prec:.2%}")
    print(f" Rappel: {rec:.2%}")

# Résultats globaux
print("\n Résultats globaux:")
print(f"Taux de Reconnaissance: {accuracy_score(y_test, y_pred):.2%}")
print(f"Précision Moyenne: {precision_score(y_test, y_pred, average='weighted'):.2%}")
print(f"Rappel Moyen: {recall_score(y_test, y_pred, average='weighted'):.2%}")
print("\n" + "="*50)

else:
    print("Aucune image de test trouvée!")

if __name__ == "__main__":
    main()

```

SYSTÈME DE RECONNAISSANCE VOCALE

```
# -*- coding: utf-8 -*-
```

```
"""
```

SYSTÈME DE CLASSIFICATION DES ÉMOTIONS VOCALES

Technologies utilisées :

- Librosa pour le traitement audio
- Scikit-learn pour les SVM
- Soundfile pour la lecture des fichiers audio

```
"""
```

```
import os
```

```
import librosa
```

```
import numpy as np
```

```
from sklearn import svm
```

```
from sklearn.metrics import precision_score, recall_score, accuracy_score
```

```
from sklearn.preprocessing import StandardScaler
```

```
from sklearn.model_selection import train_test_split
```

```
import soundfile as sf
```

```
# Configuration du système
```

```
data_path = r"R:\base de données\Audio"
```

```
classes = ["Malade", "Peur", "Secours", "Mourir", "Danger"]
```

```
sample_rate = 22050
```

```
pre_emphasis = 0.97
```

```

clip_length = 1.5

n_mfcc = 13

def preprocess_audio(file_path):
    """Charge et prétraite l'audio (normalisation, pré-accentuation, ajustement de longueur)."""
    audio, sr = librosa.load(file_path, sr=sample_rate, mono=True)

    peak = np.max(np.abs(audio))
    if peak > 0:
        audio = audio * (0.99 / peak)

    audio = np.append(audio[0], audio[1:] - pre_emphasis * audio[:-1])
    audio = librosa.effects.preemphasis(audio, coef=pre_emphasis)

    target_length = int(sample_rate * clip_length)
    if len(audio) > target_length:
        start = (len(audio) - target_length) // 2
        audio = audio[start:start+target_length]
    else:
        padding = target_length - len(audio)
        left_pad = padding // 2
        right_pad = padding - left_pad
        audio = np.pad(audio, (left_pad, right_pad), mode='constant')

    return audio

```

```

def extract_enhanced_mfcc(audio):
    """Extrait les MFCC avec dérivées temporelles (delta)."""
    harmonic = librosa.effects.harmonic(audio)

    mfccs = librosa.feature.mfcc(
        y=harmonic,
        sr=sample_rate,
        n_mfcc=n_mfcc,
        n_mels=128,
        fmax=8000,
        n_fft=2048,
        hop_length=512
    )

    delta = librosa.feature.delta(mfccs)
    delta2 = librosa.feature.delta(mfccs, order=2)

    stacked = np.vstack((mfccs, delta, delta2))
    return np.concatenate((np.mean(stacked, axis=1), np.std(stacked, axis=1)))

# Chargement des données par classe
features = []
labels = []
for label_idx, class_name in enumerate(classes):
    class_path = os.path.join(data_path, class_name)
    class_features = []

```

```

for filename in os.listdir(class_path):
    if filename.endswith(".wav"):
        file_path = os.path.join(class_path, filename)
        try:
            audio = preprocess_audio(file_path)
            features.append(extract_enhanced_mfcc(audio))
            labels.append(label_idx)
        except Exception as e:
            print(f"Ignoré {filename}: {str(e)}")

X = np.array(features)
y = np.array(labels)

# Division entre l'entraînement et le teste
X_train, X_test, y_train, y_test = train_test_split(
    X, y, test_size=0.25, random_state=42, stratify=y
)

# Normalisation des caractéristiques
scaler = StandardScaler()
X_train = scaler.fit_transform(X_train)
X_test = scaler.transform(X_test)

# Entraînement du modèle SVM
model = svm.SVC(

```

```

kernel='rbf',

C=1.5,

gamma='scale',

class_weight='balanced',

probability=True
)

model.fit(X_train, y_train)

# Évaluation

y_pred = model.predict(X_test)

# Affichage des résultats par classe

print("\n" + "="*50)

print("=== RÉSULTATS PAR CLASSE ===")

for label_idx, class_name in enumerate(classes):

    mask = (y_test == label_idx)

    class_accuracy = accuracy_score(y_test[mask], y_pred[mask])

    class_precision = precision_score(y_test, y_pred, average=None, labels=[label_idx])[0]

    class_recall = recall_score(y_test, y_pred, average=None, labels=[label_idx])[0]

    print(f"\n=== {class_name.upper()} ===")

    print(f"Taux de reconnaissance: {class_accuracy*100:.2f}%")

    print(f"Précision:          {class_precision*100:.2f}%")

    print(f"Rappel:                {class_recall*100:.2f}%")

# Calcul des métriques globales

```

```
global_accuracy = accuracy_score(y_test, y_pred)
global_precision = precision_score(y_test, y_pred, average='weighted')
global_recall = recall_score(y_test, y_pred, average='weighted')

# Affichage des résultats globaux
print("\n" + "="*50)
print("=== RÉSULTATS GLOBAUX ===")
print(f"Taux de reconnaissance global: {global_accuracy*100:.2f}%")
print(f"Précision globale:          {global_precision*100:.2f}%")
print(f"Rappel global:                {global_recall*100:.2f}%")
print("="*50)
```

الجمهورية الجزائرية الديمقراطية الشعبية
وزارة التعليم العالي والبحث العلمي

جامعة عبد الحميد بن باديس-مستغانم-
كلية العلوم و التكنولوجيا

تصريح شرفي خاص بالالتزام بقواعد النزاهة العلمية لإنجاز البحث

أنا الممضي أدناه،

الطالب(ة): ..بريك أيمن ياسين. رقم التسجيل الجامعي :.....202037028270.....

الحامل لبطاقة التعريف الوطنية رقم: 413611396 والصادرة بتاريخ:01/12/2024.....

عن.....مستغانم.....

المسجل بكلية العلوم و التكنولوجيا/ قسمالهندسة الكهربائية.....

شعبة/الكرونيك..... التخصص.. Systèmes des télécommunications.....

والمكلف بإنجاز مذكرة ماستر بعنوان:

.Le secourisme des personnes dans les bâtiments intelligents par la classification d'images et de la parole

أصرح بشرفي أنني ألتم بمراجعة المعايير العلمية والمنهجية ومعايير الأخلاقيات العلمية والنزاهة الأكاديمية المطلوبة في إنجاز البحث، وأتحمل المسؤولية الشخصية عن كل المحتوى المتضمن في البحث المذكور أعلاه .

التاريخ: 24/06/2025

إمضاء المعني


Braik Aymen

* ملحق القرار الوزاري رقم 933 المؤرخ في 28 جويلية 2016 الذي يحدد القواعد المتعلقة بالوقاية من السرقة العلمية ومكافحتها.

الجمهورية الجزائرية الديمقراطية الشعبية
وزارة التعليم العالي والبحث العلمي

جامعة عبد الحميد بن باديس - مستغانم -
كلية العلوم و التكنولوجيا

تصريح شرفي خاص بالالتزام بقواعد النزاهة العلمية لإنجاز البحث

أنا الممضي أدناه،

الطالب (ة): السيد/ السيدة رقم التسجيل الجامعي: 2020.370.296.73
الحامل لبطاقة التعريف الوطنية رقم: 887984342 والصادرة بتاريخ: 2020.06.17
عن السيد/ السيدة
المسجل بكلية العلوم و التكنولوجيا / قسم الهندسة الكهربائية

شعبة / التخصص
Electronique des Systemes embarques
والمكلف بإنجاز مذكرة ماستر بعنوان:

Le... maintenance... des... dans... les... intelligent... par...
la... classification... d'image... et... de... la... parole

أصرح بشرفي أنني ألتزم بمراعاة المعايير العلمية والمنهجية ومعايير الأخلاقيات العلمية والنزاهة الأكاديمية
المطلوبة في إنجاز البحث، وأتحمل المسؤولية الشخصية عن كل المحتوى المتضمن في البحث المذكور أعلاه .

التاريخ: 2025.06.24

إمضاء المعني


* ملحق القرار الوزاري رقم 933 المؤرخ في 28 جويلية 2016 الذي يحدد القواعد المتعلقة بالوقاية من السرقة العلمية ومكافحتها.

