

République Algérienne Démocratique et Populaire
Ministère de l'enseignement Supérieur et de la Recherche Scientifique
Université de Mohamed El Bachir El Ibrahimi de Bordj Bou Arréridj
Faculté des Mathématiques et d'Informatique
Département d'informatique



MEMOIRE

Présenté en vue de l'obtention du diplôme
Master en informatique
Spécialité : Ingénierie d'Informatique Décisionnel

THEME

L'apprentissage profond pour la reconnaissance des
macro-expressions

Présenté par :

Merrouche Said

Ben Merrouche Imad el hak

Soutenu publiquement le : 22/06/2024

Devant le jury composé de :

Président : M. Mohdeb Djamila

Examineur : M. Zouaoui Hakima

Encadreur : M. Meryem Regouid

2023/2024

People's Democratic Republic of Algeria
Ministry of Higher Education and Scientific Research
Mohamed El Bachir El Ibrahimi University of Bordj Bou Arréridj
Faculty of Mathematics and Computer Science
Department of Computer Science



THESIS

Presented for the attainment of the degree
Master in Computer Science
Specialty : Decision Engineering in Computer Science

TOPIC

Deep Learning and Macro Expression Recognition

Presented by :

Merrouche Said

Ben Merrouche Imad el hak

Defended publicly on : 22/06/2024

In front of the jury composed of :

President : M. Mohdeb Djamila

Examiner : M. Zouaoui Hakima

Supervisor : M. Meryem Regouid

2023/2024

Dédicace

Je dédie ce travail

A Mes chers parents, pour tous leurs sacrifices, leur amour, leur tendresse et leurs

prières tout au long de mes études .

À mon frère Mohammed.

A mon compagnon Imad el hak et sa famille.

A mes amis... merci beaucoup pour votre aide Merci pour toutes les valeurs de

fraternité et pour tous les moments passés ensemble.

Et à toute ma famille et à tous ceux que j'aime.

Dédicace

Je dédie ce travail

A Mes chers parents, pour tous leurs sacrifices, leur amour, leur tendresse et leurs prières tout au long de mes études .

A mes frères et à ma sœur.

A mon compagnon said et sa famille.

A mes amis... merci beaucoup pour votre aide Merci pour toutes les valeurs de fraternité et pour tous les moments passés ensemble.

Et à toute ma famille et à tous ceux que j'aime.

Remerciement

Merci, Allah, de nous avoir dotés de la capacité d'écrire et de réfléchir, de la force de croire, de la patience pour réaliser nos rêves, et de la joie de lever les mains vers le ciel en disant "Alhamdulillah". En premier lieu, ce travail ne serait pas aussi enrichissant sans l'aide et l'encadrement exceptionnels de Madame **Regouid Meryem**. Nous la remercions chaleureusement pour sa qualité d'encadrement, Nous exprimons également notre gratitude à sa patience et sa disponibilité tout au long de notre préparation de ce mémoire.

Toutes mes gratitudees s'adressent aussi aux membres de jury : Madame **Zouaoui Hakima**
Qui m'a fait l'honneur d'évaluer ce travail et de présider le jury : Madame **Mohdeb Djamil**
Pour avoir accepté d'examiner ce mémoire. Je leurs adresse mes respectueux remerciements.

Nous adressons aussi nos vifs remerciements à tous nos professeurs, nos enseignants et toutes les personnes qui ont contribué à notre formation tout au long de notre chemin d'étude.

Merci à tous ...

Résumé

La reconnaissance des émotions humaines, notamment à travers les expressions faciales, a attiré beaucoup d'attention en recherche récemment. Des techniques avancées d'apprentissage profond et d'apprentissage automatique ont été utilisées pour analyser la base de données CK+, afin de mieux comprendre et identifier les émotions.

Au cours de nos expériences, nous avons exploré deux méthodes principales pour détecter les émotions. La première méthode reposait sur des techniques d'apprentissage automatique en utilisant des algorithmes tels que les k-plus proches voisins (K-NN) et les machines à vecteurs de support (SVM). La seconde méthode s'appuyait sur l'apprentissage profond en utilisant des réseaux de neurones convolutionnels (CNN) et (DenseNet). Cette comparaison nous a permis d'évaluer l'efficacité des approches traditionnelles par rapport aux techniques modernes dans le domaine de la reconnaissance des émotions, nous offrant des perspectives approfondies sur les performances relatives de chacune.

Mots clés : biométrie, réseau de neurones convolutifs (CNN), reconnaissance des émotions faciales (REF), Machines à vecteurs de support (SVM), K-plus proche voisin (KNN), réseau convolutionnel densément connecté (DenseNet).

Abstract

Recognition of human emotions, particularly through facial expressions, has recently garnered a lot of research attention. Advanced deep learning and machine learning techniques have been employed to analyze the CK+ database in order to better understand and identify emotions.

In our experiments, we explored two primary methods for emotion detection. The first method involved machine learning techniques using algorithms such as k-nearest neighbors (K-NN) and support vector machines (SVM). The second method relied on deep learning using convolutional neural networks (CNN) and (DenseNet). This comparison allowed us to evaluate the effectiveness of traditional approaches versus modern techniques in the field of emotion recognition, providing us with deep insights into the relative performance of each.

Keywords : biometrics ,convolutional neural network (CNN), facial emotion recognition (FER), support vector machine(SVM), k-nearest neighbors(KNN), densely connected convolutional neural networks(DenseNet).

المخلص

التعرف على المشاعر الإنسانية، خاصةً من خلال تعابير الوجه، جذب الكثير من الاهتمام في الأبحاث مؤخرًا. تم استخدام تقنيات التعلم العميق والتعلم الآلي المتقدمة لتحليل قاعدة البيانات CK+ من أجل فهم المشاعر وتحديدتها بشكل أفضل.

استعرضنا في تجاربنا طريقتين رئيسيتين للكشف عن العواطف. الطريقة الأولى اعتمدت على تقنيات التعلم الآلي باستخدام خوارزميات الجيران الأقرب (K-NN) K وألات المتجهات الداعمة (SVM). بينما استندت الطريقة الثانية إلى التعلم العميق باستخدام الشبكات العصبية التلافيفية (CNN) و DenseNet. سمحت لنا هذه المقارنة بتقييم فعالية الطرق التقليدية مقارنة بالتقنيات الحديثة في مجال التعرف على العواطف، مما وفر لنا رؤى عميقة حول الأداء النسبي لكل منها.

الكلمات المفتاحية : القياسات الحيوية، الشبكة العصبية التلافيفية (CNN)، التعرف على مشاعر الوجه (FER)، آلة المتجهات الداعمة (SVM)، ك أقرب جار (KNN)، الشبكات العصبية التلافيفية المتصلة بكثافة (DenseNet).

Table des matières

Liste des Abréviations	xii
Table des figures	xiv
Liste des tableaux	xvi
Introduction Générale	1
Chapitre 1 : La Biometrie et le Système de Reconnaissance de Visage	
1.1 Introduction	3
1.2 Définition de la biométrie	3
1.3 Architecture d'un systèmes biométriques	4
1.4 La reconnaissance biométrique	5
1.5 Types de modalités biométriques	6
1.6 Identification et vérification	7
1.6.1 Application d'identification	7
1.6.2 Application d'authentification / vérification	8
1.7 Mesure de la performance d'un système biométrique	8
1.8 Conclusion	10
Chapitre 2 : Les Émotions Faciales	
2.1 Introduction	11

2.2 La reconnaissance des émotions faciales (Macro-expression)	11
2.3 Anatomie du visage	11
2.4 Les avantages du macro-expression	13
2.5 Les domaines d'application	13
2.6 Travaux connexes	15
2.6.1 Méthodes traditionnelles	15
2.6.2 L'apprentissage profond (deep learning)	16
2.7 Les bases de données des expressions faciales	19
2.7.1 MMI	19
2.7.2. CK+	20
2.7.3 FER 2013	21
2.8 Conclusion	22
Chapitre 3 : Approche Proposée	
3.1 Introduction	23
3.2 L'architecture du système de reconnaissance de macro-expression proposé	23
3.2.1 Prétraitement	24
3.2.2 Extraction de caractéristiques	26
3.2.3 Classification	28
3.2.4 L'apprentissage automatique (Machine learning)	28
3.2.4.1 K-plus proches voisins (K-NN)	29
3.2.4.1.1 Définition	29
3.2.4.1.2 Principe de KNN	29
3.2.4.1.3 Distance euclidienne	30
3.2.4.2 Machines à vecteurs supports (SVM)	31
3.2.5 Convolutional neural network (CNN)	32
3.2.5.1 Définition	32

3.2.5.2 Architecture du modèle CNN	32
3.2.6 Réseau Convolutionnel Densément Connecté (DenseNet) .	35
3.2.6.1 Définition	35
3.2.6.2 Architecture du modèle DenseNet	35
3.7 Conclusion	37
Chapitre 4 : Les Résultats Expérimentaux	
4.1 Introduction	38
4.2 Expériences et résultats	38
4.2.1 Exigences matérielles	38
4.2.2 Logiciels requis	38
4.3 Interface graphique du système de macro-expression	41
4.4 Résultats numérique	44
4.4.1 La base de donnée	44
4.4.2 Implémentation de base de données ck+	44
4.4.3 Résultats	45
4.4.4 Discussion	50
4.5 Conclusion	51
Conclusion Générale	52
Bibliographie	57

Liste des Abréviations

ADN	Acide Désoxyribonucléique
CNN	Convolutional Neural Network
CRNN	Convolutional Recurrent Neural Network
CK +	Cohn-Kanade étendu
DenseNet	Densely Connected Convolutional Neural Networks
EER	Taux d'Erreur Égal (Equal Error Rate)
FRR	Taux de Faux Rejets (False Reject Rate)
FAR	Taux de Fausses Acceptations(False Accept Rate)
FACS	Facial Action Coding System
GAN	Generative Adversarial Network
IA	l'intelligence artificielle
ILSVRC	ImageNet Large Scale Visual Recognition Challenge
ICML	International Conference on Machine Learning
JAFFE	Japanese Female Facial Expression
KNN	k-Nearest Neighbors
LBP	Local Binary Patterns
LSTM	Long Short-Term Memory
MMI	Man-Machine Interaction

REF	Reconnaissance des Émotions Faciales
RNN	Recurrent Neural Network
SVM	Support Vector Machines
UA	Unités d'Action
VGG	Visual Geometry Group

Table des figures

1	Figure 1.1 - Architecture d'un système biométrique	4
2	Figure 1.2 - les modalités biométriques	6
3	Figure 1.3 - Différentes modalités biométriques	7
4	Figure 1.4 - Illustration du FRR et du FAR	9
5	Figure 2.1 - Les muscles du visage	12
6	Figure 2.2 - Exemples d'images statiques d'expressions faciales en vue frontale dans la base de données d'expressions faciales MMI	19
7	Figure 2.3 - Exemples de la base de données CK+	20
8	Figure 2.4 - Exemples d'images FER-2013 pour la reconnaissance des émotions faciales	21
9	Figure 2.5 - L'évolution de la reconnaissance des expressions faciales en termes d'ensembles de données et de méthodes	22
10	Figure 3.1 - L'architecture proposé	24
11	Figure 3.2 - l'application de prétraitement sur une image de la base de données CK+	26
12	Figure 3.3 - la méthode d'extraction des points de repère faciaux(landmarks)	27
13	Figure 3.4 - Deep Learning comme une sorte d'apprentissage automatique.	28

14	Figure 3.5 - K-Plus Proches Voisins (K-NN) Algorithme de classification : applications et exemples concrets	30
15	Figure 3.6 - General classification hyperplane representation of SVM algorithm.	31
16	Figure 3.7 - Détail architectural de l'architecture CNN personnalisée utilisée pour CK+.	34
17	Figure 3.8 - L'architecture du réseau DenseNet.	37
18	Figure 4.1 - L'écran d'accueil de notre application	41
19	Figure 4.2 - Interface Reconnaissance des émotions faciales	42
20	Figure 4.3 - L'étape sélectionner image et prétraitement	42
21	Figure 4.4 - Étape d'extraction (Landmarks)	43
22	Figure 4.5 - Fenêtre de sélection le classificateur	43
23	Figure 4.6 - résultats classification par cnn	44
24	Figure 4.7 - diagramme des mesure de performance de nos systèmes	46
25	Figure 4.8 - Évolution de la performance du modèle d'apprentissage profond : Précision (a) et Perte (b) au fil des époques (DenseNet)	47
26	Figure 4.9 - Évolution de la performance du modèle d'apprentissage profond : Précision (a) et Perte (b) au fil des époques (DenseNet)	48
27	Figure 4.10 - Matrice de confusion pour les deux modèles CNN (a) et Dense-Net (b)	49

Liste des tableaux

1	Table 4.1 - Répartition de la base de données CK+ en train et test	45
2	Table 4.2 - Les mesures de performance de nos systèmes . .	46
3	Table 4.3 - Comparaison des performances avec d'autres modèles	50

Introduction Générale

L'interaction émotionnelle est un phénomène psychologique courant dans notre quotidien. La reconnaissance précise des émotions est cruciale pour assurer une communication, des interactions et des prises de décision efficaces. Avec l'émergence des mégadonnées et de l'intelligence artificielle, la création de systèmes de reconnaissance des émotions est devenue un projet typique, tant dans le domaine académique que dans l'industrie.

Le domaine de la reconnaissance des émotions faciales est une intersection fascinante de la vision par ordinateur, de l'apprentissage automatique et de la psychologie. Avec l'avènement des technologies de deep learning, les capacités des systèmes de reconnaissance des émotions ont considérablement progressé, permettant des applications dans divers domaines tels que la sécurité, les interfaces utilisateur intelligentes, et la recherche en psychologie. Ce mémoire se concentre sur la mise en œuvre d'un système de reconnaissance des émotions faciales en utilisant des méthodes d'apprentissage automatique et d'apprentissage profond.

L'objectif principal de ce travail est de développer un système capable de détecter et d'interpréter les émotions humaines à partir d'images faciales avec une haute précision. Pour ce faire, nous explorons différentes techniques de prétraitement des données, d'extraction de caractéristiques et de classification. Nous comparons les performances des modèles traditionnels comme le K-Nearest Neighbors (K-NN) et les Support Vector Machines (SVM) avec des modèles de deep learning tels que les Convolutional Neural Networks (CNN) et DenseNet.

Ce mémoire est structuré en plusieurs chapitres. Nous commençons au premier chapitre par la biométrie et le système de reconnaissance de visage dans ce chapitre, nous explorons les principes fondamentaux de la biométrie et son application dans les systèmes de reconnaissance de visage. Nous discutons des différentes techniques utilisées pour capturer et analyser les traits faciaux, ainsi que des défis liés à la précision et à la sécurité de ces systèmes.

après les émotions faciales dans ce chapitre , nous explorerons les fondements de la reconnaissance des émotions faciales, en mettant l'accent sur les approches basées sur le deep learning. Nous examinerons également les travaux connexes dans ce domaine et discuterons des avantages et des limites de ces méthodes.

Dans le chapitre trois approche proposé, nous commençons par décrire les méthodes de prétraitement d'images, puis explorons les techniques d'extraction de caractéristiques et aussi les méthodes de classification proposées.

Enfin le dernier chapitre présente les résultats expérimentaux obtenus à partir de nos tests et évaluations. Nous comparons les performances des différentes approches utilisées et nous présentons les discutons des avantages et des limites des différentes approches.

Chapitre 1

La Biometrie et le Système de Reconnaissance de Visage

1.1 Introduction

Le domaine de la biométrie offre des solutions innovantes pour l'identification et la vérification de l'identité des individus, utilisant des caractéristiques biologiques ou physiologiques uniques. Ces modalités biométriques, telles que les empreintes digitales, la reconnaissance faciale, ou la rétine, sont devenues omniprésentes dans notre vie quotidienne, des dispositifs de déverrouillage de smartphones aux systèmes de sécurité haut de gamme. Cependant, chaque modalité biométrique présente ses propres avantages et inconvénients, soulevant des questions cruciales en matière de sécurité, de protection de la vie privée et d'efficacité. Dans ce chapitre, nous explorerons les différentes modalités biométriques, examinant leurs caractéristiques distinctives ainsi que les défis associés à leur mise en œuvre. De plus, nous aborderons les mesures de performance essentielles pour évaluer l'efficacité et la fiabilité des systèmes biométriques.

1.2 Définition de la biométrie

La biométrie consiste à utiliser des mesures des caractéristiques biologiques d'un individu, telles que des traits comportementaux (comme la dynamique de frappe au clavier) ou des traits physiques/physiologiques (comme l'ADN), pour l'identifier ou l'authentifier. Ce concept, également appelé anthropométrie en français, englobe l'étude quantitative des variations biologiques au sein d'un groupe déterminé.

Selon différentes définitions, la biométrie peut être considérée comme la science mathématique des variations biologiques, ou comme toute caractéristique physique automatiquement mesurable, robuste et distinctive pouvant être utilisée pour identifier ou vérifier l'identité d'un individu. Dans le processus de reconnaissance, des paramètres tels que les empreintes, le visage ou l'iris sont extraits de l'image photographiée, puis comparés avec des gabarits préalablement enregistrés.

Les techniques biométriques se distinguent des autres méthodes d'identification en mesurant et reconnaissant ce que l'on est, par opposition à ce que l'on possède (comme un badge ou un document) ou ce que l'on sait (comme un mot de passe). Un système biométrique peut opérer en mode de vérification, confirmant ou niant une identité réclamée, ou en mode d'identification, déterminant l'identité d'un individu. Cette approche est de plus en plus utilisée dans diverses applications pour assurer la reconnaissance précise des personnes[1].

1.3 Architecture d'un systèmes biométrique

Un système biométrique standard se compose de quatre éléments clés :

I. Le module de capture, qui est chargé de recueillir les données biométriques d'un individu. Cela peut être réalisé à l'aide d'un appareil photo, d'un scanner d'empreintes digitales, d'une caméra de surveillance, etc.

II. Le module d'extraction de caractéristiques, qui traite les données biométriques recueillies par le module de capture et extrait uniquement les informations pertinentes pour créer une nouvelle représentation des données. Idéalement, cette représentation devrait être unique pour chaque individu et relativement constante malgré les variations intra-classe.

III. Le module de correspondance, qui compare l'ensemble des caractéristiques extraites avec le modèle stocké dans la base de données du système et évalue le niveau de similitude (ou de divergence) entre les deux.

IV. Le module de décision, qui confirme l'identité revendiquée par un utilisateur ou détermine l'identité d'une personne en fonction du degré de similitude entre les caractéristiques extraites et le(s) modèle(s) enregistré(s).

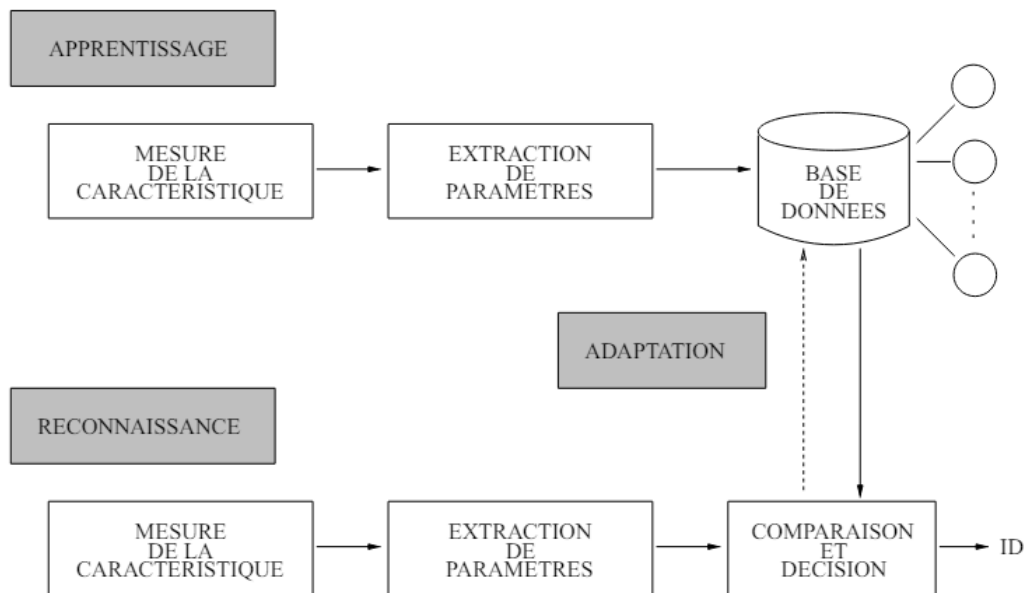


Figure 1.1 - Architecture d'un système biométrique [2].

1.4 La reconnaissance biométrique

La reconnaissance biométrique est une technique utilisée pour identifier ou vérifier l'identité d'une personne en se basant sur des caractéristiques physiques ou comportementales uniques.

Une modalité biométrique n'est rien d'autre qu'une catégorie de système biométrique dépendant du type de trait humain qu'elle prend comme entrée. Voici quelques modalités courantes utilisées pour la reconnaissance biométrique :

» I. Reconnaissance des empreintes digitales :

Cette méthode utilise les caractéristiques uniques présentes dans les empreintes digitales pour identifier une personne. Les scanners d'empreintes digitales captent et analysent les motifs, les crêtes et les vallées des empreintes digitales.

»II. Reconnaissance faciale :

La reconnaissance faciale analyse les caractéristiques du visage d'une personne, telles que la forme des yeux, du nez et de la bouche, ainsi que les distances entre ces points. Les caméras capturent l'image du visage, et des algorithmes spécialisés comparent ces caractéristiques avec celles stockées dans la base de données.

»III. Reconnaissance de la rétine ou de l'iris :

Ces méthodes se basent sur les caractéristiques uniques de la rétine ou de l'iris de l'œil. La reconnaissance de la rétine analyse les motifs de vaisseaux sanguins à l'arrière de l'œil, tandis que la reconnaissance de l'iris examine les caractéristiques colorées et les motifs du cercle coloré entourant la pupille.

»IV. Reconnaissance de la voix :

La reconnaissance vocale se concentre sur les caractéristiques uniques de la voix d'une personne, telles que la fréquence, le timbre et le rythme. Les systèmes de reconnaissance vocale convertissent la voix en un modèle numérique pour l'analyse.

»V. Reconnaissance de la main ou de la géométrie de la main :

Cette méthode utilise les caractéristiques de la main, telles que la forme, la taille et la disposition des doigts. Les scanners de main capturent ces informations pour l'identification. Chaque méthode a ses propres avantages et limitations, et le choix dépend souvent du contexte d'utilisation, de la précision requise et des préférences de sécurité. La reconnaissance biométrique est largement utilisée dans divers domaines, notamment la sécurité, le contrôle d'accès, les transactions financières et les dispositifs électroniques[3].

1.5 Types de modalités biométriques

La biométrie est largement statistique plus les données disponibles à partir de l'échantillon sont nombreuses, plus le système est susceptible d'être unique et fiable. Il peut fonctionner sur diverses modalités relatives aux mesures du corps et des caractéristiques de l'individu, ainsi qu'aux modèles de comportement. Les modalités sont classées en fonction des traits biologiques de la personne[4].

»**Biométrie physique ou morphologique** : Analyse des caractéristiques morphologiques externes du corps, telles que les empreintes digitales, le réseau veineux, la forme de la main, l'iris, la rétine de l'œil et le visage.

»**Biométrie comportementale** : Analyse des caractéristiques liées au comportement d'une personne, comme le tracé d'une signature, la voix, les pulsations cardiaques, la démarche ou la posture.

»**Biométrie biologique** : Analyse des traces ou échantillons biologiques d'une personne, tels que l'ADN, le sang, la salive, l'urine, les odeurs, etc.

⇒ Voici les trois principaux types de biométrie :

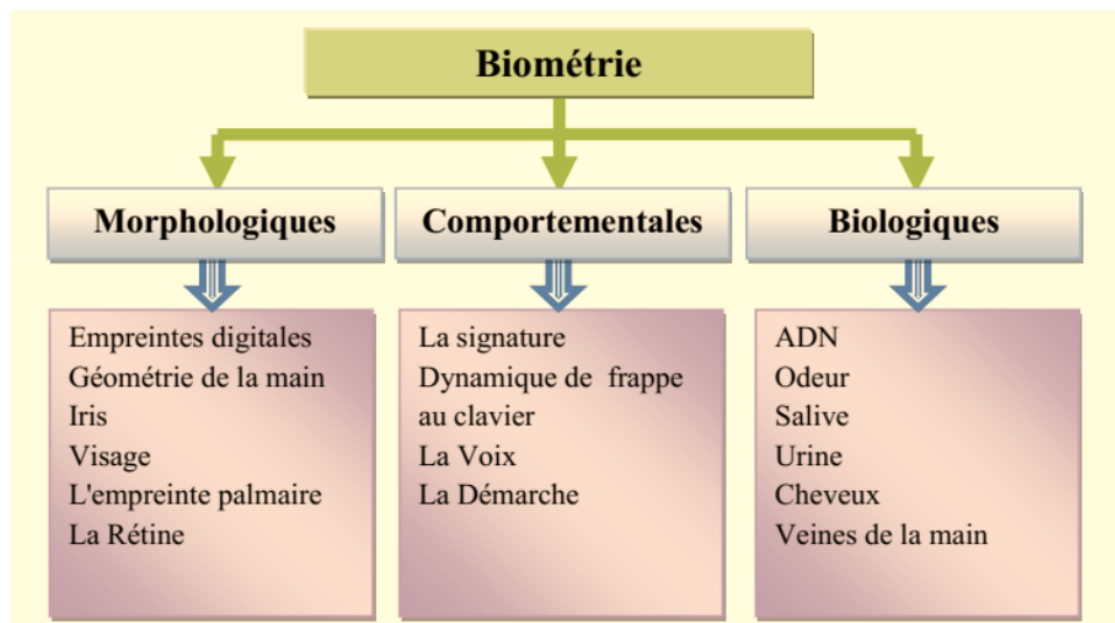


Figure 1.2 - les modalités biométriques [5].

⇒ Et cette image présente différentes modalités biométriques :

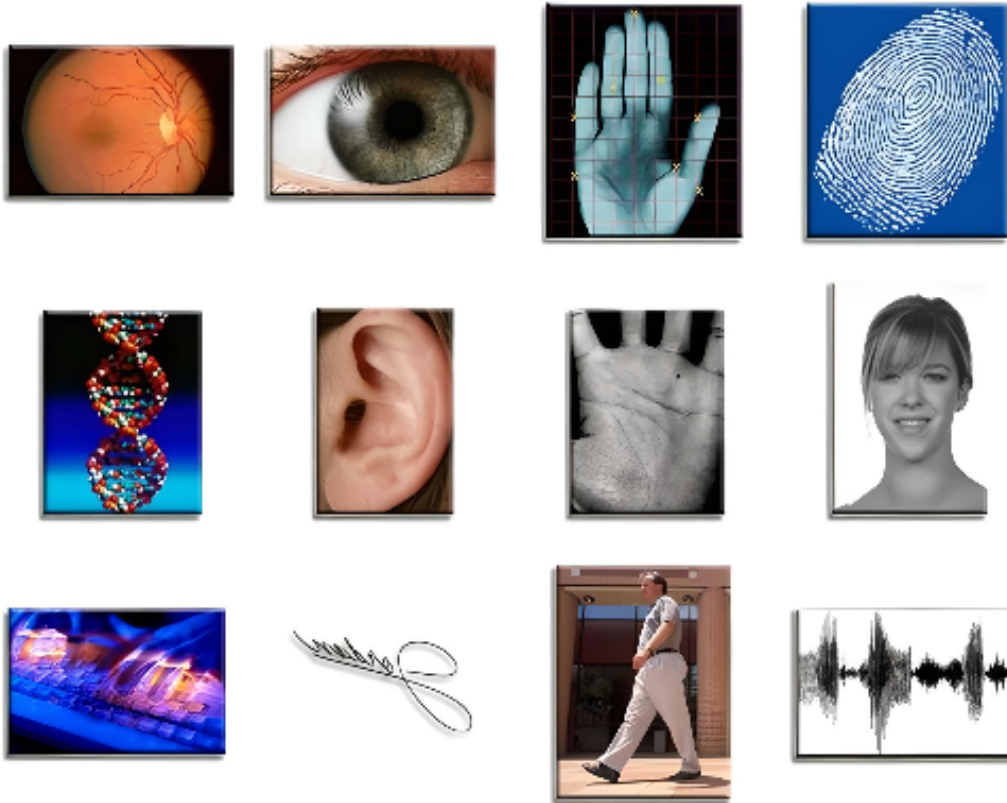


Figure 1.3 - Différentes modalités biométriques [6].

1.6 Identification et vérification

1.6.1 Application d'identification

L'identification d'une personne à plusieurs (1 :N) : La biométrie peut être utilisée pour déterminer l'identité d'une personne, même sans son consentement. Par exemple, la numérisation d'une foule avec une caméra et l'utilisation de la technologie de reconnaissance du visage peuvent contribuer à la détermination du sujet traité, en comparaison avec des profils stockés dans une ou plusieurs bases de données de référence

1.6.2 Application d'authentification / vérification

L'authentification d'une à une personne (1 :1) : dans cette configuration, la biométrie est utilisée pour vérifier l'identité d'une personne. Par exemple, on peut assurer un accès physique à un espace sécurisé dans un bâtiment, par empreinte digitale ou bien on peut garantir l'accès à un compte bancaire ou à un guichet automatique par reconnaissance de l'iris. L'authentification biométrique nécessite de comparer un échantillon biométrique préalablement enregistré (modèle biométrique) à un autre échantillon biométrique nouvellement capturé (par exemple, celui capturé lors d'une connexion) [7].

1.7 Mesure de la performance d'un système biométrique

pour comprendre comment déterminer les performances du système biométrie, nous devons définir clairement quatre critères principaux :

»Le premier critère est appelé

a. Taux de faux rejets ("False Reject Rate" ou FRR) :

Ce taux représente le pourcentage de personnes qui auraient dû être reconnues mais qui l'ont été rejeté par le système.

Il est calculé en divisant le nombre de rejets faux (lorsque des personnes authentiques sont rejetées) par le nombre total de tentatives d'accès authentiques.

$$FRR = \frac{\text{Nombre de rejets faux}}{\text{Nombre total de tentatives d'accès non authentiques}} \quad (1.1)$$

»Le deuxième critère est appelé

b. Taux de fausses acceptations ("False Accept Rate" ou FAR) :

Ce taux représente le pourcentage de personnes qui ne devraient pas être reconnues mais qui sont toujours acceptés par le système.

Il est calculé en divisant le nombre d'acceptations fausses (lorsque des personnes non authentiques sont acceptées) par le nombre total de tentatives d'accès non authentiques.

$$FAR = \frac{\text{Nombre d'acceptations fausses}}{\text{Nombre total de tentatives d'accès non authentiques}} \quad (1.2)$$

»Le troisième critère est connu sous le nom de

c. Taux d'erreur égal ("Equal ErrorRate" ou EER) :

Ce taux est calculé sur la base des deux premiers critères et constitue point commun de la mesure de la performance. Ce point correspond à l'endroit où $FRR = FAR$, qui est le meilleur compromis entre faux rejets et mensonges des soirées[8].

Il est calculé en cherchant le point où FRR et FAR sont ce qui est un bon indicateur de la performance du système.

$$EER = \frac{FAR + FRR}{2} \quad (1.3)$$

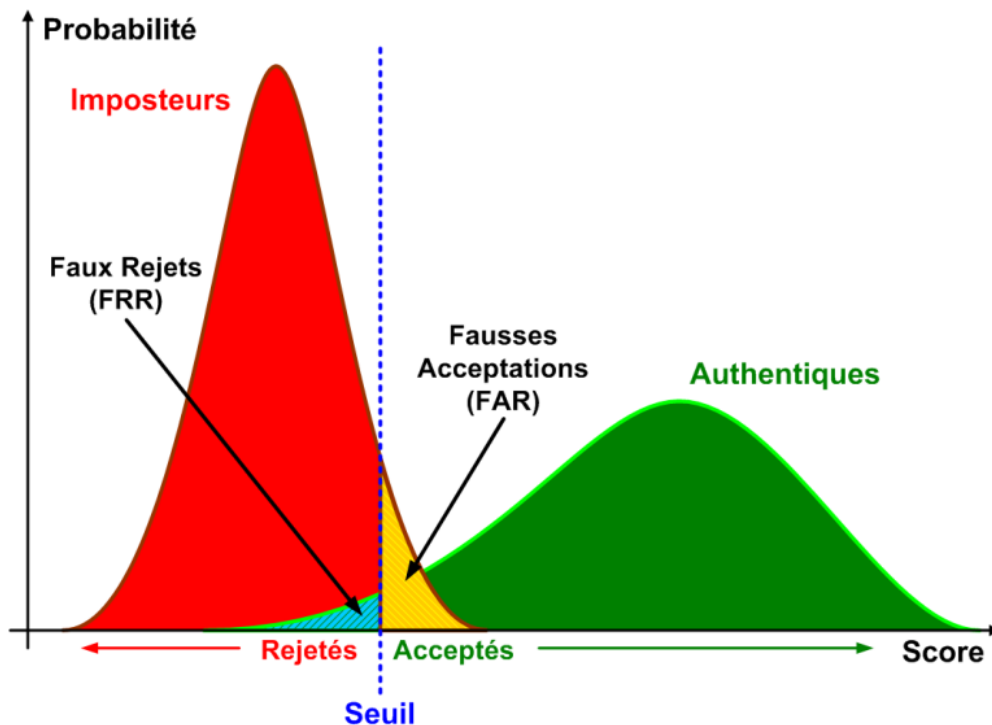


Figure 1.4 - Illustration du FRR et du FAR [6].

d. Taux de reconnaissance correct (accuracy) :

Le taux de reconnaissance correct, ou "accuracy" en anglais, est une mesure de performance utilisée pour évaluer l'efficacité d'un système, notamment dans le domaine de la biométrie. Il représente le pourcentage de décisions correctes faites par le système par rapport à l'ensemble des décisions prises. En d'autres termes, c'est la proportion de prédictions correctes sur l'ensemble des prédictions effectuées par le système. Plus le taux de reconnaissance correct est élevé, plus le système est considéré comme précis et fiable.

La précision (accuracy) est calculée à l'aide de la formule suivante :

$$\text{Accuracy} = \left(\frac{\text{Nombre de Prédictions Correctes}}{\text{Nombre Total de Prédictions}} \right) \times 100\% \quad (1.4)$$

1.8 Conclusion

La biométrie utilise des caractéristiques uniques pour identifier et authentifier les individus. Elle est largement adoptée dans des domaines comme la sécurité et les transactions financières en raison de sa précision. Cependant, une mise en œuvre réussie nécessite une bonne compréhension de ses principes et une évaluation précise de ses performances pour garantir sa fiabilité et sa sécurité. Dans le chapitre suivant, nous présenterons les fondements de la reconnaissance des émotions faciales, en mettant l'accent sur les approches basées sur le deep learning. Nous examinerons les travaux connexes, discuterons des avantages et des limites de ces méthodes, et aborderons les bases de données couramment utilisées dans cette recherche.

Chapitre 2

Les Émotions Faciales

2.1 Introduction

La reconnaissance des émotions faciales est un domaine de recherche passionnant qui vise à détecter et à interpréter les expressions émotionnelles à partir d'images faciales. Avec l'avènement du deep learning, de nouvelles méthodes ont émergé pour améliorer la précision de cette tâche. Dans ce chapitre, nous explorerons les fondements de la reconnaissance des émotions faciales, en mettant l'accent sur les approches basées sur le deep learning. Nous examinerons également les travaux connexes dans ce domaine et discuterons des avantages et des limites de ces méthodes. Enfin, nous aborderons les bases de données couramment utilisées dans la recherche sur la reconnaissance des émotions faciales.

2.2 La reconnaissance des émotions faciales (macro-expression)

La Reconnaissance des Émotions Faciales (REF ou macro-expression) est le processus d'identification et de catégorisation des émotions humaines en se basant sur les expressions faciales. En analysant les caractéristiques et les schémas du visage, les machines peuvent faire des inférences sur l'état émotionnel d'une personne. En général, les individus déduisent les états émotionnels des autres, tels que la joie, la tristesse et la colère, à partir des expressions faciales et du ton vocal [9].

Le REF présente un potentiel académique et commercial considérable, étant largement utilisé dans divers domaines comme la sécurité des systèmes, la sûreté des smartphones, les contrôles douaniers dans les aéroports et les forces de l'ordre. Les chercheurs explorent diverses approches conventionnelles du REF ainsi que des méthodes basées sur l'apprentissage profond, notamment l'utilisation de réseaux de neurones convolutifs (CNN) et de réseaux LSTM (long short-term memory). Comprendre les émotions faciales est crucial pour une communication humaine efficace, ce qui positionne le REF comme un sujet essentiel en vision par ordinateur et en intelligence artificielle[10].

2.3 Anatomie du visage

Dans les échanges en personne, les expressions faciales sont des moyens particulièrement expressifs par lesquels les individus communiquent sans mots. Elles jouent un rôle crucial dans les interactions humaines, influençant la manière dont nous interagissons avec notre environnement et les autres[11]. Les expressions faciales transmettent une gamme variée d'informations, allant de l'état émotionnel, comme la peur, la colère, la joie, la surprise, la tristesse et le dégoût, à des indices sur notre concentration, notre ennui

ou notre perplexité, reflétant ainsi notre activité cognitive. De plus, elles peuvent révéler des traits de personnalité et de tempérament.

Sur le plan physiologique, les expressions faciales impliquent l'action de 44 muscles de chaque côté du visage, également appelés muscles mimétiques, contrôlés par le nerf facial. Ces muscles permettent une grande variété de mouvements perceptibles. Grâce à la complexité de leur agencement, les expressions faciales offrent la possibilité d'exprimer une multitude de sentiments et d'émotions, tandis que la diversité des contractions musculaires permet une variété encore plus grande de modèles de mouvement[12].

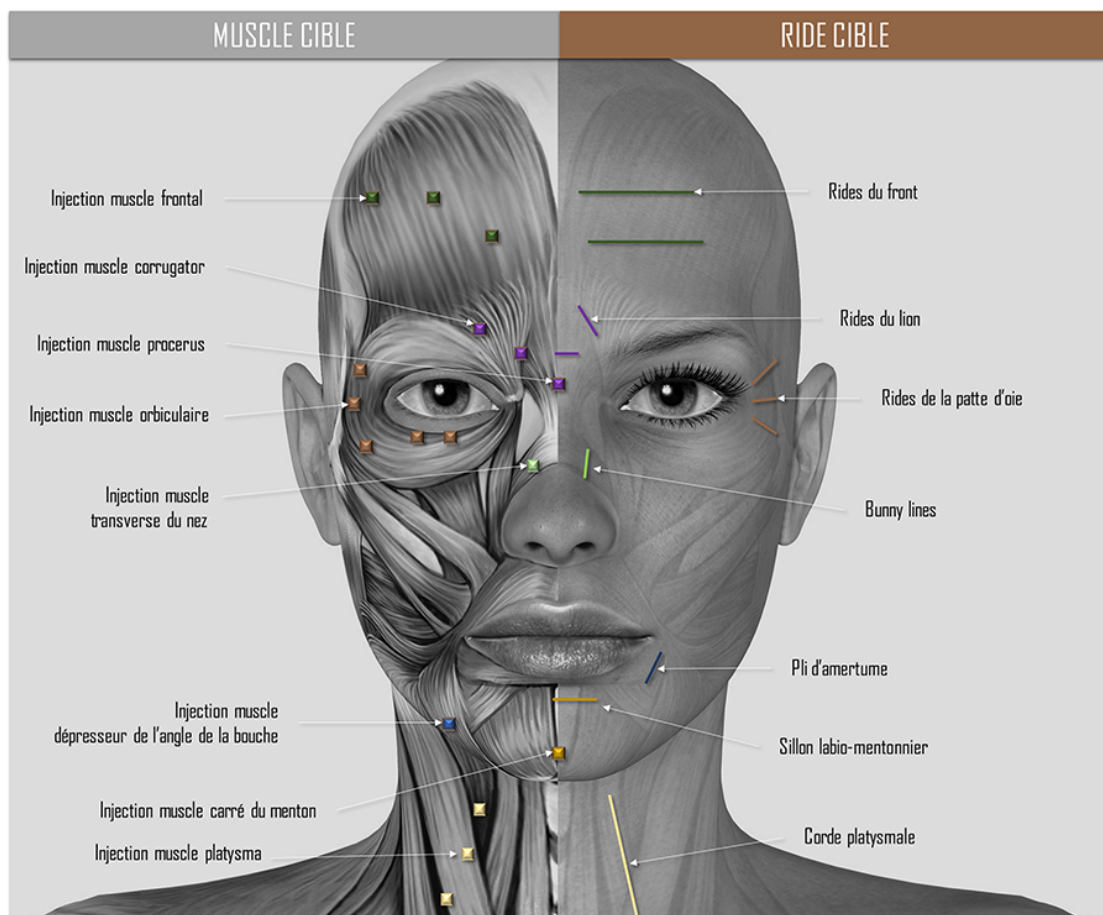


Figure 2.1 - Les muscles du visage [13].

2.4 Les avantages du macro-expression

La reconnaissance des émotions faciales présente des avantages uniques par rapport aux autres méthodes biométriques. Contrairement à d'autres techniques qui se concentrent principalement sur des caractéristiques physiologiques spécifiques, la reconnaissance des émotions faciales exploite les signaux non verbaux universels des expressions faciales, offrant ainsi une compréhension précieuse et en temps réel de l'état émotionnel actuel des individus. De plus, elle peut être réalisée à l'aide de caméras simples, ce qui facilite la collecte des données de manière non intrusive. Cette méthode transcende également les barrières culturelles et linguistiques, étant largement universelle, ce qui la rend adaptable à une utilisation mondiale. Sa polyvalence lui permet d'être appliquée dans une variété d'applications, offrant des réponses rapides et adaptatives dans des domaines tels que les interactions homme-machine, la sécurité et la santé mentale.

2.5 Les domaines d'application

La reconnaissance des émotions faciales a de nombreux domaines d'application, dont voici quelques exemples :

a. Interfaces utilisateur et expérience utilisateur (UX) :

Les applications et les dispositifs peuvent être conçus pour réagir aux émotions des utilisateurs, offrant ainsi une expérience plus personnalisée et immersive.

b. Santé mentale :

La reconnaissance des émotions faciales peut être utilisée pour diagnostiquer et surveiller les troubles mentaux tels que la dépression, l'anxiété et l'autisme, ainsi que pour évaluer l'efficacité des traitements.

c. Éducation :

Les systèmes éducatifs peuvent utiliser la reconnaissance des émotions faciales pour évaluer l'engagement des élèves et adapter les méthodes d'enseignement en conséquence, ainsi que pour fournir un retour d'information aux enseignants.

d. Publicité et marketing :

Les entreprises peuvent utiliser la reconnaissance des émotions faciales pour évaluer les réactions des consommateurs à leurs publicités, produits ou services, afin d'optimiser

leurs campagnes marketing.

e. Sécurité :

Les systèmes de sécurité peuvent utiliser la reconnaissance des émotions faciales pour détecter les comportements suspects ou les situations d'urgence, ainsi que pour contrôler l'accès à des zones sensibles.

f. Soins de santé :

Dans le domaine médical, la reconnaissance des émotions faciales peut être utilisée pour évaluer la douleur chez les patients non verbaux, ainsi que pour aider les thérapeutes à interagir avec les patients et à évaluer leur bien-être émotionnel.

g. Divertissement :

Les jeux vidéo et les applications de réalité virtuelle peuvent utiliser la reconnaissance des émotions faciales pour créer des expériences interactives plus immersives et adaptatives.

h. Recrutement et évaluation du personnel :

Les entreprises peuvent utiliser la reconnaissance des émotions faciales lors des processus de recrutement et d'évaluation du personnel pour évaluer les compétences sociales et émotionnelles des candidats.

i. Automobile :

Dans l'industrie automobile, la reconnaissance des émotions faciales peut être utilisée pour améliorer la sécurité en détectant la somnolence ou la distraction du conducteur, ainsi que pour personnaliser l'expérience de conduite.

j. Robotique :

Les robots sociaux peuvent utiliser la reconnaissance des émotions faciales pour interagir de manière plus naturelle avec les humains, en adaptant leurs réponses en fonction des expressions faciales de leurs interlocuteurs.

Ces domaines d'application montrent la diversité des applications potentielles de la reconnaissance des émotions faciales, qui peut avoir un impact significatif dans de nombreux aspects de la vie quotidienne et des industries.

2.6 Travaux connexes

2.6.1 Méthodes traditionnelles

La reconnaissance des expressions faciales (REF) est un domaine de recherche passionnant dans le domaine de l'intelligence artificielle et du deep learning. Voici un bref aperçu de l'histoire de la recherche sur la reconnaissance des expressions faciales.

Les premières tentatives de reconnaissance des expressions faciales remontent aux années 1967 par Albert Mehrabian [14] le psychologue et professeur de psychologie qui a montré que 55 % des informations émotionnelles étaient visuelles, 38% vocaux et 7% verbaux. Les changements de visage lors d'une communication sont les premiers signes qui transmettent l'état émotionnel, c'est pourquoi la plupart des chercheurs sont très intéressés par cette modalité.

En 1978, Ekman et Friesen ont joué un rôle pionnier dans l'étude de l'expression faciale en développant le facial action coding system (FACS). Ce système, largement utilisé dans la recherche en psychologie et en sciences sociales, a introduit une méthodologie systématique pour décrire les mouvements du visage. Dans le cadre du FACS, les expressions faciales sont analysées et décomposées en 46 Unités d'Action (UA), représentant différents mouvements des muscles faciaux.

Chaque UA est attribuée à un ou plusieurs muscles spécifiques, permettant une description précise et normalisée des expressions faciales. Cette approche a ouvert la voie à une compréhension plus approfondie des expressions faciales et de leur signification émotionnelle, et elle continue d'être une référence importante dans la recherche sur l'expression faciale et la reconnaissance des émotions [15].

Différentes approches traditionnelles sont employées pour extraire les traits du visage, comprenant les caractéristiques géométriques et texturales. Parmi celles-ci, on compte les motifs binaires locaux LBP [16], les unités d'action faciale FAC [17], les motifs directionnels locaux LDA [18], ainsi que les ondelettes de Gabor [19]. Ces dernières années, l'apprentissage profond s'est avéré être une méthode extrêmement efficace et fructueuse, grâce aux performances remarquables obtenues avec des architectures telles que les réseaux neuronaux convolutionnels CNN et les réseaux neuronaux récurrents RNN. Cette réussite a incité les chercheurs à adopter cette approche pour la reconnaissance des émotions humaines.

2.6.2 L'apprentissage profond (deep learning)

L'apprentissage profond, une sous-catégorie de l'apprentissage automatique, a pour objectif d'instruire les ordinateurs pour qu'ils accomplissent des tâches humaines, comme la reconnaissance d'objets visuels et l'identification de sons et d'images. Plutôt que d'organiser les données, l'apprentissage profond établit ses propres paramètres fondamentaux, permettant ainsi à la machine d'apprendre de manière autonome. L'engouement actuel pour l'apprentissage profond est en partie dû à l'excitation autour de l'intelligence artificielle. Les techniques d'apprentissage profond ont amélioré la capacité à classer, reconnaître, détecter, et bien plus encore.

L'apprentissage profond explore la structure complexe dans de vastes ensembles de données en utilisant l'algorithme de rétropropagation pour indiquer comment une machine doit ajuster ses paramètres internes qui sont utilisés pour calculer la représentation de chaque couche à partir de la représentation de la couche précédente. Les réseaux de neurones convolutionnels profonds ont permis des progrès dans le traitement des images, des vidéos, de la parole et de l'audio, tandis que les réseaux de neurones récurrents ont apporté des éclairages sur des données séquentielles telles que le texte et la parole [20].

»Voilà quelque type de l'apprentissage profond :

- **Réseaux de neurones convolutionnels (CNN).**
- **Réseaux de neurones récurrents (RNN).**
- **Réseau Convolutionnel Densément Connecté (DenseNet).**
- **Réseaux de neurones adversariaux génératifs (GAN).**

2.6.2.1 L'apprentissage profond pour la reconnaissance des expressions faciales

Au début des années 2010, l'avènement du deep learning a révolutionné le domaine de la reconnaissance des émotions faciales. Des chercheurs ont commencé à utiliser des réseaux de neurones profonds, tels que les CNN, pour extraire automatiquement des caractéristiques faciales à partir d'images brutes.

Récemment, Christian Szegedy et ses collaborateurs [21] ont présenté une architecture nommée GoogLeNet. Cette structure, comprenant 27 couches, est principalement constituée de réseaux de neurones convolutifs (CNN). Le réseau est formé en utilisant la technique de descente de gradient stochastique. Lors du défi de classification ILSVRC 2014, ce réseau a atteint un taux d'erreur de top-5 de 6,67%, lui permettant de remporter la première place.

En 2016, Mollahosseini et ses collègues ont présenté une architecture de réseau neuronal convolutif (CNN) profond pour la reconnaissance des expressions faciales (REF) à travers plusieurs bases de données disponibles. Leur approche a débuté par l'extraction des points caractéristiques du visage à partir des données, suivie de la réduction des images à une taille de 48x48 pixels. Ensuite, ils ont appliqué des techniques d'augmentation des données pour enrichir l'ensemble d'entraînement.

L'architecture utilisée dans leur étude comprenait deux couches de convolution et de mise en commun, suivies de l'ajout de deux modules de style inception. Ces modules contenaient des couches convolutionnelles de tailles 1x1, 3x3 et 5x5. De plus, ils ont introduit l'utilisation de la technique du réseau dans le réseau (network-in-network), qui permettait d'améliorer les performances locales en utilisant des couches de convolution localement. Cette technique a également contribué à atténuer le problème de surajustement[22].

En 2019, D.Y. Liliana et ses collègues ont proposé une architecture de réseau de neurones convolutifs (CNN) pour la reconnaissance des expressions faciales, visant à identifier huit classes distinctes d'expressions faciales. En utilisant la base de données CK+, ils ont entraîné le réseau avec des tailles de données d'entraînement variées. Les résultats ont montré que l'erreur quadratique moyenne diminuait à mesure que la quantité de données d'entraînement augmentait. Cette expérience a démontré que l'augmentation des données d'entraînement conduit à une réduction de l'erreur quadratique moyenne. De plus, le système a atteint un taux de précision de 92,81 %. Pour les travaux futurs, l'accent sera mis sur le perfectionnement de l'architecture du CNN afin d'obtenir des résultats encore meilleurs[23].

En 2022, Xiaopeng Hong et ses collègues ont présenté une avancée significative dans le domaine de la reconnaissance des émotions faciales en introduisant une architecture innovante de réseau neuronal récurrent convolutif (CRNN). Cette architecture a été spécifiquement conçue pour traiter des données vidéo, ce qui représente une contribution majeure compte tenu de la complexité temporelle inhérente aux séquences vidéo.

L'une des caractéristiques les plus remarquables de l'approche de Hong et al. Est sa capacité à capturer à la fois les informations spatiales et temporelles des séquences vidéo. Contrairement aux méthodes traditionnelles qui se concentrent principalement sur les informations spatiales contenues dans une image statique, le CRNN permet de prendre en compte l'évolution temporelle des expressions faciales. Cette prise en compte de la dynamique temporelle s'avère cruciale pour une reconnaissance précise des émotions, car elle permet de saisir les changements subtils dans les expressions faciales au fil du temps.

En intégrant des couches récurrentes au sein de l'architecture convolutionnelle, le modèle CRNN peut analyser les séquences vidéo frame par frame, en tenant compte des relations temporelles entre les images successives. Cette approche permet non seulement de mieux comprendre l'évolution des expressions faciales, mais aussi de prendre en compte les transitions fluides entre les différentes émotions.

Grâce à cette combinaison de convolutions spatiales et de récurrences temporelles, l'approche de Hong et ses collègues a conduit à des améliorations significatives en termes de précision de la reconnaissance des émotions faciales. Leur méthode a été évaluée sur des ensembles de données vidéo standard et a démontré des performances prometteuses, ouvrant ainsi la voie à de nouvelles applications dans des domaines tels que la surveillance intelligente, l'analyse du comportement humain et les interfaces utilisateur émotionnelles[24].

En 2023, Armanur Rahman et son équipe ont souligné l'importance des expressions faciales dans la communication émotionnelle par des moyens non verbaux. Ces dernières années, la reconnaissance automatique des expressions faciales (REF) a suscité un grand intérêt. Les chercheurs explorent les diverses applications potentielles de cette technologie, notamment dans l'éducation, la surveillance de la sécurité, les systèmes de santé intelligents, ainsi que pour la compréhension des comportements individuels et collectifs. Toutefois, la précision et la fiabilité de la reconnaissance des expressions faciales restent un défi en raison des variations telles que les différentes poses et conditions d'éclairage dans les images. Ils ont ainsi développé une méthode de classification automatique des expressions faciales basée sur le transfert d'apprentissage en profondeur. Cette approche combine des réseaux neuronaux convolutionnels (CNN) avec le modèle VGG19 pour exploiter le transfert d'apprentissage. L'entraînement du modèle a été réalisé en utilisant des techniques de deep learning modernes telles que la recherche du taux d'apprentissage optimal, le réglage fin et l'augmentation des données. Sur les ensembles de données extended cohn-kanade (CK+) et japanese female facial expression (JAFPE), le modèle développé a atteint respectivement des taux de précision de 94,8 % et 93,7 %. Ce système a déjà été testé sur une grande base de données et offre des possibilités d'assistance pour les systèmes d'éducation en ligne, la surveillance et la gestion quotidienne des systèmes de santé intelligents[25].

2.7 Les bases de données des expressions faciales

2.7.1 MMI

MMI [26] ,[27] : La base de données MMI(Man-Machine Interaction), sous contrôle de laboratoire, est constituée de 326 séquences issues de 32 sujets. Parmi celles-ci, 213 séquences sont identifiées avec six expressions de base (sans "dégoût"), tandis que 205 séquences sont enregistrées en vue frontale. Contrairement à CK+, les séquences dans MMI sont annotées avec un début, un pic et une fin, signifiant qu'elles débutent avec une expression neutre, atteignent leur point culminant au milieu, puis reviennent à l'expression neutre. En outre, MMI présente des conditions plus complexes en raison de variations interpersonnelles importantes, les sujets exécutant l'expression de manière non uniforme et beaucoup d'entre eux portant des accessoires (comme des lunettes ou une moustache). Pour les expériences, la méthode la plus utilisée consiste à sélectionner le premier cadre (visage neutre) et les trois cadres de pic dans chaque séquence frontale afin de réaliser une validation croisée à 10 plus indépendante des personnes.



Figure 2.2 - Exemples d'images statiques d'expressions faciales en vue frontale dans la base de données d'expressions faciales MMI [28].

2.7.2 CK+

CK+ : En l'an 2000, la sortie de la base de données Cohn-Kanade (CK) avait pour objectif de catalyser la recherche dans le domaine de la détection automatique des expressions faciales individuelles. Depuis lors, cette base de données est devenue l'une des plates-formes d'essai les plus largement exploitées pour le développement et l'évaluation d'algorithmes[29].

CK+ [30] L'extension de la base de données cohn-kanade (CK+), sous contrôle de laboratoire, est désormais considérée comme l'un des outils les plus utilisés pour évaluer les systèmes de reconnaissance des expressions faciales. Elle comprend 593 séquences vidéo provenant de 123 sujets, variant en durée de 10 à 60 images, montrant la transition d'une expression faciale neutre à une expression maximale. Parmi ces vidéos, 327 séquences provenant de 118 sujets sont étiquetées avec sept expressions faciales de base (colère, mépris, dégoût, peur, bonheur, tristesse et surprise), conformément au facial action coding system (FACS). Étant donné que CK+ ne prévoit pas de base de données d'entraînement, de validation ou de test spécifiques, les algorithmes évalués sur cette base de données ne sont pas uniformisés. Pour les méthodes basées sur des images fixes, la technique de sélection de données la plus courante consiste à extraire les une à trois dernières images où l'expression est à son maximum, ainsi que la première image (expression neutre) de chaque séquence. Ensuite, les sujets sont répartis en n groupes pour des expériences de validation croisée à n plis, indépendamment de la personne, avec des valeurs communes de n.



Figure 2.3 - Exemples de la base de données CK+. [31]

2.7.3 FER 2013

FER2013 : La base de données FER2013 [32] a été présentée lors des ICML 2013 Challenges in Representation Learning. Elle est considérée comme une ressource à grande échelle et sans contrainte, collectée automatiquement via l'API de recherche d'images de google. Après avoir écarté les images incorrectement étiquetées et ajusté la région rognée, toutes les images ont été enregistrées et redimensionnées en 48*48 pixels. FER2013 comprend un ensemble de 28 709 images pour l'entraînement, 3 589 pour la validation et 3 589 pour les tests. Chacune de ces images est associée à l'une des sept expressions émotionnelles suivantes : colère, dégoût, peur, bonheur, tristesse, surprise et neutralité.



Figure 2.4 - Exemples d'images FER-2013 pour la reconnaissance des émotions faciales [33]

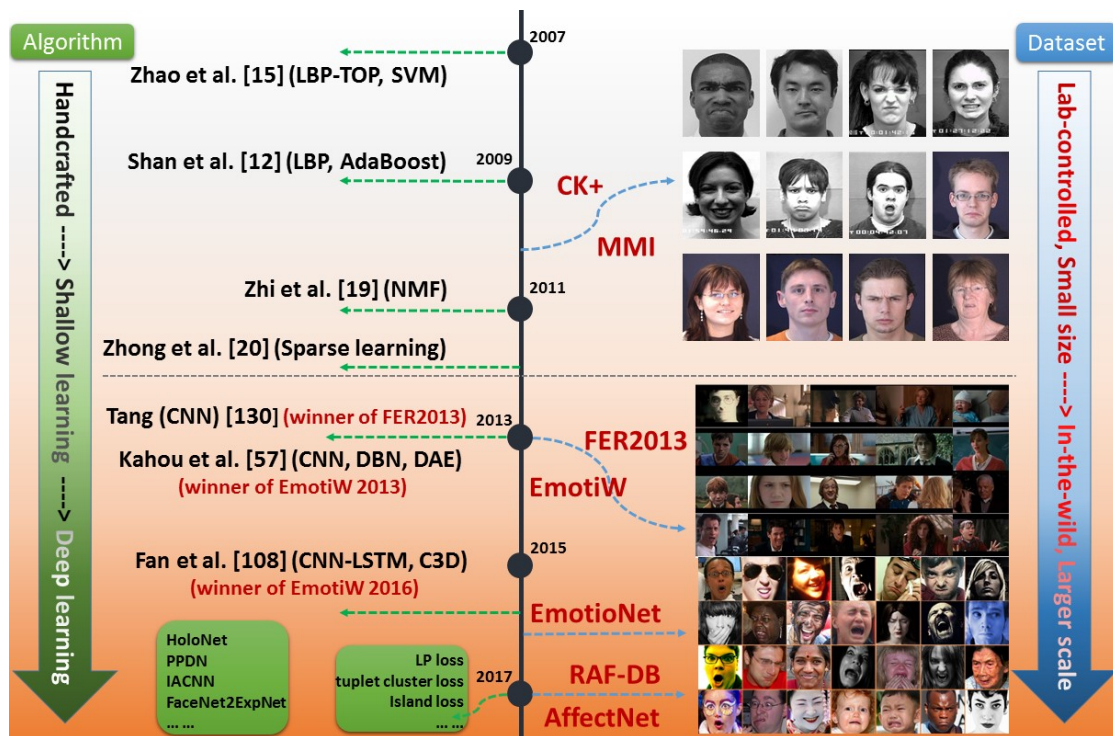


Figure 2.5 - L'évolution de la reconnaissance des expressions faciales en termes d'ensembles de données et de méthodes [34]

2.8 Conclusion

Ce chapitre a introduit les bases de la reconnaissance des émotions faciales avec le deep learning, en définissant clairement le sujet, en explorant les travaux antérieurs et les domaines d'application, en analysant les avantages de la méthode, et en présentant les bases de données couramment utilisées. Il est essentiel de suivre les développements futurs pour maximiser le potentiel de cette technologie dans la compréhension des émotions humaines et les interactions avec les systèmes intelligents. Dans le prochain chapitre, nous aborderons l'architecture détaillée de notre système de reconnaissance de macro-expressions proposé et examinerons également le rôle crucial de l'apprentissage automatique et de l'apprentissage profond dans la construction de modèles capables de reconnaître et d'interpréter les macro-expressions.

Chapitre 3

Approche Proposée

3.1 Introduction

Dans ce chapitre, nous présentons l'architecture détaillée de notre système de reconnaissance de macro-expressions proposé. Nous abordons les étapes allant du prétraitement des données à la classification finale. Nous commençons par décrire les méthodes de prétraitement d'images, puis explorons les techniques d'extraction de caractéristiques, en mettant l'accent sur l'identification des points de repère faciaux. Nous examinons également le rôle crucial de l'apprentissage automatique et l'apprentissage profond dans la construction de modèles capables de reconnaître et d'interpréter les macro-expressions. Enfin, nous passons en revue plusieurs méthodes de classification, allant des approches traditionnelles (SVM ,K-NN) aux modèles d'apprentissage profond (CNN , DenseNet).

3.2 L'architecture du système de reconnaissance de macro-expression proposé

On a propose une architecture du système de reconnaissance de macro-expression avec trois étape. La premiere étape est le prétraitement. Dans cette étape, on a appliqué quatre techniques començant par le redimensionnement des images, l'ajustement des couleurs, le filtrage de netteté et la conversion en niveaux de gris.La deuxième étape est l'extraction des caractéristiques en utilisant la méthode de land marks.Our la troisième étape qui est la classification, deux types d'apprentissage automatique avec svm et knn et l'apprentissage profond avec cnn et dense-net sont appliqués.

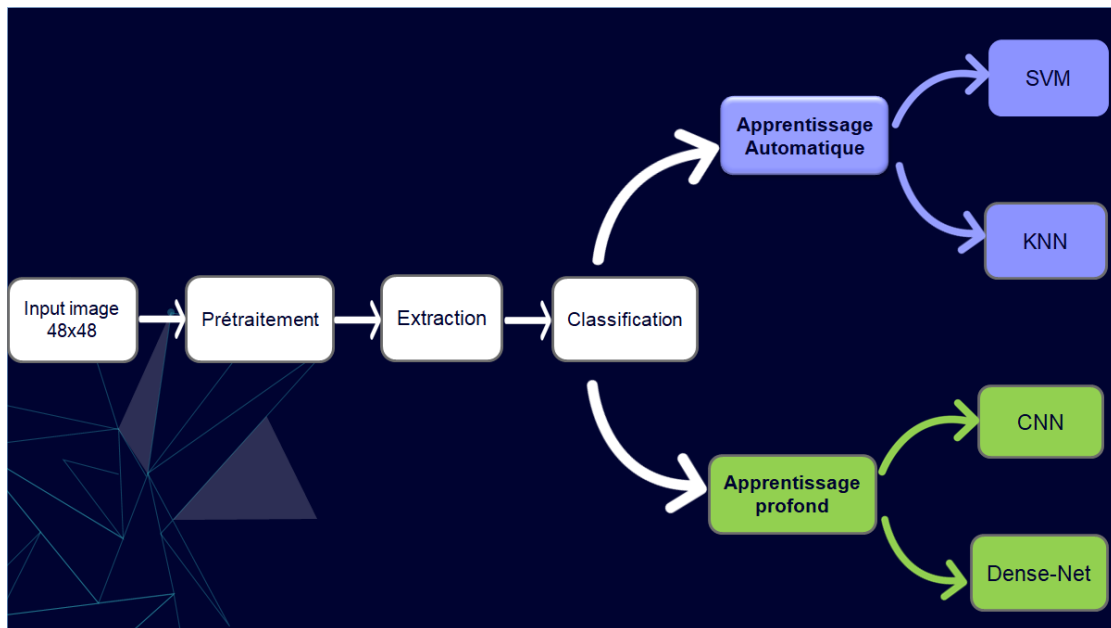


Figure 3.1 - L'architecture proposé

3.2.1 Prétraitement

Cette étape se décompose en quatre sous-étapes distinctes : redimensionnement des images, ajustement des couleurs, filtrage de netteté et conversion en niveaux de gris. Pour chaque image, elle applique les fonctions de prétraitement définies précédemment, puis enregistre l'image prétraitée dans un dossier de sortie spécifique à l'émotion. Chacune de ces sous-étapes est détaillée ci-dessous :

I. Redimensionnement des images :

Redimensionner les images à une taille standardisée de 256x256 pixels est une étape essentielle dans le prétraitement des données pour les modèles de machine Learning et apprentissage profond. Cette normalisation permet de garantir que toutes les images ont les mêmes dimensions, ce qui facilite l'entraînement des modèles et améliore leur performance. La fonction **resize-image** utilisant OpenCV est un outil simple mais puissant pour réaliser cette tâche.

II. ajustement des couleurs :

Cette technique normalise les valeurs de pixel d'une image en les divisant par 300, puis les remet à l'échelle à la plage [0, 255]. Cette transformation peut être utilisée pour ajuster les couleurs de l'image de manière à redistribuer les valeurs de pixel et

potentiellement améliorer la visibilité ou le contraste de l'image, bien que l'échelle de normalisation de 300 soit inhabituelle et peut nécessiter des ajustements supplémentaires pour des applications spécifiques.

III. filtrage de netteté :

Le filtre de netteté (sharpening) présente une approche simple mais efficace pour appliquer un filtre de netteté à une image en utilisant la convolution. Elle utilise un noyau spécifique $[-1, -1, -1; -1, 9, -1; -1, -1, -1]$ pour accentuer les différences d'intensité dans l'image, produisant ainsi un effet de netteté qui améliore les détails et les contours. En utilisant la fonction `cv2.filter2D` de OpenCV avec un argument de profondeur de sortie -1, elle assure que la sortie conserve la même profondeur que l'image d'entrée. Cette fonction peut être utile dans divers contextes, tels que le traitement d'images pour la vision par ordinateur, la reconnaissance d'objets ou la correction de certaines imperfections dans les images. En somme, la combinaison du noyau spécifique et de la méthode de convolution offre une solution pratique pour l'amélioration de la netteté des images dans diverses applications de traitement d'images.

IV. conversion en niveaux de gris :

Median-filter offre une méthode simple et efficace pour réduire le bruit dans une image en niveaux de gris en appliquant un filtre médian, avec le paramètre de taille de noyau **kernel-size** fixé par défaut à 3. En utilisant la fonction `cv2.MedianBlur` de la bibliothèque OpenCV, elle remplace chaque pixel de l'image par la valeur médiane des pixels voisins, ce qui permet de réduire les effets du bruit tout en préservant les contours et les détails importants de l'image.

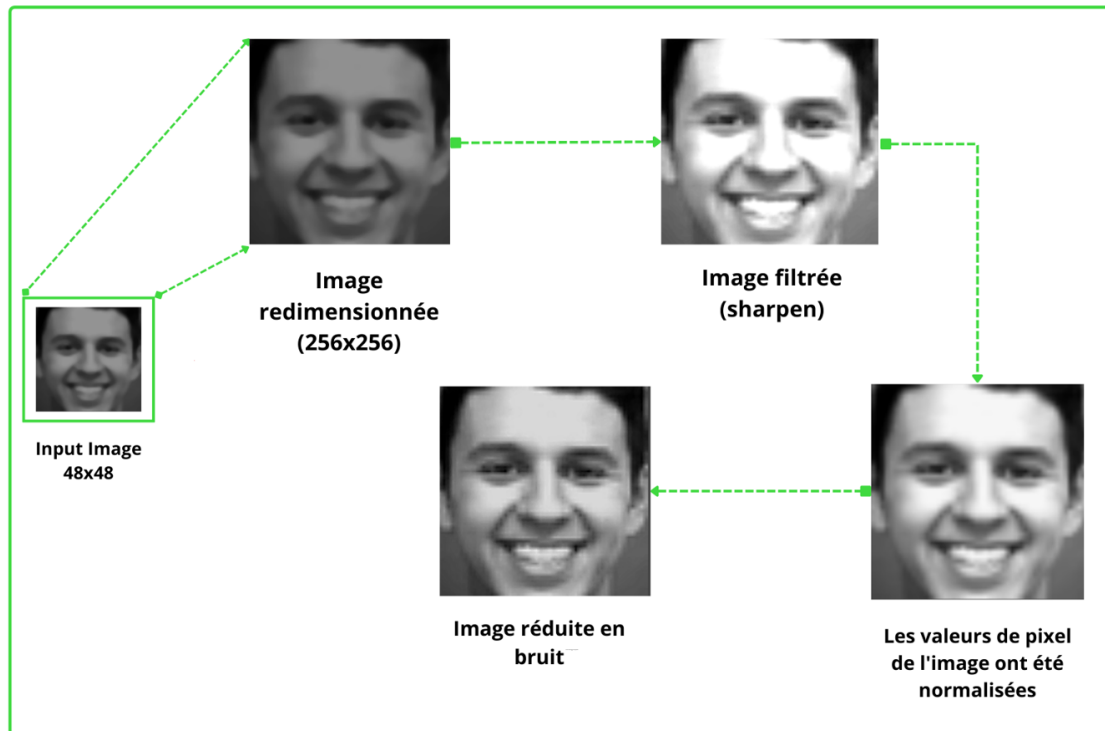


Figure 3.2 - l'application de prétraitement sur une image de la base de données CK+

3.2.2 Extraction de caractéristiques

Pour l'étape d'extraction des caractéristiques, on a appliqué la méthode des points de repère faciaux (landmarks). Elle permet d'extraire les points de repère faciaux à partir d'une image en utilisant des techniques avancées de détection de visage et de prédiction de forme. Cette fonction est cruciale pour de nombreuses applications telles que la reconnaissance faciale, l'analyse d'expression faciale et la réalité augmentée. Ces points de repère faciaux sont enregistrés dans un fichier CSV structuré. Cela facilite l'utilisation ultérieure des données pour diverses tâches d'analyse et de traitement d'images. Cette technique permet de construire des systèmes plus complexes pour des applications avancées telles que l'analyse comportementale, la détection d'émotions en temps réel, ou même des applications interactives basées sur la reconnaissance faciale.

La méthode de fonctionnement de l'extraction des points de repère faciaux se déroule en plusieurs étapes :

I. Initialisation du détecteur de visage et du prédicteur de forme : Tout d'abord, le détecteur de visage et le prédicteur de forme sont initialisés en utilisant le modèle pré-entraîné **shape-predictor-68-face-landmarks**. Le détecteur de visage localise les visages dans une image donnée, tandis que le prédicteur de forme prédit les positions des 68 landmarks faciaux une fois un visage détecté.

II. Détection de visage et prédiction des landmarks : Pour chaque image traitée, le détecteur de visage identifie les visages présents. Ensuite, pour chaque visage détecté, le prédicteur de forme prédit les positions des 68 landmarks faciaux.

III. Extraction des coordonnées des landmarks : Les coordonnées des landmarks prédits sont ensuite extraites sous forme de points. Chaque landmark est représenté par une paire de coordonnées (x, y) dans l'image, où x représente la position horizontale et y représente la position verticale du landmark par rapport à l'origine de l'image.

IV. Utilisation des landmarks : Les coordonnées des landmarks peuvent être utilisées pour diverses applications telles que l'analyse d'expression faciale, la reconnaissance faciale, la réalité augmentée, etc. Les informations fournies par ces landmarks permettent de caractériser et d'analyser la structure du visage dans l'image.[35]

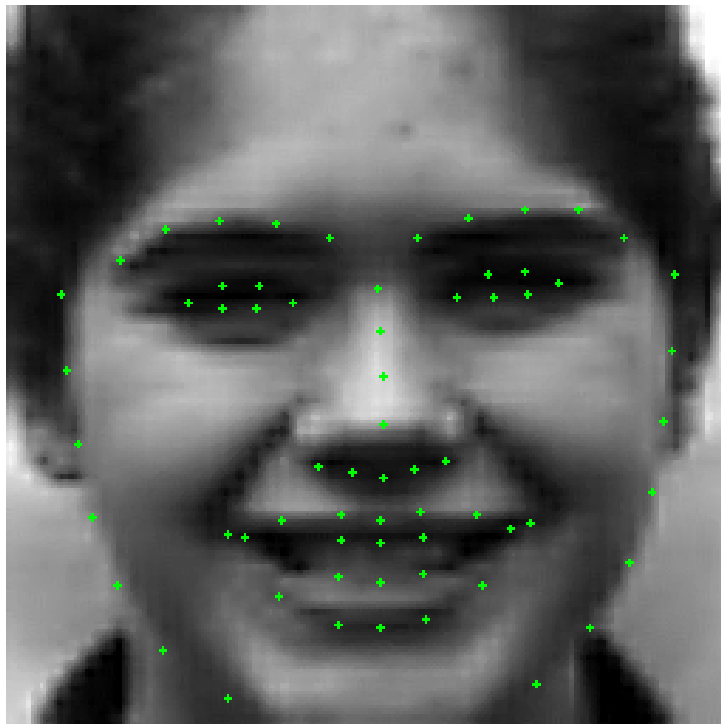


Figure 3.3 - la méthode d'extraction des points de repère faciaux(landmarks)

3.2.3 Classification

Dans cette étape, nous avons évalué les performances de notre système proposé en utilisant l'apprentissage automatique avec des méthodes de classification traditionnelles telles que le KNN (k-plus proches voisins) et le SVM (machines à vecteurs de support), ainsi que l'apprentissage en profondeur avec les réseaux CNN et DenseNet.

3.2.4 L'apprentissage automatique (Machine learning)

L'apprentissage automatique est une branche de l'intelligence artificielle (IA) et de l'informatique qui se concentre sur l'utilisation de données et d'algorithmes pour imiter le processus d'apprentissage humain, améliorant progressivement la précision des prédictions. Cette technique de modélisation repose sur des données pour programmer des machines afin qu'elles apprennent à partir de ces informations et construisent des modèles représentatifs des classes, après un processus d'apprentissage. Les données utilisées peuvent être variées, incluant des documents, des images, des sons, des vidéos, etc.[36]

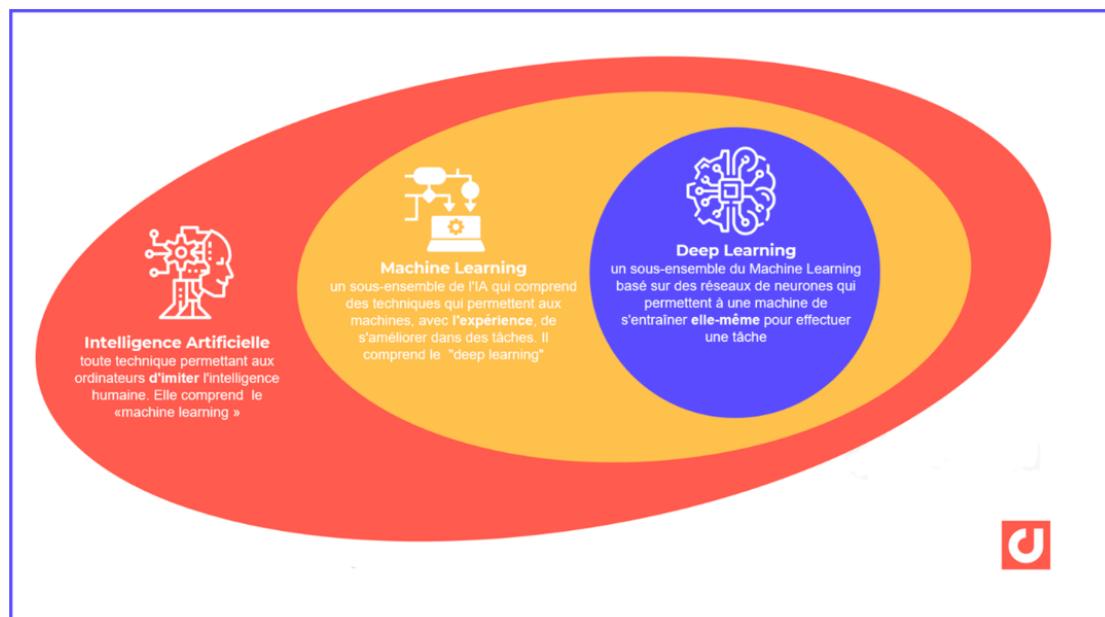


Figure 3.4 - Deep Learning comme une sorte d'apprentissage automatique [37].

L'apprentissage automatique est un pilier essentiel du domaine en pleine expansion de la science des données. En utilisant des méthodes statistiques, les algorithmes sont formés pour effectuer des classifications ou des prédictions et pour découvrir des informations cruciales dans les projets d'exploration de données. Ces informations orientent ensuite la prise de décision dans les applications et les entreprises, impactant de manière significative les mesures de croissance clés.

Avec l'expansion continue des volumes de données, la demande pour les scientifiques des données sur le marché ne cesse de croître.

3.2.4.1 K-plus proches voisins (K-NN)

3.2.4.1.1 Définition

K-Nearest Neighbors (KNN) est un algorithme d'apprentissage automatique non paramétrique utilisé pour la classification et la régression. Il est basé sur l'idée que des points similaires dans les données doivent être proches les uns des autres. KNN fonctionne en trouvant les K observations les plus proches dans l'espace des caractéristiques et en prenant une décision basée sur les valeurs de ces voisins.

3.2.4.1.2 Principe de KNN

a. Stockage des Données : KNN est un algorithme basé sur la mémoire, ce qui signifie qu'il stocke toutes les instances d'apprentissage et les utilise pour la prédiction sans créer de modèle explicite pendant la phase d'entraînement.

b. Calcul des Distances : Pour prédire la classe ou la valeur d'une nouvelle observation, KNN calcule la distance entre cette nouvelle observation et toutes les autres instances dans les données d'apprentissage.

c. Sélection des Voisins : Les K instances les plus proches (les "voisins") sont sélectionnées en fonction des distances calculées.

d. Prédiction :

» **Classification :** La classe la plus fréquente parmi les K voisins est attribuée à la nouvelle observation.

» **Régression :** La valeur moyenne ou pondérée des K voisins est utilisée pour la prédiction.

3.2.4.1.3 Distance euclidienne

La distance la plus couramment utilisée dans KNN est la distance euclidienne soient deux données représentées par deux vecteurs x_i et x_j [38], la distance entre ces deux données est donnée par :

$$d(x_i, x_j) = \sqrt{\sum_{k=1}^d (x_{ik} - x_{jk})^2} \quad (3.1)$$

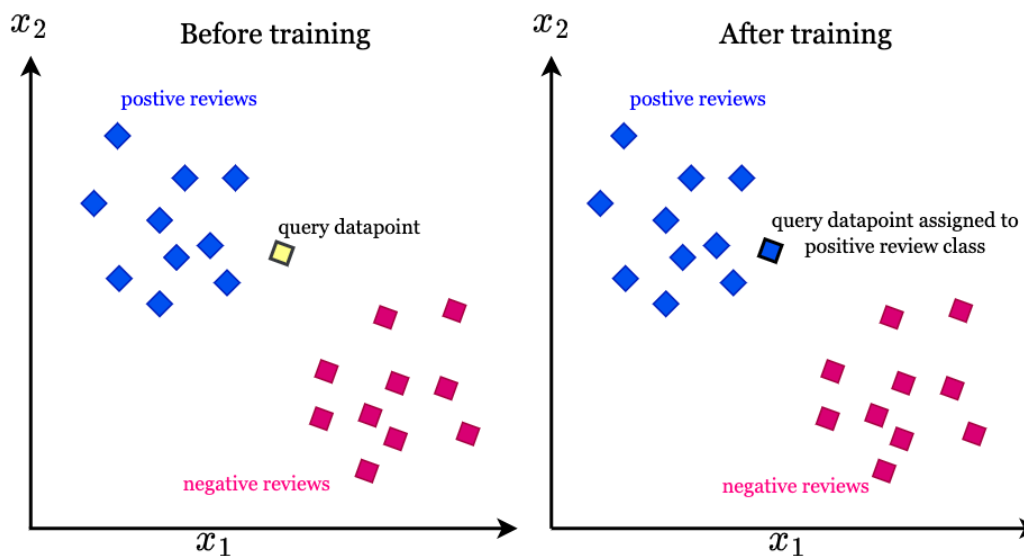


Figure 3.5 - K-Plus Proches Voisins (K-NN) Algorithme de classification : applications et exemples concrets [39].

» L'image précédente illustre le processus de classification avant et après l'entraînement avec l'algorithme des k plus proches voisins (KNN)

- le point jaune a été attribué à la classe des avis positifs (indiqué par le changement de couleur du jaune au bleu).
- KNN a examiné les K voisins les plus proches du point jaune et a déterminé que la majorité d'entre eux appartenaient à la classe des avis positifs. KNN utilise

la proximité spatiale entre les points de données pour effectuer la classification. Après l'entraînement, l'algorithme KNN attribue des étiquettes aux points non étiquetés en fonction de la majorité de leurs voisins les plus proches. Cela signifie que le point non étiqueté est classé dans la classe qui prédomine parmi ses K voisins les plus proches. En d'autres termes, KNN "vote" pour la classe la plus fréquente autour du point à classer.

3.2.4.2 Machines à vecteurs supports (SVM)

Les Machines à Vecteurs de Support (SVM) sont des algorithmes d'apprentissage automatique supervisé couramment utilisés pour les tâches de classification et de régression. Bien qu'ils soient principalement préférés pour la classification, leur capacité à modéliser les relations complexes les rend également utiles pour la régression.

Le principe de base des SVM est de trouver un hyperplan dans un espace à N dimensions (où N est le nombre de caractéristiques) qui sépare les points de données en deux classes distinctes. Cet hyperplan est choisi de manière à maximiser la marge, c'est-à-dire la distance entre les points de données les plus proches de chaque classe et l'hyperplan lui-même. Ces points sont appelés vecteurs de support, d'où le nom de l'algorithme.

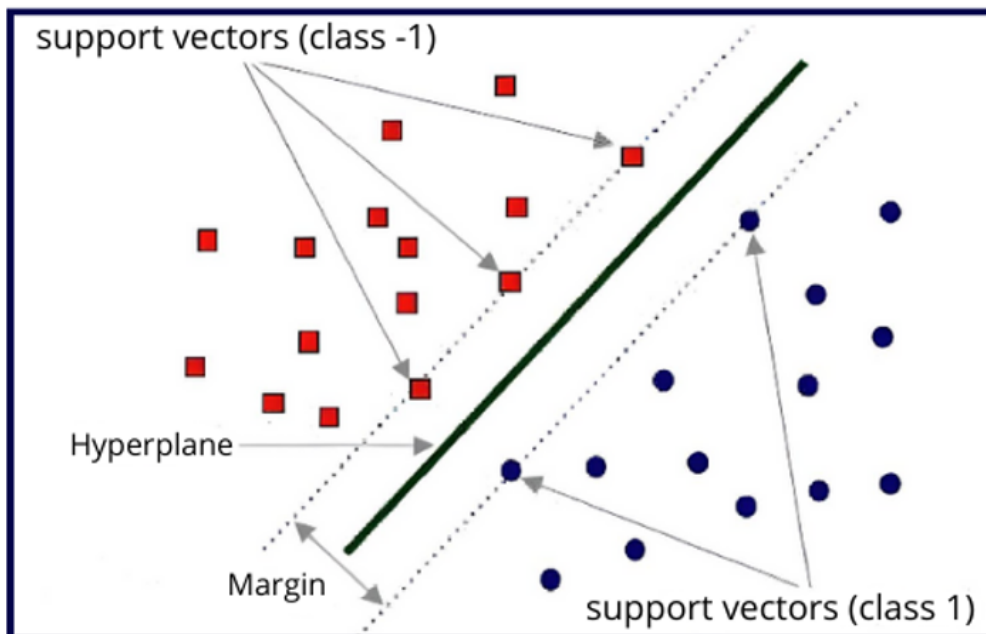


Figure 3.6 - General classification hyperplane representation of SVM algorithm [40].

» La machine à vecteurs de support (SVM) trouve un **hyperplan optimal** qui sépare les classes tout en maximisant la marge. **Les vecteurs de support** (points les plus proches de l'hyperplan) sont essentiels pour construire le modèle SVM. L'objectif est de trouver un équilibre entre la séparation des classes et la marge.

3.2.5 Convolutional neural network (CNN)

3.2.5.1 Définition

Les Réseaux de Neurones Convolutifs (CNN) sont des modèles d'apprentissage profond particulièrement efficaces pour le traitement des données ayant une structure en grille, comme les images. Les CNN sont capables d'extraire automatiquement des caractéristiques significatives à différents niveaux de complexité grâce à l'utilisation de couches de convolution, de pooling et de couches entièrement connectées.

3.2.5.2 Architecture du modèle CNN

Voici la description détaillée de l'architecture de notre modèle CNN proposé pour la reconnaissance des émotions faciales :

»**Couche Convolutionnelle 1D (512 filtres, taille du noyau 3, activation ReLU)**

- Cette couche effectue la convolution sur les données d'entrée.
- 512 filtres sont utilisés pour extraire des caractéristiques des données d'entrée.
- La fonction d'activation ReLU (Rectified Linear Unit) est appliquée pour introduire de la non-linéarité.

»**Couche de Pooling 1D (MaxPooling, taille du pool 2)**

- Cette couche réduit la dimensionnalité de la sortie de la couche précédente en prenant la valeur maximale dans chaque fenêtre de pooling.

»**Couche de Normalisation par Lots**

- Cette couche normalise les activations de la couche précédente, ce qui accélère l'apprentissage et stabilise le processus.

»**Couche Convolutionnelle 1D (1024 filtres, taille du noyau 3, activation ReLU)**

- Une autre couche de convolution pour extraire des caractéristiques plus complexes.

»**Couche de Pooling 1D (MaxPooling, taille du pool 2)**

- Réduit la dimensionnalité de sortie.

»**Couche de Normalisation par Lots**

— Normalise les activations.

»**Couche Convolutionnelle 1D (1024 filtres, taille du noyau 3, activation ReLU)**

— Une autre couche de convolution.

»**Couche d'aplatissement (Flatten)**

— Convertit la sortie de la couche précédente en un vecteur à une seule dimension pour l'entrée des couches entièrement connectées.

»**Couche Dense (1024 neurones, activation ReLU)**

— Couche entièrement connectée avec 1024 neurones.
— Applique la fonction d'activation ReLU.

»**Couche Dropout (0.5) Réduit le surapprentissage en désactivant aléatoirement 50% des neurones pendant l'entraînement.**

»**Couche Dense (512 neurones, activation ReLU)**

— Une autre couche entièrement connectée avec 512 neurones.

»**Couche Dropout (0.5)**

— Réduit le surapprentissage.

»**Couche Dense de sortie (7 neurones, activation Softmax)**

— Couche entièrement connectée avec 7 neurones de sortie, un pour chaque classe d'émotion.

*La fonction d'activation softmax est utilisée pour obtenir des probabilités pour chaque classe.

⇒ **Compilation du Modèle**

*Fonction de perte : Cross-entropy catégorique sparse.

*Optimiseur : Adam avec un taux d'apprentissage de 0,0001 et un taux de décroissance de $1e-8$.

⇒ **Entraînement du Modèle**

*Nombre d'époques : 175.

*Validation : 10% des données d'entraînement utilisées comme ensemble de validation.

*Arrêt précoce : Surveillance de la perte de validation avec une patience de 20 époques.

⇒ **Évaluation du Modèle**

» Le modèle est évalué sur l'ensemble de test pour obtenir la précision et la perte de généralisation.

Cette architecture utilise des couches convolutionnelles pour extraire des caractéristiques importantes des données d'entrée, suivies de couches entièrement connectées pour effectuer la classification finale des émotions. Les couches de normalisation par lots et de dropout sont utilisées pour améliorer la généralisation et prévenir le surapprentissage.

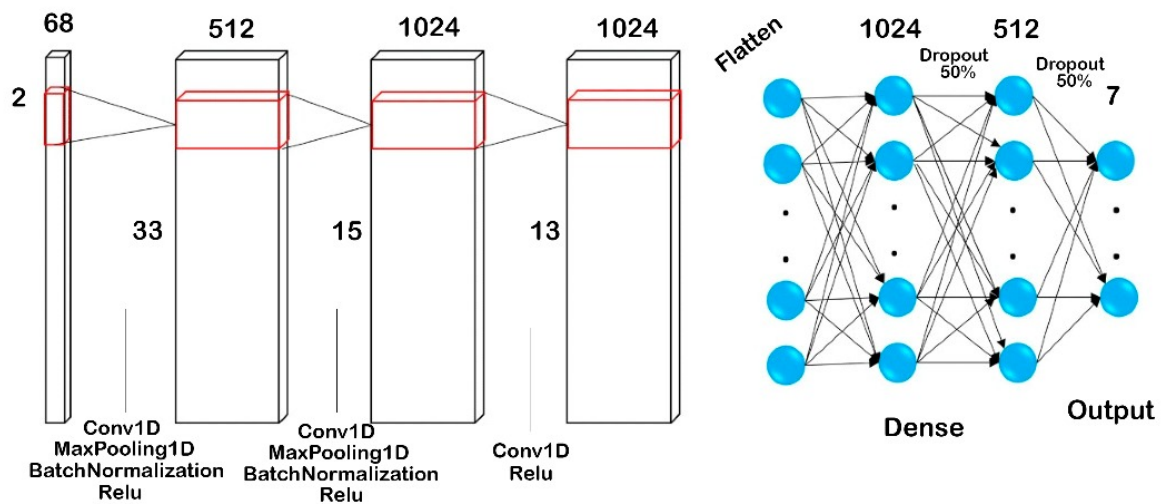


Figure 3.6 - Détail architectural de l'architecture CNN personnalisée utilisée pour CK+.

3.2.6 Réseau Convolutionnel Densément Connecté (Dense-Net)

3.2.6.1 Définition

DenseNet, ou réseau convolutionnel densément connecté, est un type de réseau de neurones convolutif qui se distingue par ses connexions directes entre chaque couche et toutes les couches précédentes. Chaque couche reçoit en entrée les caractéristiques de toutes les couches précédentes, ce qui favorise la réutilisation des caractéristiques et réduit le problème de la disparition du gradient.

3.2.6.2 Architecture du modèle DenseNet

Voici la description détaillée de l'architecture de notre modèle dense-net proposé pour la reconnaissance des émotions faciales :

»Couche d'Entrée (Input)

- Prend en entrée des données de forme (68, 2) correspondant aux coordonnées x et y des 68 points clés faciaux.

»Convolution Initiale

- Une couche de convolution avec 2 fois le taux de croissance (growth rate), un noyau de taille 7, et une activation ReLU.
- Suivie d'une normalisation par lots et d'une max pooling pour réduire la dimension.

»Blocs Dense

- Répétition de blocs densément connectés. Chaque bloc consiste en une série de couches de convolution (conv-block) concaténées.
- Les blocs sont définis par la fonction dense-block, qui empile un nombre spécifié de couches conv-block.

»Couches de Transition

- Intercalées entre les blocs dense, ces couches réduisent la dimensionnalité des fonctionnalités en appliquant une combinaison de normalisation par lots, activation ReLU, et convolution avec un taux de réduction (reduction rate) de 0.5.
- Une couche de mise en commun moyenne (AveragePooling) est utilisée pour réduire la taille.

»Couche de Classification

- Après les derniers blocs dense, une couche de normalisation par lots, activation ReLU, et une opération de mise en commun globale (GlobalAveragePooling) sont appliquées.
- Une couche dense avec une activation softmax est utilisée pour obtenir les probabilités de chaque classe d'émotion en sortie.

⇒ **Compilation du Modèle**

Fonction de perte : Cross-entropy catégorique sparse. Optimiseur : Adam avec un taux d'apprentissage de 0.0005 et un taux de décroissance de $1e-7$.

⇒ **Entraînement du Modèle**

- * Nombre d'époques : 50.
- * Validation : 10% des données d'entraînement sont utilisées comme ensemble de validation.
- * Arrêt précoce : Surveillance de la perte de validation avec une patience de 20 époques.

⇒ **Évaluation du Modèle**

* Le modèle est évalué sur l'ensemble de test pour obtenir la précision et la perte de généralisation.

Cette architecture denseNet exploite la connectivité dense entre les couches pour favoriser le flux d'informations et l'utilisation efficace des caractéristiques extraites des données d'entrée. Les couches de transition réduisent la complexité du modèle, ce qui peut aider à prévenir le surapprentissage.

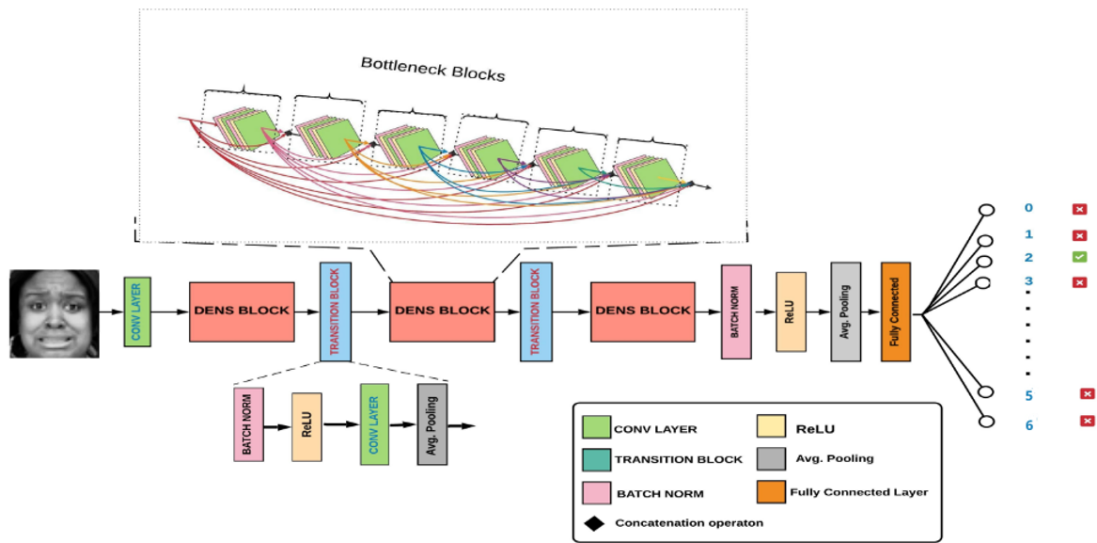


Figure 3.8 - L'architecture du réseau DenseNet[41].

3.7 Conclusion

Ce chapitre a étudié les techniques d'apprentissage automatique et profond, notamment k-NN, SVM, CNN et DenseNet, en mettant l'accent sur le prétraitement des données pour améliorer la précision des modèles. L'exemple de la classification des expressions faciales avec la base de données CK+ a illustré l'application de ces techniques. La maîtrise de ces méthodes permet de développer des solutions performantes pour diverses tâches de classification et de reconnaissance. Dans le prochain chapitre, nous vérifierons la validité de nos résultats et en discuterons en détail afin d'identifier les points forts et les faiblesses de nos approches, et d'explorer des pistes d'amélioration potentielles.

Chapitre 4

Les Résultats Expérimentaux

4.1 Introduction

Ce chapitre vise à vous guider à travers les étapes de mise en œuvre de notre approche dans un système de reconnaissance des émotions faciales. Nous commencerons par présenter les ressources, le langage de programmation et l'environnement de développement utilisés. Ensuite, nous examinerons en détail les différentes phases de la création de notre modèle. Pour finir, nous partagerons les résultats des tests effectués, en comparant les performances de modèles d'apprentissage profond comme DenseNet et CNN avec celles de modèles d'apprentissage automatique comme SVM et KNN.

4.2 Expériences et résultats

Les exigences techniques de ce projet sont les suivantes :

4.2.1 Exigences matérielles :

⇒ **Nom de l'appareil** : hp elitebook 840 g5

Processeur : Intel(R) Core(TM) i5-8250U CPU @ 1.60GHz 1.80 GHz

Mémoire RAM : 8,00 Go

disque dur : ssd 256 NVME

Type du système : Système d'exploitation 64 bits, processeur x64

⇒ **Nom de l'appareil** : HP Laptop 15-dw3xxx

Processeur : 11th Gen Intel(R) Core(TM) i7-1165G7 @ 2.80GHz 2.80 GHz

Mémoire RAM : 16,00 Go

disque dur : ssd 256 NVME

Type du système : Système d'exploitation 64 bits, processeur x64

4.2.2 Logiciels requis

a. système opérateur :

Windows 10 Professionnel , Windows 11 Professionnel

b. Langage de programmation :

Python (version 3.12)

c. Environnement de développement intégré (IDE) :

PyCharm (version 2023.3.3)

d. LaTeX (Typesetting System on Overleaf) :

Typesetting System on Overleaf : une plateforme collaborative en ligne pour la rédaction, l'édition et la publication de documents LaTeX.

e. Bibliothèques :

I. Opencv (version 4.9.0.80) :

Définition : Une bibliothèque de boîte à outils pour l'apprentissage automatique.

Utilité : Utilisée pour la détection et la reconnaissance de visages, inclut des algorithmes pour la machine learning, la vision par ordinateur et l'analyse de données.

II. dlib (version 19.24.2) :

Définition : Une bibliothèque de boîte à outils pour l'apprentissage automatique.

Utilité : Utilisée pour la détection et la reconnaissance de visages, inclut des algorithmes pour la machine learning, la vision par ordinateur et l'analyse de données.

III. numpy (version 1.26.4) :

Définition : Une bibliothèque pour le calcul scientifique.

Utilité : Utilisée pour des opérations sur des tableaux multidimensionnels (matrices) et des fonctions mathématiques de haut niveau.

IV. matplotlib (version 3.8.3) :

Définition : Une bibliothèque pour la visualisation de données.

Utilité : Permet de créer des graphiques statiques, animés et interactifs en Python, souvent utilisée pour tracer des courbes, des histogrammes, des scatter plots, etc.

V. pandas (version 2.2.1) :

Définition : Une bibliothèque pour la manipulation et l'analyse de données.

Utilité : Fournit des structures de données flexibles et expressives comme les DataFrames, permettant une manipulation facile des données, leur nettoyage et leur analyse.

VI. seaborn (version 0.13.2) :

Définition : Une bibliothèque pour la visualisation statistique des données.

Utilité : Basée sur Matplotlib, elle offre des graphiques plus attrayants et informatifs, avec des fonctionnalités pour tracer des distributions de données et des relations statistiques.

VII. tensorflow (version 2.16.1) :

Définition : Une bibliothèque pour l'apprentissage automatique et l'intelligence artificielle développée par Google.

Utilité : Utilisée pour construire et entraîner des modèles de machine learning et de deep learning, particulièrement pour les réseaux de neurones.

VIII. keras (version 3.3.3) :

Définition : Une interface utilisateur de haut niveau pour les réseaux de neurones, fonctionnant au-dessus de TensorFlow.

Utilité : Facilite la construction et l'entraînement de modèles de deep learning avec une syntaxe simple et intuitive.

IX. Tkinter :

Définition : Tkinter est la bibliothèque standard de Python pour la création d'interfaces graphiques utilisateur (GUI). Elle permet de concevoir des applications dotées d'interfaces graphiques en utilisant le langage de programmation Python.

Utilité : Tkinter est utilisé pour créer des interfaces graphiques utilisateur (GUI). Il permet de développer des applications de bureau interactives, de prototyper rapidement des applications GUI, et est couramment utilisé dans l'enseignement et l'apprentissage de la programmation des interfaces graphiques en Python.

4.3 Interface graphique du système de macro-expression

- » la première fenêtre contient deux boutons :
 - **Launch** : pour lancer application
 - **Exit** : pour quitter application

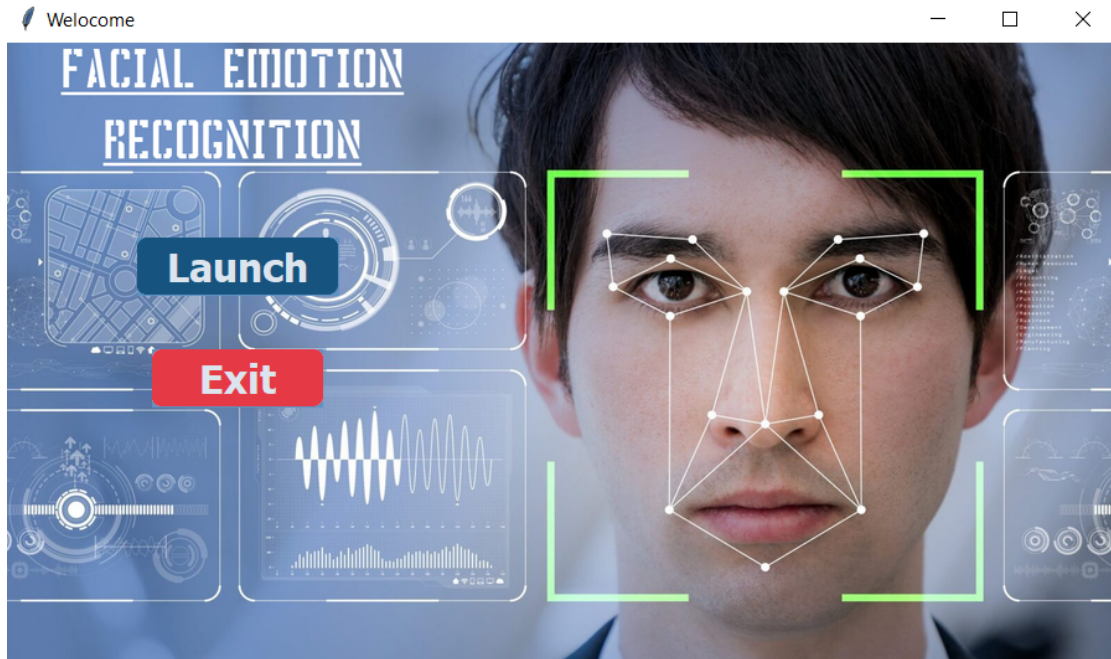


Figure 4.1 - L'écran d'accueil de notre application

- » la deuxième fenêtre contient quatre boutons :
 - select image : pour sélectionner image depuis votre pc
 - preprocess image : pour applique la prétraitement sur image sélectionnée
 - Extract landmarks : pour fait l'extraction après la prétraitement
 - Classify image : pour fait la classification

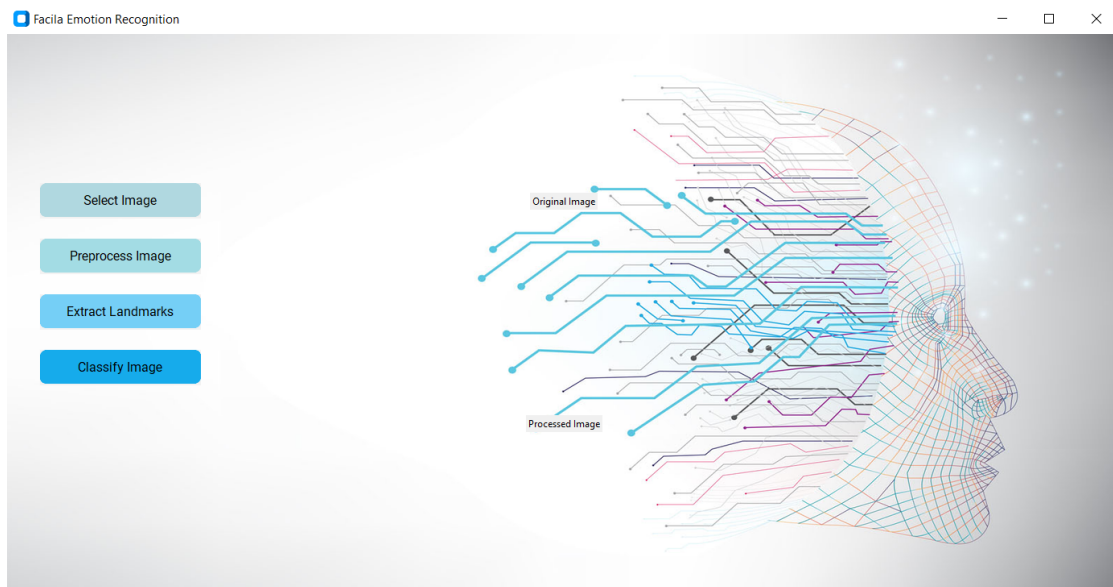


Figure 4.2 - Interface Reconnaissance des émotions faciales

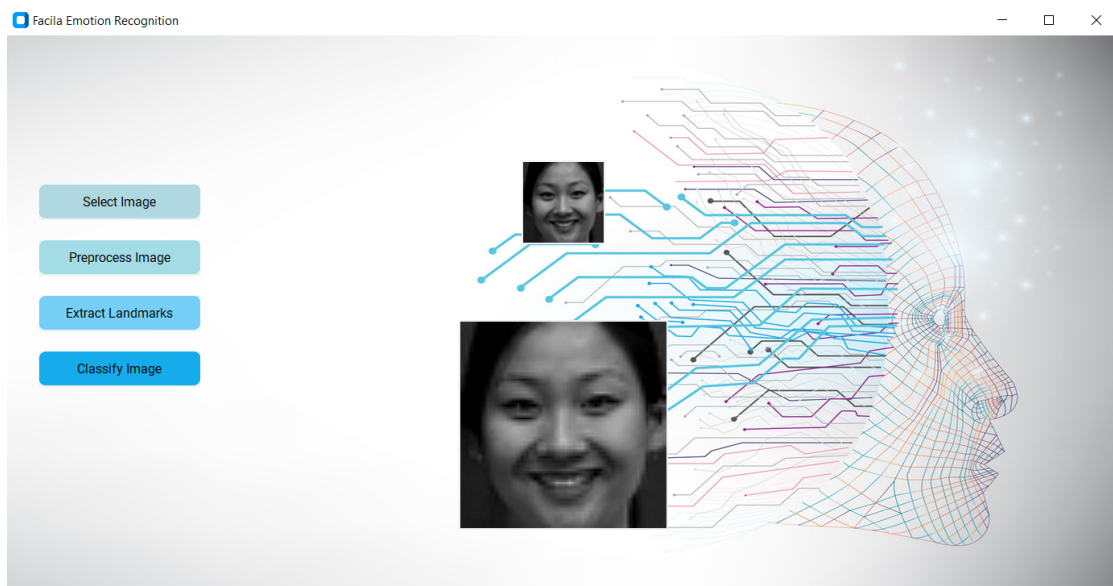


Figure 4.3 - L'étape sélectionner image et prétraitement

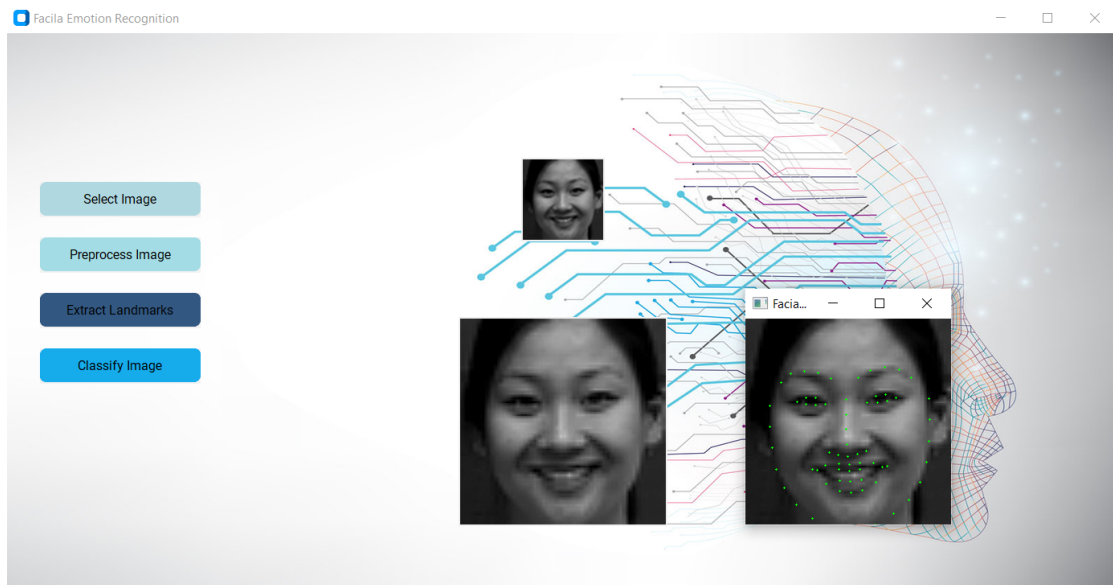


Figure 4.4 - Étape d'extraction (Landmarks)

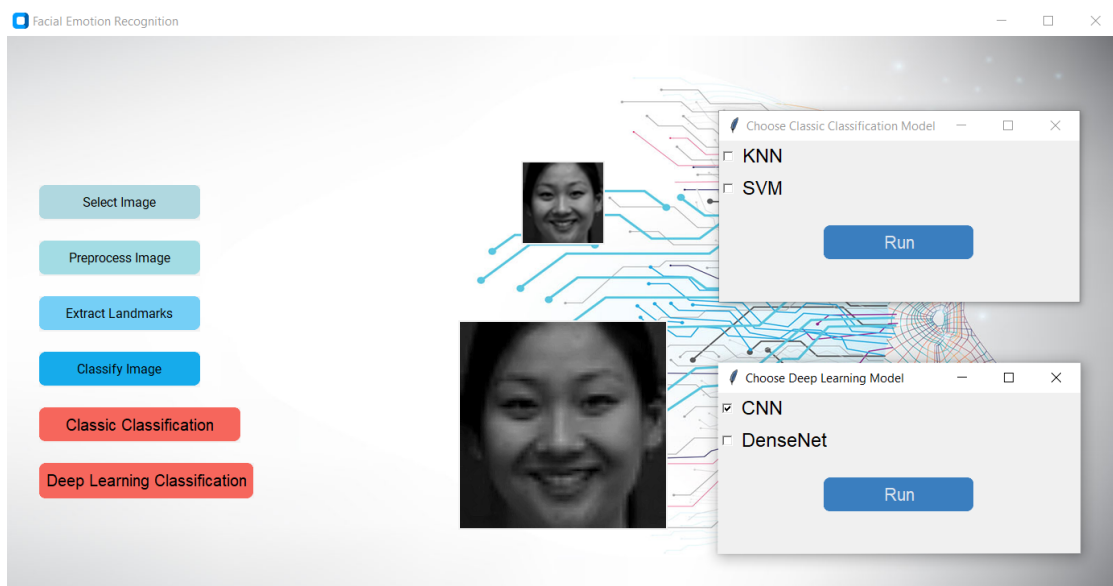


Figure 4.5 - Fenêtre de sélection le classificateur

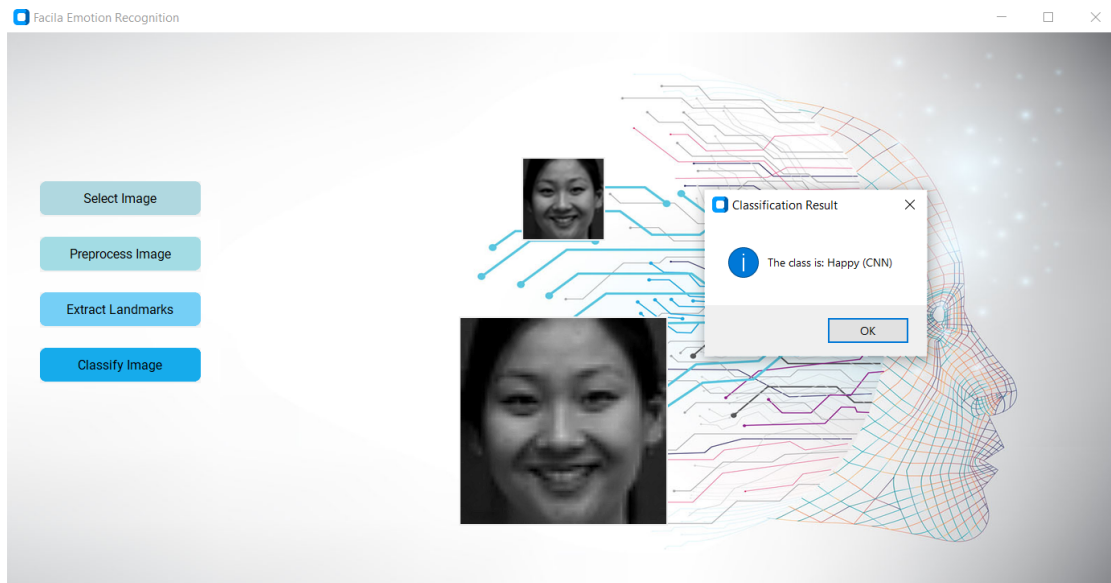


Figure 4.6 - résultats classification par cnn

4.4 Résultats numérique

4.4.1 La base de donnée

Le choix de CK+ (Extended Cohn-Kanade) pour notre projet de la reconnaissance des émotions faciales avec l'apprentissage profond est judicieux pour plusieurs raisons :

- CK+ est une base de données riche et bien documentée, Elle offre une combinaison de diversité, qualité et standardisation.
- Utilisation extensive : La base de données CK+ est largement considérée comme la base de données de classification des expressions faciales contrôlée en laboratoire la plus utilisée, et est utilisée dans la majorité des méthodes de classification des expressions faciales.
- Les caractéristiques font du CK+ un excellent choix pour l'apprentissage profond et la reconnaissance des expressions faciales. De plus, de nombreux travaux de recherche ont utilisé CK+ pour leurs expériences, ce qui facilite la comparaison des résultats.

4.4.2 Implémentation de base de données ck+

L'ensemble de données est Ck + téléchargé depuis kaggle se compose d'images de visages en niveaux de gris de 48x48 pixels. Les visages ont été automatiquement

enregistrés afin que le visage soit plus ou moins centré et occupe à peu près la même quantité d'espace dans chaque image.

La tâche consiste à classer chaque visage en fonction de l'émotion montrée dans l'expression faciale dans l'une des sept catégories (0 = anger, 1 = contempt, 2 = disgust, 3 = fear, 4 = happy, 5 = sadness, 6 = surprise).

Elle contient 981 images expressions faciales.

» Pour les valeurs d'usage on a 90% training et 10% pour le test :

Table 4.1 - Répartition de la base de données CK+ en train et test

EMOTION	Train	Test
Anger	121	14
Disgust	159	18
Fear	67	8
Happy	186	21
Sadness	76	8
Surprise	224	25
Contempt	49	5
TOTAL	882	99

4.4.3 Résultats

On a présenté, par un tableau et un graphique, les mesures de performance de nos systèmes qui comparent les performances de différents systèmes de reconnaissance des émotions utilisant des modèles d'apprentissage automatique (SVM, KNN) et profond (CNN, DenseNet). Les métriques de performance évaluées sont le Taux de Faux Acceptions (FAR), le Taux de Faux Rejets (FRR), le Taux d'Erreur Égale (EER) et la Précision (Accuracy).

Table 4.2 - Les mesures de performance de nos systèmes

	FAR	FRR	EER	Accuracy
SVM	0.31%	1.59%	0.95%	97.9%
KNN	0,50%	2,38%	1,44%	96.9%
CNN	0%	0%	0%	100%
DenseNet	0.36%	1.81%	1.09%	98.9%

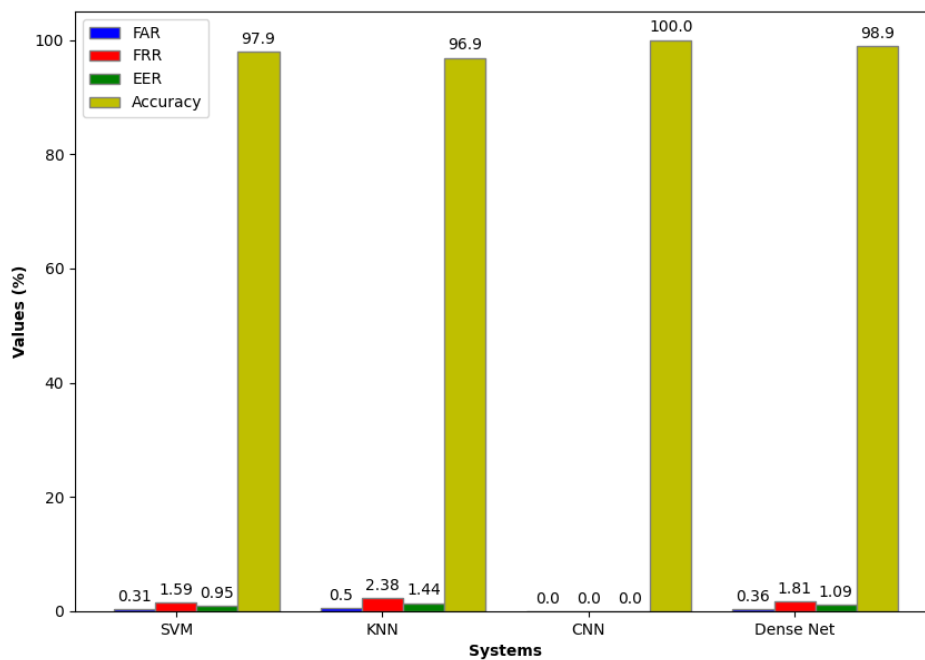


Figure 4.7 - diagramme des mesure de performance de nos systèmes

L'analyse des performances des différents modèles de reconnaissance des expressions faciales révèle des résultats intéressants. Le modèle SVM (Support Vector Machine) présente un taux d'acceptation erronée (FAR) de 0,31% et un taux de rejet erroné (FRR) de 1,59% , avec un taux d'erreur égal (EER) de 0,95% . Bien que ces chiffres soient relativement bas, ils sont légèrement supérieurs à ceux du modèle CNN (Convolutional Neural Network). Le modèle KNN (K-Nearest Neighbors), quant à lui, affiche un FAR de 0,50% et un FRR de 2,38% , indiquant une tendance à rejeter incorrectement des entrées légèrement plus élevée que le SVM, avec un EER de 1,44% et une précision de 96,9% .

En comparaison, le modèle CNN se distingue par une performance exceptionnelle, atteignant une précision de 100% sans aucune erreur d'acceptation ou de rejet fausse, ce qui se traduit par des taux FAR, FRR, et EER de 0% . Cela démontre la capacité supérieure du CNN à classifier correctement les expressions faciales sans erreurs. Le Dense Net, bien qu'il affiche une précision impressionnante de 98,9% , montre un léger accroissement des erreurs avec un FAR de 0,36% , un FRR de 1,81% , et un EER de 1,09% , ce qui reste néanmoins une performance remarquable.

» le modèle CNN se révèle être le plus efficace pour la reconnaissance des expressions faciales selon ces données, suivi de près par le Dense Net. Les modèles SVM et KNN montrent des performances honorables mais sont surpassés par les capacités avancées du CNN et du Dense Net. Ces résultats soulignent l'importance de choisir le bon modèle pour des tâches spécifiques, le CNN étant le choix optimal pour ce projet en particulier.

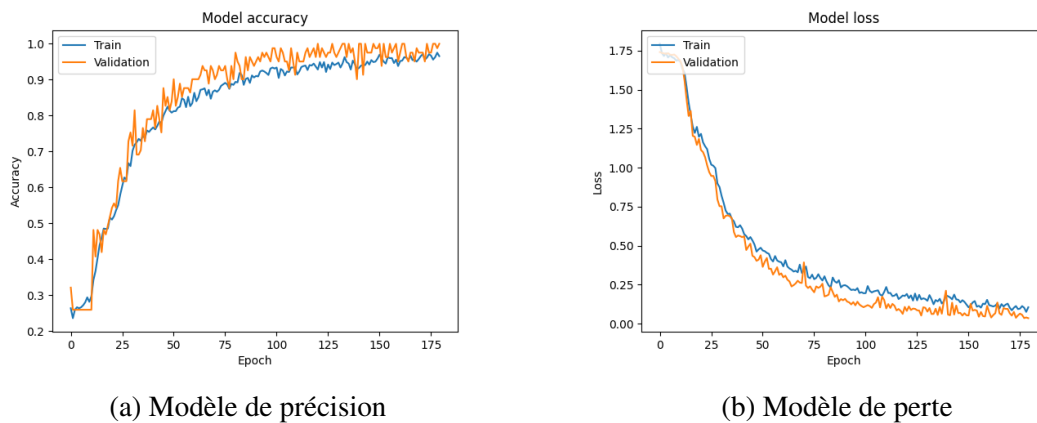


Figure 4.8 - Évolution de la performance du modèle d'apprentissage profond : Précision (a) et Perte (b) au fil des époques (DenseNet)

On remarque d'après les formes de figure (25) la Précision du modèle (Model accuracy) la précision de l'entraînement (ligne bleue) augmente rapidement après 20 époques et se stabilise dans la valeur 1, indiquant un bon apprentissage au fil des époques. La précision de la validation (ligne orange) suit une tendance similaire, mais avec une légère divergence par rapport à la ligne d'entraînement pour perte du modèle (Model loss) la perte d'entraînement (ligne bleue) diminue de manière significative jusqu'à la valeur 0,15 ce qui est attendu pendant l'apprentissage. La perte de validation (ligne orange) diminue également jusqu'à la valeur 0,1.

Finalement les deux graphiques montrent une amélioration de la performance du modèle avec l'augmentation des époques et donner excellence résultats par la méthode apprentissage profond CNN.

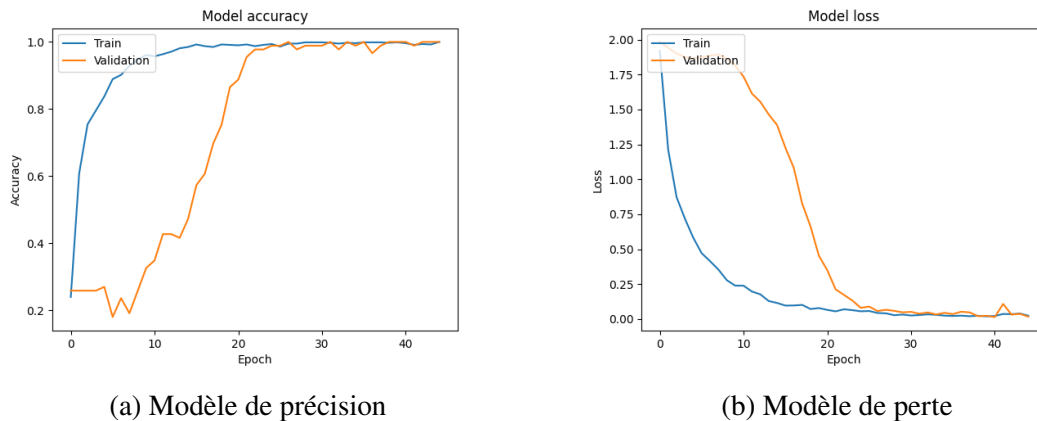


Figure 4.9 - Évolution de la performance du modèle d'apprentissage profond : Précision (a) et Perte (b) au fil des époques (DenseNet)

On remarque d'après les formes de figure (26) le Graphique de Perte du Modèle (Model loss) la perte d'entraînement (Train) et la perte de validation (Validation) diminuent rapidement au début, ce qui est un bon signe que le modèle apprend correctement. Après environ 15 époques, la perte de validation se stabilise en 0,2 , ce qui peut indiquer que le modèle ne s'améliore plus significativement sur les données de validation. Pour graphique de précision du modèle (Model accuracy) la précision d'entraînement et de validation augmente également rapidement, ce qui correspond à l'amélioration de la performance du modèle. On remarque aussi un écart de se forme après environ 5 époques où la précision d'entraînement continue de s'améliorer tandis que la précision de validation plafonné jusqu'à atteindre la valeur "1".

Finalement l'écart croissant entre la perte de validation et la perte d'entraînement, ainsi que l'écart entre la précision de validation et d'entraînement, Le modèle devient très bon sur les données d'entraînement et donne bon résultats par méthode apprentissage profond dense net .

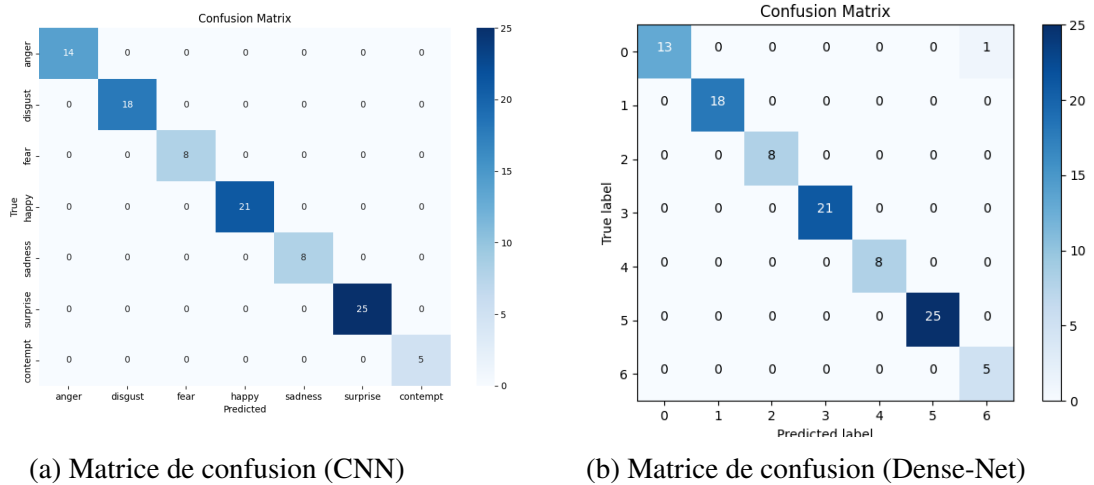


Figure 4.10 - Matrice de confusion pour les deux modèles CNN (a) et Dense-Net (b)

»Matrice CNN (a)

Le modèle a montré une performance exceptionnelle en identifiant avec précision toutes les émotions. Cette matrice met en lumière la capacité du modèle à différencier avec finesse les diverses émotions, prouvant ainsi sa compréhension approfondie des spécificités propres à chaque émotion. L'efficacité du modèle se reflète également dans sa capacité à gérer des nuances subtiles, ce qui est crucial pour une analyse émotionnelle sophistiquée.

»Matrice Dense-Net (b)

le modèle a correctement classifié presque toutes les images testées. Les valeurs élevées sur la diagonale principale de la matrice montrent que la majorité des prédictions étaient correctes pour chaque classe. Les quelques valeurs non nulles hors de la diagonale représentent des erreurs de classification, mais elles sont très peu nombreuses, ce qui est cohérent avec le taux de précision élevé. Cela suggère que le modèle est très bien entraîné et qu'il généralise bien sur les données qu'il n'a pas vues pendant l'entraînement.

Table 4.3 - Comparaison des performances avec d'autres modèles

Method	Accuracy
Niu et al [42]	93.2%
Mirnalinee TT et al [43]	98.49%
Notre approche	100%

Pour mieux comprendre les performances de notre modèle, nous avons comparé notre approche avec d'autres méthodes dans le tableau ci-dessus.

Le tableau présente une comparaison des taux de précision entre notre modèle et ceux de deux autres études de référence.

Niu et al. A obtenu une précision de 93,2%. Bien que cette précision soit respectable, elle est inférieure à celle obtenue par notre modèle. Cela pourrait être dû à des différences techniques de prétraitement ou les architectures de modèles. Mirnalinee TT et al. Cette étude rapporte une précision de 98,49%. Cette précision est très élevée et reflète une performance efficace du modèle utilisé par ces chercheurs. Toutefois, notre modèle a réussi à surpasser cette performance.

4.4.4 Discussion

Dans cette étude, nous avons comparé les performances de différents systèmes de reconnaissance des émotions en utilisant à la fois des modèles d'apprentissage automatique (SVM et KNN) et des modèles d'apprentissage profond (CNN et DenseNet). Les métriques de performance évaluées incluent le Taux de Faux Acceptions (FAR), le Taux de Faux Rejets (FRR), le Taux d'Erreur Égale (EER) et la Précision (Accuracy).

Les résultats montrent clairement que les modèles d'apprentissage profond (CNN et DenseNet) surpassent les modèles d'apprentissage automatique traditionnels (SVM et KNN) dans toutes les métriques de performance. Cela peut être attribué à la capacité des réseaux neuronaux profonds à apprendre des représentations complexes et à capturer les caractéristiques fines des données d'entrée, ce qui est crucial pour la tâche de reconnaissance des émotions.

DenseNet, en particulier, a montré les meilleures performances globales, avec les taux de FAR et FRR les plus bas, l'EER le plus faible et la précision la plus élevée. Cette performance peut être due à l'architecture dense de DenseNet, qui permet une meilleure réutilisation des caractéristiques et une propagation plus efficace des gradients.

Cette comparaison met en évidence l'efficacité supérieure des modèles d'apprentissage profond par rapport aux modèles d'apprentissage automatique pour la reconnaissance des émotions. Les résultats obtenus avec CNN et DenseNet, en particulier, montrent une capacité exceptionnelle à discriminer et reconnaître les émotions avec une grande précision, faisant d'eux les choix privilégiés pour les applications nécessitant une compréhension fine des états émotionnels.

Après la comparaison entre les méthodes de classification on a fait comparaison avec autre résultat on remarque que Notre modèle a atteint une précision de 100% mieux que les autres . Ce résultat est remarquable et suggère que notre modèle est extrêmement efficace pour classer les émotions dans les ensembles de données utilisés. Cette performance pourrait être attribuée à une architecture de réseau de neurones optimisée, une sélection rigoureuse des paramètres, et un ensemble de données bien préparé.

Les résultats de cette étude ont des implications importantes pour le développement de systèmes de reconnaissance des émotions dans des applications réelles, telles que les interfaces utilisateur adaptatives, les systèmes de tutorat intelligent, et les plateformes de bien-être émotionnel. En choisissant des modèles d'apprentissage profond, les développeurs peuvent créer des systèmes plus précis et fiables, capables de mieux comprendre et répondre aux émotions des utilisateurs.

4.5 Conclusion

En conclusion, ce chapitre démontre l'efficacité des réseaux de neurones convolutifs (CNN) pour la reconnaissance des émotions. DenseNet a montré des performances légèrement supérieures, mais le modèle CNN reste compétitif et améliorable. La comparaison avec d'autres approches d'apprentissage automatique, comme SVM et KNN, souligne la robustesse de l'apprentissage profond dans ce domaine.

Conclusion Générale

La reconnaissance des émotions faciales constitue une avancée majeure dans le domaine de l'intelligence artificielle, fusionnant habilement les domaines de la vision par ordinateur, de l'apprentissage automatique et de la psychologie. L'objectif central de cette étude était de concevoir un système capable de détecter et d'interpréter les émotions humaines à partir d'images faciales avec une précision remarquable.

Dans cette recherche, une attention particulière a été portée au développement d'un système de reconnaissance des émotions faciales en tirant parti des techniques d'apprentissage automatique et des méthodes de deep learning. Un processus méthodique a été suivi, comprenant diverses étapes de prétraitement telles que le redimensionnement des images, l'ajustement des couleurs, le filtrage de netteté et la conversion en niveaux de gris. Pour l'extraction des caractéristiques, la méthode des landmarks a été employée, tandis que pour la classification, une variété d'algorithmes ont été explorés, incluant K-NN et SVM pour les approches traditionnelles, ainsi que des réseaux de neurones convolutionnels (CNN) et l'architecture DenseNet pour les méthodes de deep learning.

Les résultats expérimentaux ont démontré la supériorité des modèles de deep learning par rapport aux méthodes traditionnelles telles que K-NN et SVM, en termes de précision et de robustesse. En particulier, l'architecture CNN a montré des performances légèrement meilleures que les DenseNet, le modèle dense net reste compétitif et présente des possibilités d'amélioration grâce à ses connexions denses qui facilitent l'apprentissage de caractéristiques complexes et améliorent la propagation des gradient. Cependant le modèle CNN présente aussi des possibilités d'amélioration par le biais d'ajustements supplémentaires et d'une optimisation fine des hyperparamètres.

Les méthodes de prétraitement des données, en particulier l'identification et l'alignement des points de repère faciaux, ont également été déterminantes dans l'amélioration des performances du modèle. En utilisant la base de données CK+, des résultats probants ont été obtenus, illustrant ainsi l'application pratique de ces techniques et validant l'efficacité de l'approche adoptée pour la reconnaissance précise des émotions à partir des images correspondantes.

En conclusion, cette étude confirme le potentiel considérable des techniques de deep learning dans le domaine de la reconnaissance des émotions faciales, ouvrant ainsi la voie à des applications plus sophistiquées et interactives. Les perspectives futures de recherche incluent l'exploration de nouvelles architectures de réseaux neuronaux, l'intégration de données multimodales telles que l'audio et la vidéo, ainsi que l'adaptation des modèles à des environnements réels pour des applications en temps réel. Ainsi, cette recherche contribue de manière significative à l'amélioration continue des systèmes intelligents et à une meilleure compréhension des émotions humaines.

Bibliographie

- [1] I. Benchennane, *Etude et mise au point d'un procédé biométrique multimodale pour la reconnaissance des individus*. PhD thesis, Université Mohamed Boudiaf des sciences et de la technologi, 2016.
- [2] F. Perronnin and J.-L. Dugelay, "Introduction à la biométrie : Authentification des individus par traitement audio-vidéo," *Revue Traitement du Signal*, vol. 19, no. 4, 2002. l'Institut Eurocom, Département des Communications Multimédia.
- [3] A. K. J. . A. R. K. Nandakumar, *Biometric Recognition*. James Wayman, 2011.
- [4] Techniques de l'Ingénieur, *La biométrie - Techniques et usages*. PhD thesis, 2004. https://www.techniques-ingenieur.fr/?fbclid=IwAR0TzLtC-Nfy9ZE4WSjpztzx0_f8c2KzBier9XpM41MAdiDbC6_TWtSDqs.
- [5] F. Touka, "Reconnaissance de la paume de la main," Master's thesis, Ecole Nationale Supérieure d'Informatique (ESI) Oued-Smar, Alger, 2010.
- [6] E. Herrat, R. Alaoui, and H. Bouzahir, "Système d'identification biométrique par fusion multimodale," 2010.
- [7] Y. Kabbara, *Caractérisation des images à Rayon-X de la main par des modèles mathématiques : application à la biométrie*. PhD thesis, Université Paris-Est, Université Libanaise, 2016.
- [8] R. M. Bolle, J. H. Connell, S. Pankanti, N. K. Ratha, and A. W. Senior, "The relation between the roc curve and the cmc," in *Proceedings of the Fourth IEEE Workshop on Automatic Identification Advanced Technologies*, 2005.
- [9] I. Azizan and K. Fatimah, "Facial emotion recognition : A brief review," in *International Conference on Sustainable Engineering, Technology and Management 2018 (ICSETM-2018)*, 2018.
- [10] M. Harshini Linga, V. Komma, R. Yamala, and Gnanadeep, *Automatic Music Player Based on the Facial Expression*. PhD thesis, Oct 2023.

- [11] K. GHANEM, *Reconnaissance des Expressions Faciales à Base d'Informations Vidéo ; Estimation de l'Intensité des Expressions Faciales*. PhD thesis, Université Mentouri de Constantine, Oct 2010.
- [12] D. A. Chanti, *Automatic Analysis of Macro and Micro Facial Expressions : Detection and Recognition via Machine Learning. Signal and Image processing*. Thèse de doctorat, Université Grenoble Alpes, 2019.
- [13] Cosmétothèque, “La beauté musclée,” 2021. Posted on 8 AVRIL 2021.
- [14] C. Marechal *et al.*, “Survey on ai-based multimodal methods for emotion detection,” in *High-Performance Modelling and Simulation for Big Data Applications : Selected Results of the COST Action IC1406 cHiPSet* (J. Kołodziej and H. González-Vélez, eds.), pp. 307–324, Cham : Springer International Publishing, 2019.
- [15] W. Mellouka and W. Handouzia, “Facial emotion recognition using deep learning : review and insights,” in *Proceedings of the Conference on Facial Emotion Recognition*, (Leuven, Belgium), August 9-12 2020.
- [16] C. Shan, S. Gong, and P. W. McOwan, “Facial expression recognition based on local binary patterns : A comprehensive study,” *Image and Vision Computing*, vol. 27, pp. 803–816, May 2009.
- [17] M. H. Alkawaz, D. Mohamad, A. H. Basori, and T. Saba, “Blend shape interpolation and facs for realistic avatar,” *3D Research*, vol. 6, p. 6, Jan 2015.
- [18] T. Jabid, M. H. Kabir, and O. Chae, “Robust facial expression recognition based on local directional pattern,” *ETRI Journal*, vol. 32, no. 5, pp. 784–794, 2010.
- [19] S. Zhang, L. Li, and Z. Zhao, “Facial expression recognition based on gabor wavelets and sparse representation,” in *2012 IEEE 11th International Conference on Signal Processing*, vol. 2, pp. 816–819, Oct 2012.
- [20] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” *Nature*, vol. 521, pp. 436–444, 2015.
- [21] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, “Going deeper with convolutions,” *CoRR*, vol. abs/1409.4842, 2014.
- [22] A. Mollahosseini, D. Chan, and M. H. Mahoor, “Going deeper in facial expression recognition using deep neural networks,” in *2016 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pp. 1–10, Mar 2016.

- [23] “2018 international conference of computer and informatics engineering (ic2ie),” in *IOP Conf. Series : Journal of Physics : Conf. Series*, vol. 1193, p. 012004, IOP Publishing, 2019.
- [24] S. Dong, Y. Gong, J. Shi, M. Shang, X. Tao, X. Wei, X. Hong, and T. Zhou, “Brain cognition-inspired dual-pathway cnn architecture for image classification,” *IEEE Transactions on Neural Networks and Learning Systems*, Jan 2023.
- [25] A. Sultana, S. K. Dey, and M. A. Rahman, “Facial emotion recognition based on deep transfer learning approach,” May 2023.
- [26] M. Pantic, M. Valstar, R. Rademaker, and L. Maat, “Web-based database for facial expression analysis,” in *Multimedia and Expo, 2005. ICME 2005. IEEE International Conference on*, pp. 5–pp, IEEE, 2005.
- [27] M. Valstar and M. Pantic, “Induced disgust, happiness and surprise : an addition to the mmi facial expression database,” in *Proc. 3rd Intern. Workshop on EMOTION (satellite of LREC) : Corpora for Research on Emotion and Affect*, p. 65, 2010.
- [28] M. Pantic, M. Valstar, R. Rademaker, and L. Maat, “Web-based database for facial expression analysis,” 2005.
- [29] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, and Z. Ambadar, “The extended cohn-kanade dataset (ck+) : A complete dataset for action unit and emotion-specified expression,” *Robotics Institute, Carnegie Mellon University, Pittsburgh, PA, 152131 Department of Psychology, University of Pittsburgh, Pittsburgh, PA, 152602*, 2010.
- [30] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, and I. Matthews, “The extended cohn-kanade dataset (ck+) : A complete dataset for action unit and emotion-specified expression,” in *Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on*, pp. 94–101, IEEE, 2010.
- [31] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, and Z. Ambadar, “The extended cohn-kanade dataset (ck+) : A complete dataset for action unit and emotion-specified expression,” tech. rep., Robotics Institute, Carnegie Mellon University, 2010.
- [32] I. J. Goodfellow, D. Erhan, P. L. Carrier, A. Courville, M. Mirza, B. Hamner, W. Cukierski, Y. Tang, D. Thaler, and D.-H. Lee, “Challenges in representation learning : A report on three machine learning contests,” in *International Conference on Neural Information Processing*, pp. 117–124, Springer, 2013.
- [33] M. A. Chaudhary, R. Nunes, K. Nasrollahi, M. Rehm, and T. B. Moeslund, “Deep emotion recognition through upper body movements and facial expressions,” 2020.

- [34] S. Li and W. Deng, "Deep facial expression recognition : A survey," *IEEE Transactions on Affective Computing*, October 2018.
- [35] Y. Qiu and Y. Wan, "Facial expression recognition based on landmarks," in *2019 IEEE 4th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC)*, (Lanzhou, China), IEEE, 2019. School of Information Science and Engineering, Lanzhou University.
- [36] J. Hurwitz and D. Kirsch, *Machine Learning For Dummies®*, IBM Limited Edition. Wiley, 2016.
- [37] DigiMind, "Inteligencia artificial : escucha redes sociales, ejemplos concretos." <https://blog.digimind.com/es/insight-driven-marketing/inteligencia-artificial-escucha-redes-sociales-ejemplos-concretos>, 2023. Accessed on : feb 21.
- [38] F. Chamroukhi, "Classification supervisée : Les k-plus proches voisins," tech. rep., Université de Caen, Statistics & Data Science Lab of Mathematics Nicolas Oresme UMR CNRS 6139, 2023. Available at <https://chamroukhi.com/courses/2014-2015/i41/Projets/projet-2-i41-Knn.pdf>.
- [39] LinkedIn Pulse, "K-nearest neighbors (knn) algorithm for classification." <https://www.linkedin.com/pulse/k-nearest-neighbors-knn-algorithm-classification-nrxgf>.
- [40] H. R. Baghaee, D. Mlakic, S. Nikolovski, and T. Dragicević, "Support vector machine-based islanding and grid fault detection in active distribution networks," *IEEE Journal of Emerging and Selected Topics in Power Electronics*, vol. XX, May 2019.
- [41] A. Ghosh, F. Sultana, A. Sufian, and A. Chakrabarti, "Fundamental concepts of convolutional neural network," January 2020.
- [42] B. Niu, Z. Gao, and B. Guo, "Facial expression recognition with lbp and orb features," *Computational Intelligence and Neuroscience*, vol. 2021, pp. 1–10, 2021.
- [43] S. Benisha and T. Mirnalinee, "Human facial emotion recognition using deep neural networks," *The International Arab Journal of Information Technology*, vol. 20, p. 303, May 2023.