République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
Université Chadli Bendjedid
Faculté des Sciences et de la Technologie
Département d'Informatique

الجمهورية الجزائرية الديمقراطية الشعبية
وزارة التعليم العالي والبحث العلمي
جامعة الشاذلي بن جديد
كلية العلوم و التكنولوجيا
قسم الاعلام الالي

جامعة الشاذلي بن جديد
UNIVERSITE CHADLI BENDJEDID

MEMOIRE

Présenté par

KHERIF hana

Pour l'obtention de diplôme de

MASTER

Filière : Informatique

Spécialité : Systèmes Informatiques Intelligents

Thème

L'analyse des émotions multimodale

Soutenu le : 30 / 06 / 2022

Devant le Jury composé de :

Qualité	Nom et Prénom	Grade	Université
Président	Mr. CHEMAM chaouki	MCB	Chadli Bendjedid El-Tarf
Rapporteur	**Melle. ZIANI amel**	**MCB**	**Chadli Bendjedid El-Tarf**
Examinateur	Mme. MAATALLAH majda	MCB	Chadli Bendjedid El-Tarf

Année Universitaire : 2021/2022

Remerciement

Tous d'abord, je tiens à remercier le bon Dieu de m'avoir, Accordé toute la détermination, la volonté et la force pour que je puisse réaliser ce modeste travail.

Je remercie infiniment mon encadrante Dr ZIANI Amel pour ses conseils.

Puis tout notre respect et nos remerciements vont vers les membres du jury qui vont pleinement consacrer leur temps et leur attention afin d'évaluer notre travail, qui espérons le sera à la hauteur de leur attente.

Je remercie mes très chers parents, qui ont toujours été là pour moi.

Et mes deux frères et mes sœurs et mon fiancée.

Enfin, nous remercions tous les professeurs du département de l'informatique de l'université Chadli Bendjedid.

Dédicace

Je dédie ce modeste travail :
A ma chère maman et mon cher papa que
J'adore pour leur soutien et leur
dévouement tout au long de mes études.
A mes chers adorables frères que j'aime.
A mes chers sœurs que j'aime;
Et a mon fiancé;
Et a tous mes amis (es).

Résumé

L'être humain reflète constamment ces sentiments et ces émotions, qui sont des composantes cruciales de l'existence. afin d'effectuer le calcul émotionnel dans le but de reconnaître les émotions et d'analyser les sentiments, certains faits qui peuvent être appelés données émotionnelles sont nécessaires pour analyser ces émotions et sentiments. nous appelons ces types de données des modalités. Il peut s'agir de la parole, du texte ou des expressions faciales.
Ce projet a pour but d'implémenter plusieurs modèles de reconnaissance d'émotions et d'analyse de sentiments et de comparer entre les différents résultats obtenus, et ensuite faire une fusion entre ces deux approches que l'on appelle la fusion multimodale afin d'essayer d'améliorer ces derniers tout en utilisant les principes fondamentaux du Deep Learning implémentés avec le langage python. Concernant la reconnaissance des émotions, nous avons choisis les architectures CNN et concernant l'analyse des sentiments dans le texte nous avons choisis l'architecture CNN et LSTM .

Mots clés : La reconnaissance des émotions, L'analyse des sentiments, Données émotionnelles, La fusion multimodale, Deep Learning, CNN, LSTM .

Abstract

The human being constantly reflects these feelings and emotions, which are crucial components of existence in order to perform the emotional calculation in order to recognize emotions and analyze feelings, some facts that can be called emotional data are needed to analyze these emotions and feelings.we call these types of modality data. It can be speech, text or facial expressions.
This project aims to implement several models of recognition of emotions and analysis of feelings and to compare between the different results obtained, and as a result to achieve a fusion between these two approaches that we call multimodal fusion in order to try to improve these while using the fundamental principles of Deep Learning implemented with the python language. Regarding the recognition of emotions, we have
choose the CNN architectures and concerning the analysis of feelings in the text we chose the CNN and LSTM architecture.

Key words : Emotion analysis, Sentiment analysis, emotional data, Multimodal fusion, Deep Learning, CNN, LSTM .

تلخيص

يعكس الإنسان باستمرار هذه المشاعر والعواطف، وهي مكونات حاسمة للوجود. من أجل إجراء الحساب العاطفي من أجل التعرف على المشاعر وتحليل المشاعر، هناك حاجة إلى حقائق معينة يمكن تسميتها بالبيانات العاطفية لتحليل تلك المشاعر والمشاعر. نسمي هذه الأنواع من طرائق البيانات. يمكن أن تكون هذه تعابير الكلام أو النص أو الوجه.
يهدف هذا المشروع إلى تنفيذ العديد من نماذج التعرف على المشاعر وتحليلها والمقارنة بين النتائج المختلفة التي تم الحصول عليها، ثم دمج هذين النهجين المسمى الاندماج متعدد الوسائط من أجل محاولة تحسينها مع استخدام المبادئ الأساسية للتعلم العميق المنفذ مع لغة الثعبان. فيما يتعلق بالاعتراف بالمشاعر، اخترنا بنية *CNN* وفيما يتعلق بتحليل المشاعر في النص اخترنا بنية *CNN* و*LSTM*

Table des matières

Liste des figures

Liste des tableaux

Liste des acronymes

CNN	Convolutional Neural Network
SVM	Support Vector Machine
Hz	Hertz
CK+	Cohn-Kanade dataset
NN	Neural Network
IA	Neural Network
DL	Deep Learning
MLP	Multi layer perception
DNN	Deep Neural Network
TL	Transfer Learning
GRU	Gated Recurrent Unit
LSTM	Long-Short Term Memory
GAN	Generative Adversarial Network
FER	Facial Expression Recognition
GPU	Graphics Processing Units

Le calcul émotionnel est une expression composée de deux termes : « calcul », qui signifie mesure ou calcul, et « affectif », qui signifie émotion.L'étude des émotions et des sentiments est une méthode basée sur le calcul affectif qui vise à créer des systèmes capables d'identifier et d'analyser les émotions humaines. Cette analyse utilise des données pour le traitement, comme le texte, la parole et les expressions faciales. Il est également possible d'utiliser plus d'une modalité tout au long de l'analyse, appelée analyse multimodale [1].

Ces études sont encore utiles aujourd'hui puisqu'elles peuvent découvrir et résoudre des problèmes majeurs dans un environnement spécifique, comme la prévention du suicide dans n'importe quelle communauté d'une part, et pour le bénéfice des entreprises et des clients en ce qui concerne les alertes produit ou service publicisé d'autre part.

Problématique et motivation

L'identification des émotions et l'analyse des sentiments nécessitent des ensembles de données vastes, bien présentés et efficaces en termes de taille, de présentation et d'efficacité. En ce qui concerne la détection des émotions, les données utilisées sont des images, qui doivent avoir des expressions faciales non ambiguës afin d'accomplir un apprentissage efficace et des prédictions précises. Dans le cas de l'analyse émotionnelle, les données modifiées sont représentées sous forme textuelle, et plusieurs facteurs influencent sur les prédictions, comme l'orthographe, le langage et le langage utilisé (courant, soutenu, etc.).

Même après avoir traité toutes les données, il peut être difficile de trouver des exemples d'ensembles de données; il est donc essentiel de trouver une solution à ce problème et de s'efforcer d'améliorer les résultats obtenus.

Afin d'améliorer la précision des prédictions faites en utilisant deux types de données différents, fusion multimodale entre les résultats de l'analyse des sentiments (dans le texte) et la reconnaissance des émotions (dans les images) est suggéré comme remède à la mauvaise qualité des résultats obtenus en utilisant un ensemble de données de modalité unique.

Contenu du mémoire

Ce mémoire est organisé comme suit :

Chapitre 1 : La reconnaissance d'émotions et l'analyse des sentiments : notions de base et concepts.

Premièrement, nous allons introduire le domaine de la reconnaissance des émotions et d'analyse des sentiments en présentant les concepts de bases de chacune des deux, et nous allons entamer le principe de la fusion multimodale ainsi que certaines de ces techniques.

Chapitre 2 : Le Deep Learning pour la reconnaissance des émotions et l'analyse des sentiments .

Ce chapitre est consacré au deep learning dans lequel nous allons présenter les bases de ce domaine d'intelligence artificielle ainsi que les techniques de classification utilisées.

Chapitre 3 : Conception et implémentation

Dans ce dernier chapitre, nous combinerons les aspects conception et mise en œuvre de la solution proposée, y compris les différents résultats reçus pour chaque modèle appliqué, ainsi que des tableaux de comparaison entre eux, et, bien sûr, la mise en œuvre de la solution proposée.

1. Introduction

L'identification et l'analyse des émotions deviennent importantes à l'ère de l'intelligence artificielle et de l'Internet des objets [2].
L'émotion humaine peut être reconnue à travers une variété de modalités, y compris le mot, qui fournit un moyen direct de communication entre les humains et permet la transmission intuitive des émotions; L'objectif est de modéliser et d'identifier l'émotion à partir de ces expressions visuelles visibles, ainsi que l'étude des expressions faciales, qui est également considérée comme une façon de reconnaître les émotions humaines. L'émotion humaine, d'autre part, est un phénomène complexe.
En effet, il est concevable d'intégrer les nombreuses études indiquées ci-dessus, en fonction de plus d'une modalité ; il s'agit d'une analyse multimodale, qui a le potentiel d'augmenter significativement par rapport aux travaux récents [3].
Nous expliquerons ce que signifie reconnaître les émotions, analyser les sentiments et les opinions, et fournir les processus qui caractérisent les nombreuses phases qui conduisent à la reconnaissance des émotions à partir d'un visage ou d'une voix, ainsi que la reconnaissance des sentiments à partir d'un texte dans ce chapitre.

2. Qu'est-ce qu'une émotion

Les expressions faciales sont influencées par une variété de facteurs, y compris l'émotion.
Les émotions peuvent être exposées de diverses façons, y compris le langage corporel, la parole et les expressions faciales.
Une émotion est souvent accompagnée d'une expression faciale (dont la force peut être influencée dans une certaine mesure par les individus), mais l'inverse n'est pas vrai : il est possible d'imiter une émotion sans la ressentir. Bien que les manifestations individuelles et culturelles varient, il y a un nombre limité d'émotions mondialement reconnues. [3]

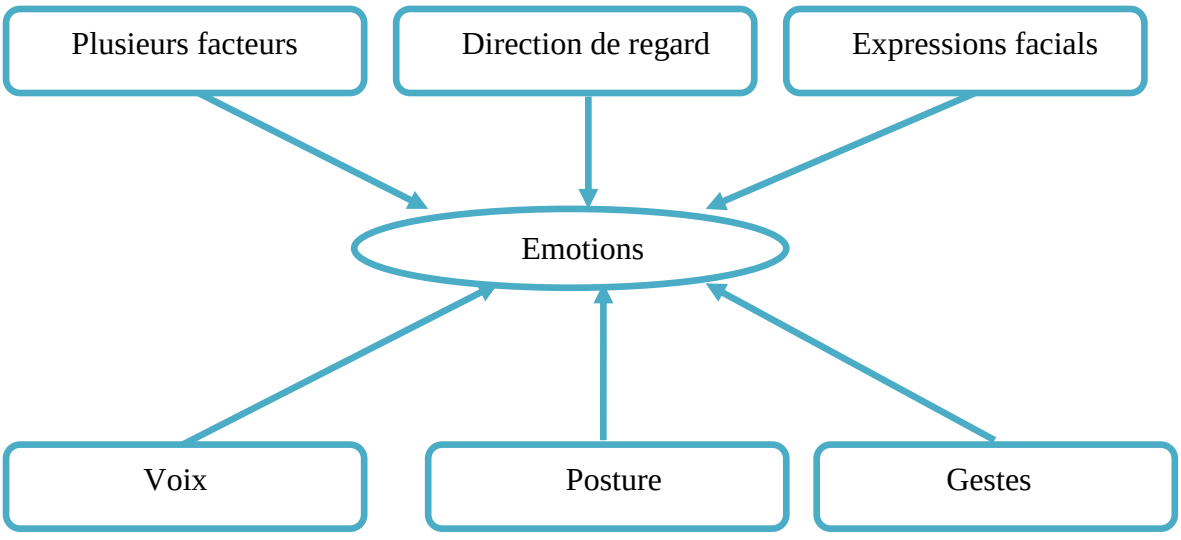


Figure.1.1 Générateurs de l'emotion[3].

Le psychologue Robert Plutchik a étudié la question en profondeur et a trouvé six émotions primitives[4]:
- **La joie :** c'est une émotion agréable , être heureux implique d'être content, content, ravi, joyeux, heureux...
- **La tristesse :** est un sentiment déprimant, elle implique d'être blessé, insatisfait, désespéré, et démuni...
- **Le dégoût** : est une émotion désagréable , blessé, méprisé, ou effrayé...
- **La peur :** est une émotion négative, l'anxiété, terrifiée, suspecte, bouleversée, inquiète... sont synonymes de peur.
- **La colère :** est une émotion déprimante, Être agité, agressif, furieux, dérangé, en colère et hostile sont tous synonymes de colère.
- **Surprise :** une émotion mitigée qui peut être désagréable ou positive selon le scénario.

La surprise doit être agitée, déconcertée, ou perturbée, mais elle peut aussi signifier être éblouie, surprise...

3. Qu'est-ce qu'un sentiment

Bien que le terme "sentiment" ait été initialement utilisé pour exprimer le sens physique du toucher à la suite de l'expérience ou de la perception, il est également utilisé pour désigner une variété de sentiments différents, y compris "un sentiment de chaleur" et de sensibilité en général.
La nominalisation du verbe à sentir est le sentiment.
Sentiment était un mot latin qui signifie "sentir," "entendre," ou "entendre et sentir."
Le terme est généralement utilisé en psychologie pour désigner l'expérience subjective consciente d'une personne de l'émotion. La phénoménologie et l'hétérophénoménologie sont des théories philosophiques qui offrent une base pour comprendre les sentiments. De nombreuses écoles de psychothérapie comptent sur le thérapeute pour arriver à une certaine forme de compréhension des sentiments du client, pour lesquels des techniques existent. [5]

4. Qu'est-ce qu'une opinion

Un point de vue que peut être positif, négatif, ou une nuance de ces extrêmes; la polarité d'une opinion est définie par des valeurs décrivant le degré de positivité (ou négativité), par exemple, une valeur entre 1 et 5, où 1 dénote une polarité très négative et 5 dénote une polarité très positive. Cette évaluation suppose que le texte est homogène, qu'il est complètement positif ou complètement négatif (peut-être neutre), selon l'objectif.
- **Accord :** un point de vue favorable dans lequel une personne est d'accord avec au moins une autre personne sur un sujet (définitivement acceptable; le recyclage est devenu un must)
- **Désaccord :** un point de vue négatif de quelqu'un qui n'est pas d'accord. [6]

5. Les émotions dans les images

En 1968, le psychologue (Albert Mehrabian 1968) a noté que les caractéristiques non verbales telles que l'expression faciale et le ton de la voix transmettent les sentiments plus efficacement

que les mots. Il a révélé la méthode de calcul de la contribution de chaque communication à l'effet expressif global : impact total = 7 % verbal + 38 % voix + 55 % visage.
Le visage humain est notre principale façon de connaître les états émotionnels d'une personne en tant que système de communication d'entrées-sorties multi-signaux (Keltner et al. 2003).
Keltner et de nombreux autres chercheurs croient que les signaux faciaux rapides communiquent les émotions (Ekman et Friesen, 2003; Ambady et Rosenthal, 1992; Keltner et Ekman, 2000) et les traits personnels (Ambady et Rosenthal, 1992; Keltner et Ekman, 2000) sur quatre classes de signaux faciaux. (statiques, lents, artificiels, rapides) .
Une image est une abstraction qui peut parfois libérer l'émotion; dans notre cas, cette émotion est identifiée à partir de visages humains, qui ont une variété de caractéristiques qui aident à décider de l'émotion connexe, tels que les sourcils, les yeux, les joues, le front et les lèvres [6].

5.1. Les expressions faciales de base

La détection de l'émotion d'une personne est basée sur des mouvements faciaux visibles.
De plus, lorsqu'une émotion survient, les activités faciales sont stimulées pendant une courte période. Par conséquent, la détection des expressions faciales est considérée comme une étape naturelle avant l'identification des émotions [7].
Les six mouvements faciaux fondamentaux chez l'homme sont représentés dans le diagramme ci-dessous.

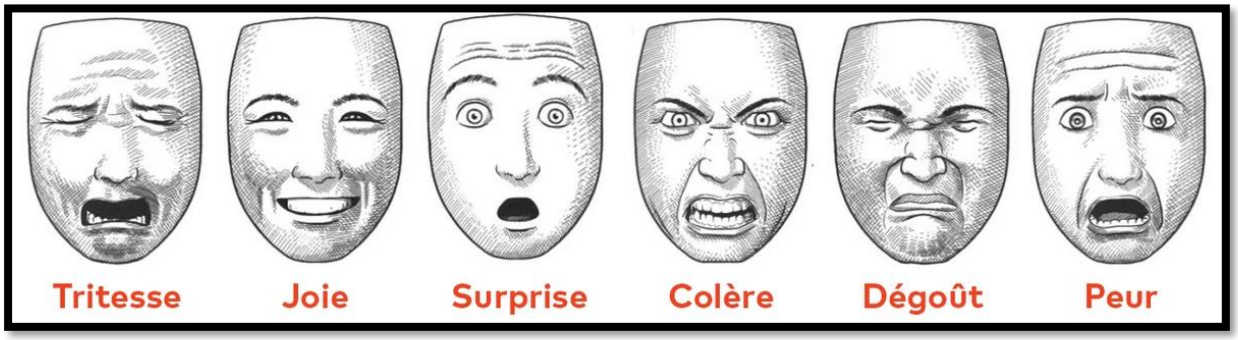


Figure.1.2 Les expressions faciales de base[W1].

5.2. Les caractéristiques des expressions faciales de base

Les visages humains peuvent révéler des émotions à travers divers mouvements du visage qui accentuent la structure des sourcils, des yeux, des joues et de la bouche, en particulier l'apparence (ou l'absence) des dents.

Les qualités les plus essentielles de chacune des six émotions fondamentales mentionnées ci-dessus sont incluses dans le tableau ci-dessous.

Tableau.1.1 Les caractéristiques des expressions faciales de base

L'expression faciale	**Les déformations survenues sur le visage**
Tristesse	• Les coins intérieurs des sourcils sont légèrement élevés pour donner cette forme / \. • Les paupières recouvrent une partie du champ de vision. • La bouche est serrée mais elle descend , légèrement les coins seront

	étirés vers le bas .
Dégout	• Les coins intérieurs des sourcils sont légèrement abaissés . • La bouche est fermée mais on peut remarquer que la lèvre supérieure remonte. • la réduction du champ de vision : l'oeil est à demi - ouvert .
Joie	• Les yeux sont légèrement plissés , c'est - à - dire que la paupière inférieure couvre en partie l'œil . • La bouche est ouverte , c'est un mouvement horizontal . • Les lèvres sont donc étirées , toujours dans un mouvement latéral La personne peut montrer les dents si elle le désire , donc on parle d'un mouvement vertical de la bouche .
Colère	• La paupière recouvre une partie de l'œil donc les yeux seront presque fermés . Quant à la bouche , • elle reste fermée mais assez serrée , sinon , elle s'ouverte verticalement. • Les sourcils ont tendance à se rejoindre , ils sont froncés , plissés . De plus leur partie intérieure est abaissée légèrement.
Peur	• Les yeux sont grands ouverts , écarquillés . • Ce mouvement des yeux a pour effet le redressement des sourcils . • La bouche est ouverte mais cela reste néanmoins un mouvement horizontal . Les lèvres sont donc étirées , toujours dans un mouvement latéral La paupière est entièrement levée . • La pupille est visible dans sa totalité . • Le champ de vision est au maximum . La personne semble fixer quelque chose comme si elle ne pouvait s'en détacher
Surprise	• Les yeux sont grands ouverts . • La bouche est ouverte verticalement . • Les sourcils sont soulevés

5.3. Le processus général de la reconnaissance des émotions dans les images :

a. La détection de visage

Est de découvrir s'il y a des visages dans une image .La plupart des approches d'analyse faciale exigent cela comme une étape de prétraitement essentielle et de base. Les techniques de reconnaissance des modèles sont les plus couramment utilisées. En effet, le défi peut être considéré comme la détection de caractéristiques communes à tous les visages humains : il s'agit de comparer une image à un modèle générique d'un visage et de déterminer s'il y a ou non une correspondance. Le nombre de faces d'une image est indiqué par la sortie d'un détecteur de visage. De plus, la majorité des détecteurs de visage d'aujourd'hui fonctionnent également comme des localisateurs de visage, renvoyant la position des visages identifiées (une boîte englobante par exemple) [8].

Les principales difficultés sont la robustesse aux différentes identités, poses du visage, expressions faciales et aux variations d'illumination [9] [10].

b. Extraction des caractéristiques d'un visage

Suite à la détection d'un visage dans une image, l'étape suivante consiste à extraire les traits du visage représentés, souvent appelés points distinctifs ou traits du visage ("Points de repère"). Ces points sont utilisés pour encadrer les yeux, les lèvres, le nez et les sourcils, entre autres. Un rectangle enveloppant boîte retournée par un détecteur de visage qui localise ce dernier est généralement le point de départ pour l'identification de la caractéristique faciale. Les caractéristiques géométriques telles que les formes des composantes faciales, les distances faciales, etc., peuvent être extraites et utilisées pour calculer les emplacements ou les traits d'apparence.

Le système a du mal à extraire les traits du visage car il y a tellement de variété dans les types de visage[11].

c. Classification des expressions

Dans un système de reconnaissance faciale, c'est la phase finale. Reconnaître la collection de six expressions de proto-type est la première étape. L'étude dans ce domaine est divisée en trois catégories : les méthodes globales, les approches locales et les approches hybrides. Chaque technique a ses propres avantages et limites en ce qui concerne les circonstances ambiantes, les changements d'échelle, les orientations de l'image, les positions de tête, et ainsi de suite[12].

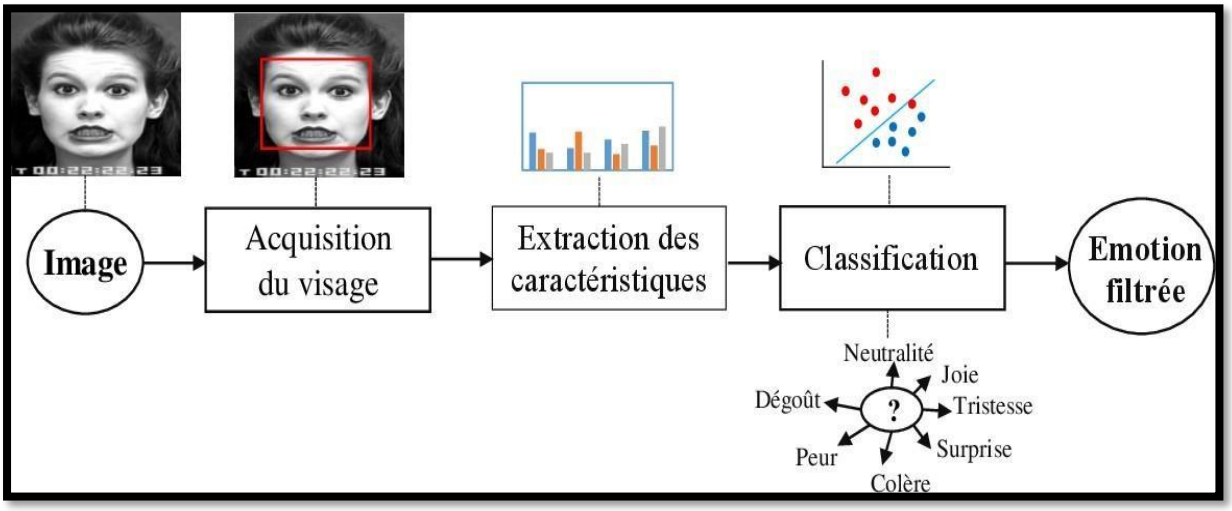


Figure.1.3 Composantes du système d'analyse des expressions faciales [12].

6. Les sentiments dans le texte

De multiples données textuelles ont circulé sur le web depuis sa création, ce qui explique le lancement de diverses recherches sur l'exploration de ces données afin de contribuer la majorité du temps à la prise de décision. Est-il possible d'avoir un sentiment mondial sur le bonheur d'un certain produit par les clients d'une entreprise de fabrication ? C'est une question fondamentale à laquelle on peut répondre dans le domaine de l'analyse des sentiments par l'intelligence artificielle.

Il y a un film à voir ?

Y a-t-il un endroit où aller ?

L'analyse des sentiments et l'étude des points de vue sont des approches d'IA qui extraient les sentiments d'un groupe de personnes et explorent leurs opinions, ce qui peut nous aider à déterminer si un produit est excellent ou horrible, par exemple :

ses caractéristiques, ainsi que des informations complémentaires à son sujet les deux méthodes d'examen des sentiments, à savoir explorer les points de vue et d'enquêter sur les sentiments, sont décrites dans la figure ci-dessous[13].

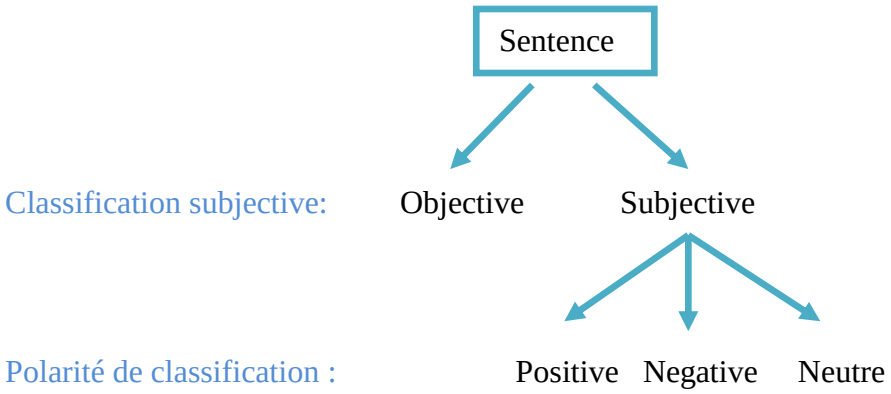


Figure.1.4 Les points de vue et d'enquêter sur les sentiments[14].

a. La classification des expressions d'un document :

Cette analyse classe les expressions d'un document source en deux types :
- Les faits (expressions objectives).
- Les opinions (expressions subjectives de sentiments) [13].

Par conséquent, il faut bien savoir qu'il n'est pas aussi facile de réaliser cette identification et extraction des faits/opinions [13].
- La classification des expressions citée précédemment en faits ou opinions.
- La classification de polarité des opinions qui peuvent être donc positives ou négatives

Une fois la tokénisation terminée, l'étape suivante consiste à supprimer toute information superflue de l'analyse, comme les symboles, les nombres, la ponctuation, et les mots vides sans signification, tels que "à" "nous" "un" et "le."
Enfin, les mots avec des fins de conjugaison, le pluriels, et d'autres suffixes ou préfixes, tels qu'extrait, extrait, extrait, doivent être transformés afin de compter le nombre d'occurrences d'un terme spécifique dans un texte.

- **Le prétraitement :**

Pour commencer, les données textuelles sont prétraitées pour séparer les phrases en un seul ensemble de mots séparés individuellement, un processus connu sous le nom de tokenisation. Il est important de noter que lorsque nous travaillons avec des mots individuels et séparés, nous ignorons la relation entre les mots qui composent les différentes phrases.
Lorsqu'on évalue les sentiments dans des ensembles de données aussi volumineux pour étudier les termes des différents articles, ainsi que leur incidence, cette approche est la plus couramment utilisée et la plus pratique.
Une fois la tokénisation terminée, l'étape suivante consiste à supprimer toute information superflue de l'analyse, comme les symboles, les nombres, la ponctuation, et les mots vides sans signification, tels que « à » « nous » « une » « et » « les ».

Enfin, les mots avec des fins de conjugaison, le pluriel s, et d'autres suffixes ou préfixes, tels que ´extraire´, ´extraite´, ´extraction´, doivent être transformés afin de compter le nombre d'occurrences d'un mot spécifique dans un texte[13].

- **L'extraction des caractéristiques :**

Cette phase consiste simplement à créer une liste de mots extraits du texte afin de construire un vecteur d'attributs qui peut être utilisé pour appliquer la classification sélectionnée[13].

- **La classification :**

Le classificateur est conçu pour évaluer le sentiment d'une phrase ou d'un ensemble de phrases, c'est-à-dire si une personne parle favorablement, négativement ou neutrement , dans cette phase, qui représente la dernière partie de l'analyse des sentiments à partir d'un texte[13].

7. L'analyse des émotions multimodale

Seuls quelques travaux dignes de mention ont abordé l'analyse des émotions multimodal, la majorité des travaux importants se concentrant sur l'évaluation du sentiment dans les vlogs.
Morency et al., au meilleur de connaissance[14].
Il y a quatre phases dans le processus de reconnaissance d'émotion multimodale[15] [16]:
L'étape initiale consiste à identifier l'ensemble de données qui sera utilisé pour exécuter la thérapie, ce qui peut inclure :

- L'enregistrement des individus dans divers états émotionnels
- Reconnaissance faciale
- Extraction de voix
- Reconnaissance de texte...

Ensuite, le trait qui ont le lien le plus fort (rapport réciproque) avec les émotions sont découverts et récupérés. Ces dernières sont les qualités visuelles, auditives ou textuelles extraites des ensembles de données précédents. La prochaine étape consistera à combiner les nombreux attributs, ce qui se traduira par un modèle plus efficace.
Il peut y avoir de caractéristique non pertinentes dans le vecteur caractéristique qui rendent un modèle plus complexe . En conséquence, il est préférable d'appliquer des techniques de réduction de la dimension de sorte que le modèle final deviendra plus simple.
La dernière étape après la réduction de la dimensionnalité du modèle est l'étape de classification, qui en déduit l'émotion finale. Pour assurer la robustesse du modèle , il est nécessaire de choisir un modèle de classification très efficace .

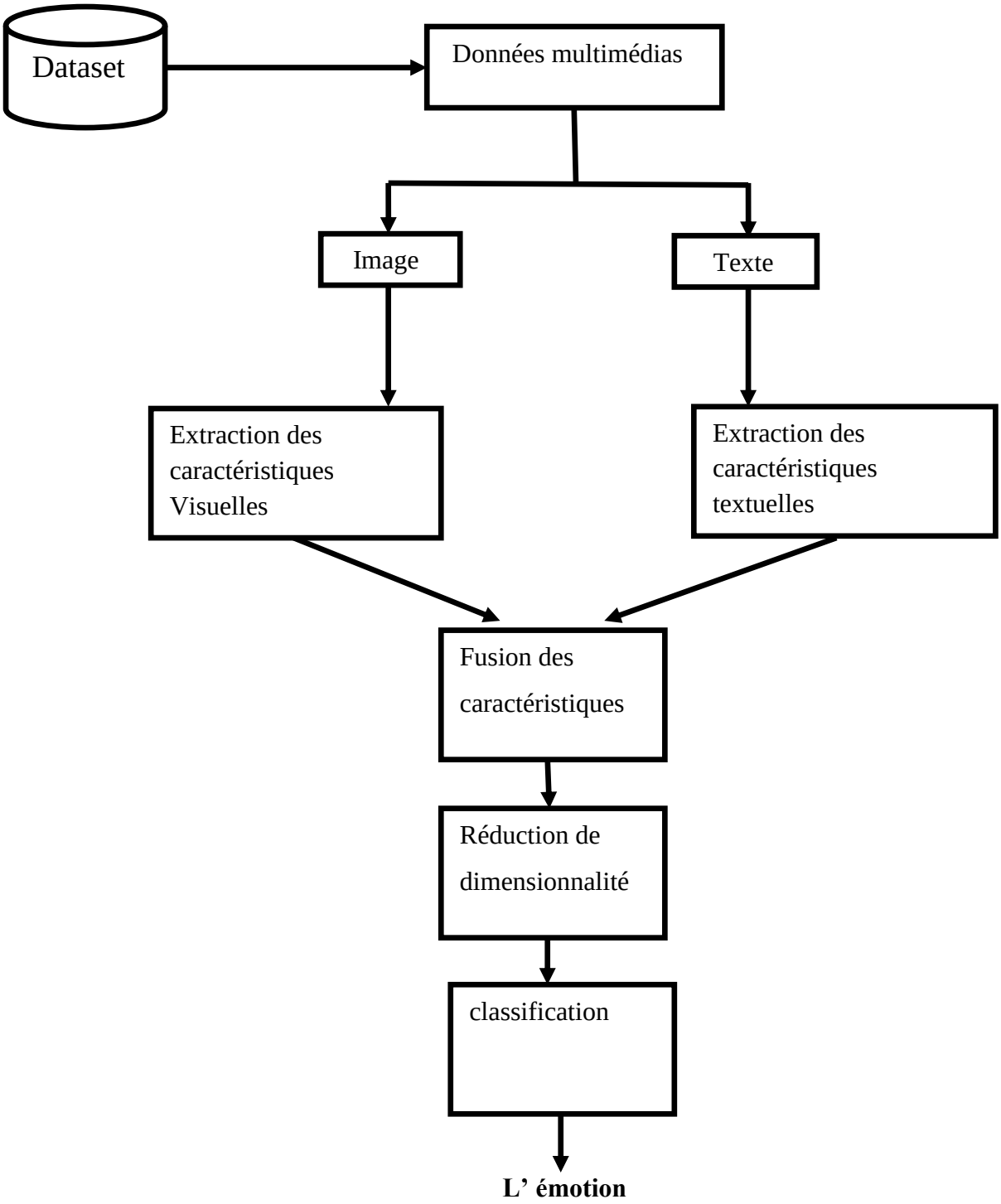


Figure.1.5 Le processus général de l'analyse multimodale [15]

8. Quelques travaux réalisés dans le domaine de la reconnaissance des émotions et de l'analyse des sentiments :

Dans cette partie, nous allons passer en revue différents ensembles de données qui sont utilisés pour l'identification et l'analyse des émotions dans une modalité et des contextes multimodaux.
Les données utilisées dans un calcul émotionnel peuvent être classées comme naturelles, selon qu'elles sont auditives ou visuelles. En ce qui concerne les bases de données textuelles, elles reflètent souvent les commentaires des utilisateurs exprimés au public sur diverses plateformes et

réseaux sociaux à l'égard d'un produit ou d'un service, de diverses questions politiques ou de sous-titres de films ou d'émissions de télévision.
Nous avons énuméré certains des ensembles de données utilisés dans diverses études dans le domaine de l'identification des émotions et de l'analyse des sentiments dans le tableau ci-dessous.
Nous avons décrit la modalité qui compose chaque collecte de données, ainsi que la taille du système et le type de données.

Tableau.1.2 Les bases de données les plus connues dans le domaine de la reconnaissance des émotions et de l'analyse des sentiments [16]

Nom	**Modalité**	**Taille**	**Type de données**
IMDB review	Texte	50000	Critiques de films
Blogs	Texte	252	Revues de produits et poste politique
Geneva Affective Picture Database (GAPED)	Image	754 photos	Photos
Japanese and Caucasian Facial Expressions of Emotion (JACFEE) (Biehl et al., 1997)	Image	56	Facial expression
Cohn-Kanade AUCoded Facial expressions database (CK+)	Vidéo	97 sujets	Expressions faciales
Humaine database	Audio Vidéo Texte	50 clips	Agi et naturel Enregistrements
ICT-MMMO (Wollmer et al 2013)	Audio Video Text	370 videos	visualiser des vidéos

Le tableau ci-dessous donne un aperçu de certaines recherches effectuées par différents chercheurs au cours des années précédentes, ainsi qu'une brève description de la technique utilisée et de l'ensemble de données, De plus, décrire le processus de classification et les taux d'exactitude liés à chaque emploi.

Tableau.1.3 Les principaux travaux réalisés dans le domaine de la reconnaissance des émotions et de l'analyse des sentiments [16]

Chercheurs	**Modalité**	**Dataset**	**La technique de classification utilisée**	**Taux de précision**
Richard et al 2013	Text	Stanford Sentiment TreeBank	Deep Learning	80.70%
Hu et al 2015	Text	Client reviews on TripAdvisor and Amazon	Deep Learning	87.50%
Guntuku et al 2016	Image	Flickr images	Fonction de sélection Régression ordinale	80%
Kaya et al 2017	Video	EmotiW 2015 EmotiW 2016	CNN	52.11%
Gupta 2017	Vidéo	MMI dataset	NN	66,15%
Sawata et al 2017	Audio	Multiple Features Database (MFD)	Noyau discriminatoire Préservation des localités Corrélation canonique Analyse (KDLPCCA).	81.4%

9. Le concept de la fusion multimodale

Les techniques de fusion multimodale (MF) sont utilisées dans une variété de domaines d'étude.
La fusion précoce est un type d'approche de fusion qui intègre de nombreuses modalités pour créer de nouvelles modalités en ajoutant des données liées de façon latente.
La fusion précoce est un type d'approche de fusion qui intègre de nombreuses modalités pour créer de nouvelles modalités en ajoutant des données liées de façon latente.
Au cours des premières étapes de l'analyse multimédia, il est mis en œuvre dans l'espace de présentation[17].

9.1. Les techniques de la fusion multimodale

La fusion multimodale est un terme utilisé pour plus d'une technique. Dans cette section, nous allons examiner certains d'entre eux et donner un bref aperçu de leurs architectures.

a. La fusion des caractéristiques

Une unité d'analyse unique est utilisée pour exécuter l'opération d'analyse dans la fusion au niveau des caractéristiques. Les caractéristiques collectées à partir des données d'entrée sont regroupées ici avant d'être transmises à l'unité d'analyse. Comme la majorité de la fusion de fonctionnalités est effectuée en série ou en parallèle, le résultat final est un vecteur de grande dimension. Il s'agit d'un problème important en matière de fusion des fonctions. Cette difficulté

peut être résolue en utilisant une technique de sélection/transformation des caractéristiques pour réduire les caractéristiques extraites.
- En termes de données visuelles, la fusion multimodale combine plusieurs caractéristiques, telles que la couleur de la peau, les caractéristiques, et ainsi de suite, en un vecteur appelé caractéristiques vectorielles, qui est énorme en taille et sera envoyé dans l'unité de détection du visage.
- Les qualités textuelles s'expriment soit directement sous la forme d'un texte, soit en général à partir de la conversion d'une déclaration orale en texte.
- En ce qui concerne les qualités audio, la transformation de Fourrier est couramment utilisée (FFT) par fusion précoce de MEG, de réponses périphériques, d'activité faciale et de caractéristiques visuelles auditives basées sur le contenu.
Dans, une fusion de niveau caractéristique est présentée pour améliorer la catégorisation de quatre modalités physiologiques, y compris l'EEG du système nerveux central et le PERI du système nerveux périphérique[16].

b. La fusion des décisions

Dans cette méthode, les jugements sont d'abord portés sur les unités d'analyse en fonction des caractéristiques individuelles. Pour clarifier, au niveau sémantique, une décision est le résultat d'une unité d'analyse. Les décisions locales sont ensuite combinées en utilisant une unité de fusion de décision pour créer un vecteur de décision fusionné. Le vecteur de décision fusionné est ensuite évalué pour en arriver à un choix final concernant la tâche à accomplir. Par rapport à la fusion de fonctionnalités, la fusion de niveau de décision offre un certain nombre d'avantages. Au niveau sémantique, par exemple, les jugements ont normalement la même représentation; cependant, les caractéristiques de diverses modalités (comme l'audio et la vidéo) peuvent avoir des représentations distinctes. En conséquence, faire la fusion de décision devient facile.
En outre, l'évolutivité (c.-à-d. une amélioration ou une détérioration en douceur) des modalités utilisées tout au long du processus de fusion est possible avec la fusion au niveau de la décision.
Dans le cas de la fusion des fonctions, cependant, les deux éléments susmentionnés sont difficiles à réaliser.
Afin d'accroître l'exactitude de la classification, la fusion au niveau de la décision a été effectuée en intégrant les résultats des classificateurs en utilisant la croyance de Bayes pour l'intégration de l'EEG et de l'information périphérique. Un algorithme de fusion de décision a été utilisé pour fusionner les résultats de chaque modalité pour une plus grande précision dans la catégorisation de l'EEG, périph données physiologiques globales et MCA[16].

c. Fusion Multimodal Hybride

L'approche de fusion multimodale hybride combine les techniques de fonctionnalité et de niveau de décision, lui donnant le meilleur des deux mondes (fusion précoce et tardive). En conséquence, de nombreux types de problèmes analytiques multimodaux sont inspirés pour appliquer la technique de fusion hybride par les chercheurs. a élaboré une stratégie de fusion hybride pour l'indexation sémantique des ressources multimodales à l'aide d'indices visuels et textuels, en tenant compte de la fusion précoce normalisée et de la fusion tardive contextuelle. Chaque entrée du vecteur combiné est normalisée avant d'être fusionnée comme un acte de fusion précoce normalisée dans cette méthode. Pour exploiter le lien contextuel, un second classificateur de couche basé sur SVM est utilisé. Une

stratégie de fusion basée sur le noyau SVM a également été étudiée, où les fonctions du noyau sont déterminées en fonction des modalités[16].

d. Fusion au niveau du modèle

La fusion au niveau du modèle utilise une fusion transparente des données pour intégrer les données expérimentales de modalités distinctes. Ce type de modèle a été créé par des chercheurs pour répondre à leurs besoins d'étude et à l'espace problématique connexe. Pour modéliser les caractéristiques de corrélation de trois composants HMM, les scientifiques ont utilisé des flux audio-visuels et un modèle Hidden Markov (HMM) triplé. Les auteurs ont suggéré un modèle de Markov caché fusionné multi-flux pour la reconnaissance des effets audio-visuels. En utilisant l'entropie maximale et les principes d'information mutuelle maximale, ce système a créé une connexion idéale entre plusieurs flux. Les auteurs ont proposé de combiner les modalités auditives et visuelles pour l'identification des émotions à l'aide d'un module de réseau neuronal. La topologie du réseau bayésien a été suggérée pour reconnaître les émotions des modalités audiovisuelles. a utilisé une approche probabiliste pour fusionner les deux modalités mentionnées ci-dessus[16].

e. Fusion basée sur des règles

La méthode de fusion fondée sur des règles utilise généralement un certain nombre de règles fondamentales pour fusionner des données multimodales. Les méthodes statistiques fondées sur des règles comprennent la fusion linéaire pondérée (somme et produit), MAX, MIN, AND, OR et le vote majoritaire, pour n'en nommer que quelques-unes. Avant la fusion des données multimodales, des poids normalisés sont attribués à chaque modalité considérée. Par conséquent, comparativement à d'autres méthodes, la méthode de fusion linéaire pondérée est moins coûteuse en calcul. Cependant, pour une exécution optimale, les poids doivent être normalisés adéquatement, ce qui est le fardeau que cette méthode a, en plus du fait qu'elle est sensible aux valeurs aberrantes. La décision prise par la majorité des classificateurs est utilisée dans la fusion des votes à la majorité. Afin d'obtenir des décisions optimales, des règles personnalisées spécifiques à l'application sont élaborées en fonction des données recueillies à partir d'un certain nombre de modalités et de la conséquence ultime probable.
Il existe également des règles personnalisées qui sont élaborées pour une application spécifique.
Lorsque la qualité de l'arrangement chronologique entre les modalités distinctes est acceptable, les méthodes fondées sur des règles donnent de meilleurs résultats[16].

f. Fusion basée sur la classification

Cette méthode de fusion utilise diverses techniques de classification pour catégoriser les observations multimodales en un des nombreux groupes prédéfinis. Les machines vectorielles de support, l'inférence bayésienne, la théorie de Dempster-Shafer, les réseaux bayésiens dynamiques, les réseaux neuronaux et le modèle d'entropie maximale sont parmi les techniques utilisées. En termes d'apprentissage automatique, ces stratégies peuvent être classées comme modèles génératifs et discriminatoires. Un système d'approche de fusion tardive pour détecter les concepts sémantiques dans les données vidéo utilisant une méthode d'apprentissage discriminante pour fusionner plusieurs modalités a été proposé dans une étude. La fusion linéaire gradient-descente-optimisation (GLF) et la fusion non linéaire du super-noyau sont deux

méthodes décrites par les auteurs pour extraire les informations multimodales des données vidéo (NLF). Dans GLF, chaque modalité reçoit sa propre matrice noyau, qui fournit une image partielle de la cible. Après cela, les matrices de noyau séparées sont fusionnées en utilisant des poids optimaux. Pour clarifier, les poids idéaux sont déterminés à l'aide de la technique de descente en pente. Les films cibles sont ensuite classés en utilisant SVM sur la matrice du noyau fusionné. L'approche NLF n'est utilisée que lorsqu'il existe un mélange non linéaire de données multimodales[16].

g. Fusion basée sur l'estimation

Les méthodes de fusion fondées sur l'estimation comprennent le filtre Kalman, le filtre Kalman étendu et les méthodes de fusion par filtre de particules. Le filtre Kalman fonctionne mieux pour les systèmes linéaires, tandis que le filtre Kalman étendu fonctionne mieux pour les systèmes non linéaires. Les filtres à particules sont une méthode de simulation robuste pour obtenir la distribution d'état dans les modèles non linéaires et non-Gaussiens d'état-espace. Pour obtenir une estimation bayésienne optimale, la fusion par filtre à particules nécessite un grand nombre de données. Ces méthodes ont été principalement appliquées avec des données multimodales pour améliorer l'estimation d'état d'un objet en mouvement. Le suivi des objets en est un exemple ; pour ce faire, différentes modalités comme l'audio et la vidéo sont combinées pour estimer la position de l'objet[16].

10.Conclusion

Ce chapitre sert d'introduction au domaine de la reconnaissance des émotions faciales et textuelle, ainsi qu'à l'analyse des sentiments dans le texte. Nous avons commencé par discuter des définitions de base, telles que les définitions des émotions, des sentiments et des opinions, avant de passer au processus général de reconnaissance des émotions dans les images et la voix, ainsi que le processus général d'analyse des sentiments dans le texte. Et nous avons discuté de ce qu'est la fusion multimodale et des diverses disciplines de recherche qu'elle englobe, ainsi que des approches de fusion les plus connues et souvent utilisées, comme la fusion d'entités, la fusion décisionnelle, la fusion hybride et la fusion de modèles.
Dans le chapitre suivant, nous examinerons l'apprentissage profond, qui est un aspect essentiel de l'intelligence et de l'apprentissage automatique, ainsi que ses architectures fondamentales.

1. Introduction

L'intelligence artificielle (IA) et l'apprentissage automatique sont devenus extrêmement populaires ces dernières années en raison de leur intégration dans un large éventail d'applications informatiques. Parmi les différents algorithmes d'apprentissage automatique, le Deep Learning (DL), qui est couramment utilisé dans les dernières années [21].

Nous expliquerons d'abord la relation entre l'intelligence artificielle, l'apprentissage automatique et l'apprentissage profond, ainsi que les différences et les principes de chaque approche. Nous présenterons ensuite quelques techniques de classification du deep learning, telles que le réseau de neurone convolutif et le réseau de neurone récurrents, que nous testerons afin de comparer les résultats obtenus pour chaque technique.

2. Qu'est-ce qu'un réseau de neurone

Un réseau neuronal est l'association de neurones formels d'une manière plus ou moins. Les principaux réseaux se distinguent par leur architecture (nombre de neurones, présence ou absence de boucles de rétroaction dans le réseau), leur niveau de complexité (nombre de neurones, type de neurones (fonctions de transition ou d'activation), et enfin par la cible : les apprentis-supervisés ou non supervisés, l'optimisation, les systèmes dynamiques[W2].

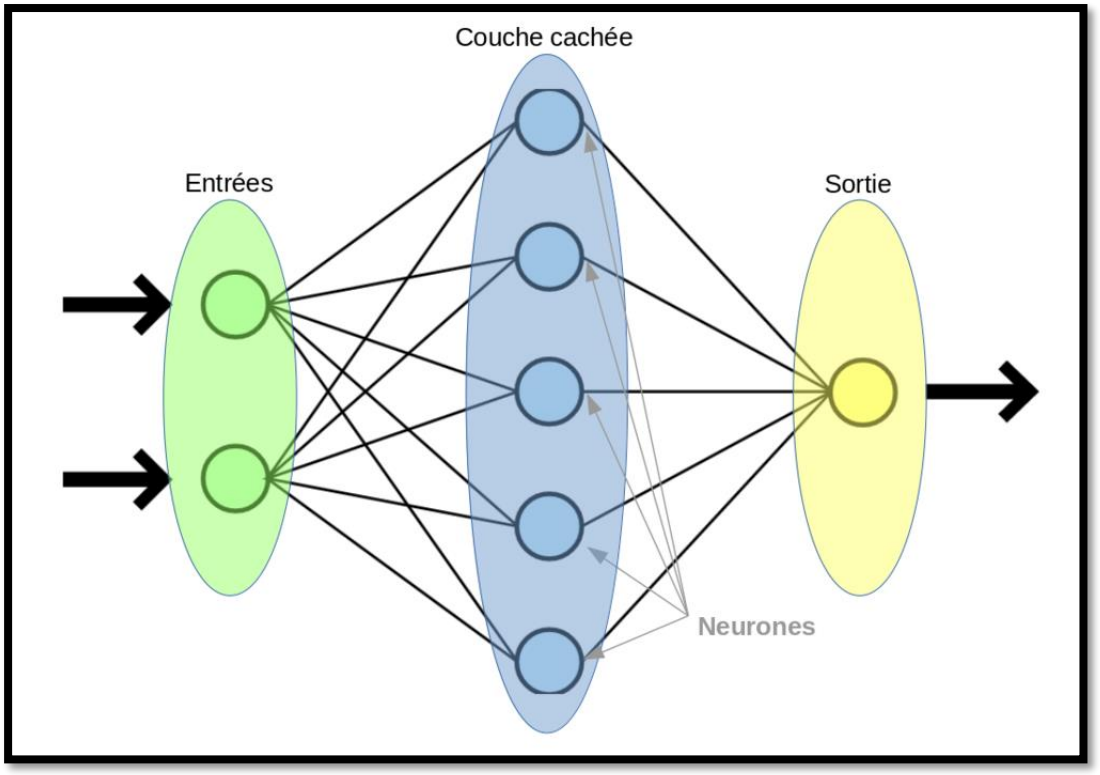


Figure.2.1 L'architecture générale des réseaux de neurones[W2].

2.1. Les avantages des réseaux de neurones

La fascination des réseaux neuronaux découle des caractéristiques qu'ils présentent; néanmoins, ils ont également une variété de limites ou d'inconvénients[18].

Les utilisations potentielles de la méthode neurale peuvent être déduites de ces qualités[18].

a. Avantages :

i. Application des connaissances empiriques : L'apprentissage à partir d'exemples (méthode d'apprentissage empirique) est un processus très facile qui produit de bons résultats comparativement à d'autres stratégies d'apprentissage automatisées[19]

ii. Robustesse : La mémoire est dispersée dans les réseaux neuronaux, et elle est corrélée à une carte d'activation neuronale. Parce que la perte d'un élément n'équivaut pas à la perte d'une connaissance apprise, cette carte fonctionne comme une forme de codage de faits stockés, donnant à ces réseaux l'avantage de résister au bruit (défaillances) [19]

iii. Dégradation progressive : Les réseaux, de par leur nature même, ne fonctionnent pas de façon tout ou rien, et leurs performances tendent à se détériorer dans le temps en cas de problème (bruit, panne, entrée inconnue...). Les réseaux permettent de généraliser les connaissances de la base d'apprentissage et sont moins sensibles aux perturbations que les systèmes symboliques[19]

iv. Parallélisme massif : Les réseaux sont constitués d'un certain nombre d'unités de traitement des données qui peuvent fonctionner en parallèle. Bien que la plupart des implémentations de réseau de connexion se fassent de manière séquentielle sur des simulateurs, il est possible de créer des implémentations (logicielles ou matérielles) qui utilisent la capacité d'activer des unités en même temps[19]

v. Considérations relatives au temps et à la non-linéarité : Les réseaux neuronaux sont un type de réseau informatique.Les avantages de pouvoir tenir compte de la non-linéarité (les fonctions d'activation sont souvent non linéaires). Certains réseaux peuvent également inclure des facteurs temporels (dans le cas de réseaux récurrents) [19]

b. Inconvénients :

i. Difficulté dans le choix de l'architecture et des paramètres : Il n'y a pas de moyen automatisé de choisir l'architecture optimale pour un problème particulier[19] . Algorithmes d'apprentissage Les connexionnistes sont généralement très dépendants de l'état initial du réseau (poids aléatoires d'initialisation) et de la configuration de la base d'apprentissage.

ii. Initialisation et codage Problème : Algorithmes d'apprentissage Les connexionnistes dépendent généralement beaucoup de l'état initial du réseau (poids aléatoires d'initialisation) et de la configuration de la base d'apprentissage. Un mauvais choix de poids pour démarrer le réseau, la technique de codage des données, ou même la séquence de l'entrée pourrait entraver l'apprentissage et causer des problèmes avec la convergence du réseau à une solution satisfaisante [19].

iii. Manque de convivialité : La collection de valeurs de poids synaptiques ainsi que la façon dont elles sont interconnectées connaissance du code reçu par le réseau. Il est très difficile pour une personne de les comprendre immédiatement. Lorsque la connaissance est cachée et incompréhensible pour l'utilisateur ou l'expert. Un réseau ne peut pas expliquer pourquoi il est arrivé à une conclusion particulière [19].

iv. Manque d'exploitation des connaissances théoriques : Les réseaux ne permettent pas l'utilisation des connaissances théoriques sur le sujet de la situation. Ils sont engagés dans la manipulation empirique de la connaissance. La conversion de règles en exemples est une approche simple pour tirer profit des connaissances théoriques (prototypes) [19]

3. Définition de l`apprentissage profond (deep learning) :

L'apprentissage profond est un ensemble d'approches d'apprentissage qui ont contribué à des progrès substantiels en intelligence artificielle au cours des dernières années. Un logiciel examine une collecte de données pour créer des règles pour tirer des inférences à partir de nouvelles données dans l'apprentissage automatique. L'apprentissage profond est basé sur des "réseaux de neurones artificiels", qui sont composés de milliers d'unités (appelées "neurones") qui conduisent des tâches modestes et de base. La sortie d'une première couche de "neurones" est utilisée pour alimenter les calculs d'une deuxième couche, et ainsi de suite.

Par exemple, les premiers niveaux d'unités en reconnaissance visuelle détectent des lignes, des courbes, des angles... les couches supérieures identifient des formes et des combinaisons de formes, d'objets et de situations.L'apprentissage profond a évolué grâce à la puissance informatique et au développement d'énormes ensembles de données (parfois appelés "big data")[20].

3.1. Les performances de l'apprentissage profond :

La majorité des secteurs et des entreprises ont déjà utilisé l'AD dans plusieurs domaines, à condition que la performance humaine et les capacités ne dépassent pas l'AD dans de nombreux domaines [21].

Cela suggère que le DL pourrait être utilisé pour résoudre les problèmes suivants :

- L'expertise humaine fait défaut.

- Mise à jour continue du problème.

- Le problème est énorme, dépassant de loin les capacités de pensée humaine [21].

En matière de performance, nous avons [21] :

- **Une approche d'apprentissage universelle** : La capacité du DL à fonctionner dans pratiquement tous les domaines d'application en fait une technique d'apprentissage universelle.
- **Résilience :** Nous discutons de la robustesse des modèles en ce qui concerne les changements dans les données d'entrée typiques via l'apprentissage automatique basé sur les tâches.
- **Généralisation :** La même technique DL peut être utilisée pour de nombreux types de données et d'applications, un processus connu sous le nom de TL ou Transfer Learning, qui est une option très bénéfique quand il n'y a pas assez de données.
- **Évolutivité :** Ils sont très évolutifs, et des recherches sur l'évolution des Frameworks pour les réseaux ont été menées. Des milliers de nœuds peuvent être construits.

3.2. La catégorisation de l'apprentissage profond :

Il existe trois grands types d'apprentissage profond, selon la façon dont les architectures et les algorithmes sont destinés à être utilisés :

a. Réseaux profonds pour l'apprentissage non supervisé :

Ces réseaux sont conçus pour saisir un niveau élevé d'association entre les données observées à des fins d'analyse ou de synthèse lorsqu'aucune information sur les étiquettes de sortie n'est fournie.[22]

Les auto encodeurs en sont un bon exemple :
Les auto encodeurs sont une sorte de DNN (Deep Neural Network) qui n'a pas de classe et produit des vecteurs de sortie avec les mêmes dimensions que les vecteurs d'entrée. il est fréquemment utilisé pour l'encodage de données.[22]
Un auto encodeur type contient une couche d'entrée (couche L1) qui représente les données ou les vecteurs caractéristiques, une ou plusieurs couches cachées (couche L2) qui représentent la caractéristique modifiée, et une couche de sortie qui correspond à la couche d'entrée (couche L2). L'auto encodeur est appelé deep lorsque le nombre de couches cachées est supérieur à un.
Les dimensions des couches peuvent être modestes (où la compression est le but) ou grandes (lorsque le but est d'élargir la dimension de l'espace) [22].

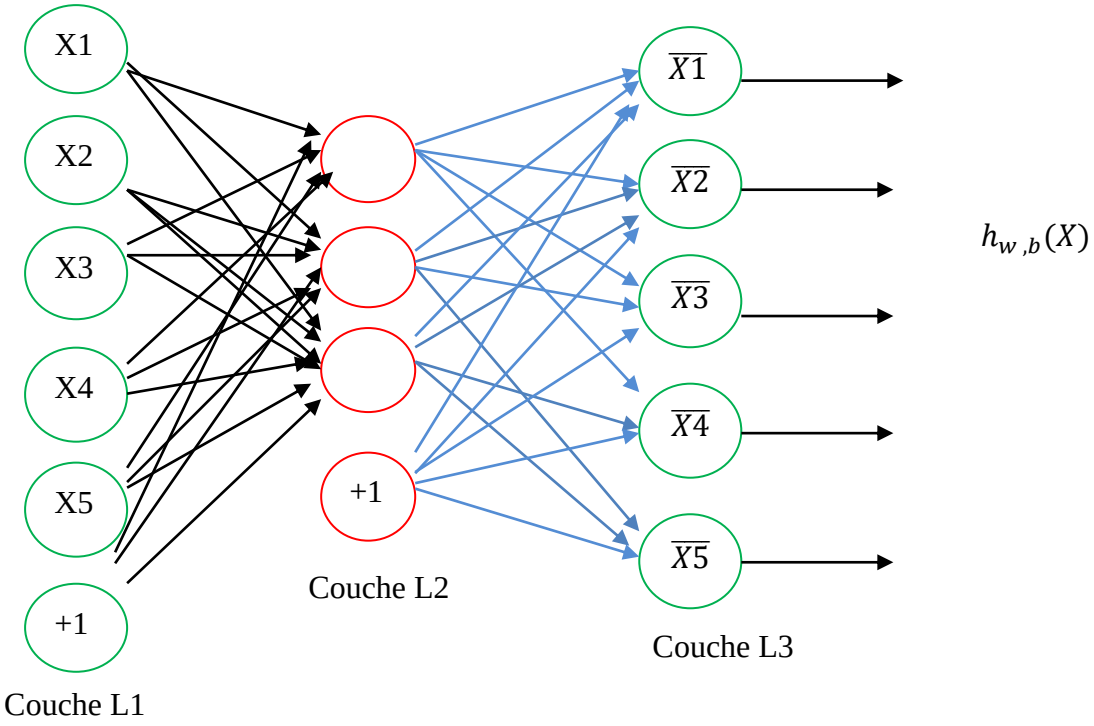


Figure.2.2 L'architecture des autoencoders [22].

Réseaux profonds pour l'apprentissage supervisé : Ces réseaux sont conçus pour donner un pouvoir discriminant direct pour la classification des modèles, souvent en décrivant les distributions postérieures des classes conditionnées par des données observables[22].

Exemple : DSN (Distributed Synchronous Network (Deep Stacking Networks),
DSN (Deep Stacking Networks) a été lancé avec une architecture quelque peu différente de , DNN (Deep Neural Networks). Il est composé de petits sous-réseaux connectés par une seule couche cachée.
Chaque teinte représente un sous-réseau, également connu sous le nom de module.

La sortie de n'importe quel module peut être répliquée à des niveaux plus élevés pour former une architecture complexe (les lignes pointillées indiquent les couches de copie). Dans les modules, la même conception est répétée, avec une couche d'entrée linéaire suivie d'une couche non linéaire couplée à une couche de sortie linéaire. La couche d'entrée et la couche cachée sont reliées par la matrice de poids de la couche inférieure, que nous appelons *W*. La matrice de poids de la couche supérieure, indiquée par la lettre *U*, relie la couche cachée à la couche de sortie[22].

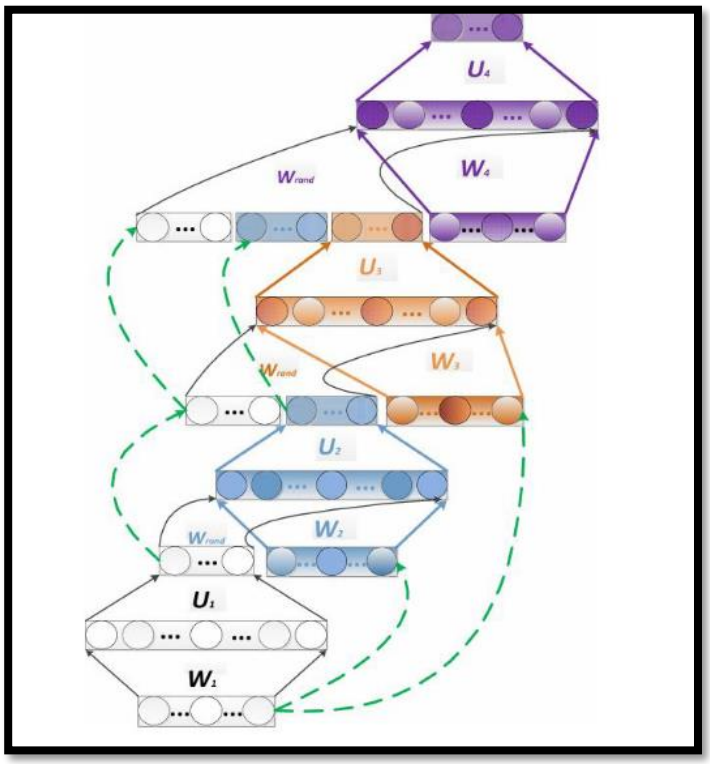


Figure.2.3 L'architecture des réseaux DSN[22].

b. Réseaux profonds hybrides :

L'objectif de cette catégorie est la discrimination, qui est facilitée par les résultats de réseaux profonds non supervisés, tels que les auto-encodeurs.Pour le dire autrement, les autoencodeurs sont utilisés pour apprendre le DNN[22].

- SAE-DNN (Stacked Autoencoder - Deep Neural Network) comme exemple :

SAE-DNN est une combinaison de deux réseaux neuronaux profonds : SAE et DNN. SAE est composé de nombreux niveaux d'auto-encodeurs (voir section 2.9.3.1), avec les sorties de chaque couche couplées aux entrées de la couche suivante.
Ces autocollants servent à déplacer les couches cachées du DNN une par une (Figure 2.4).
La formation sur les RSN comporte deux étapes (figure 2.4). Tout d'abord, "autoencodeur 1" est formé sans supervision. Deuxièmement, après avoir appris, la couche DNN h1 est initialisée en utilisant des poids "auto encodeur 1". Ensuite, les poids de la première "couche H1" deviennent des entrées pour la deuxième couche (h2), et ainsi de suite. L'étude du DNN commence dans la deuxième étape avec l'instructeur[22].

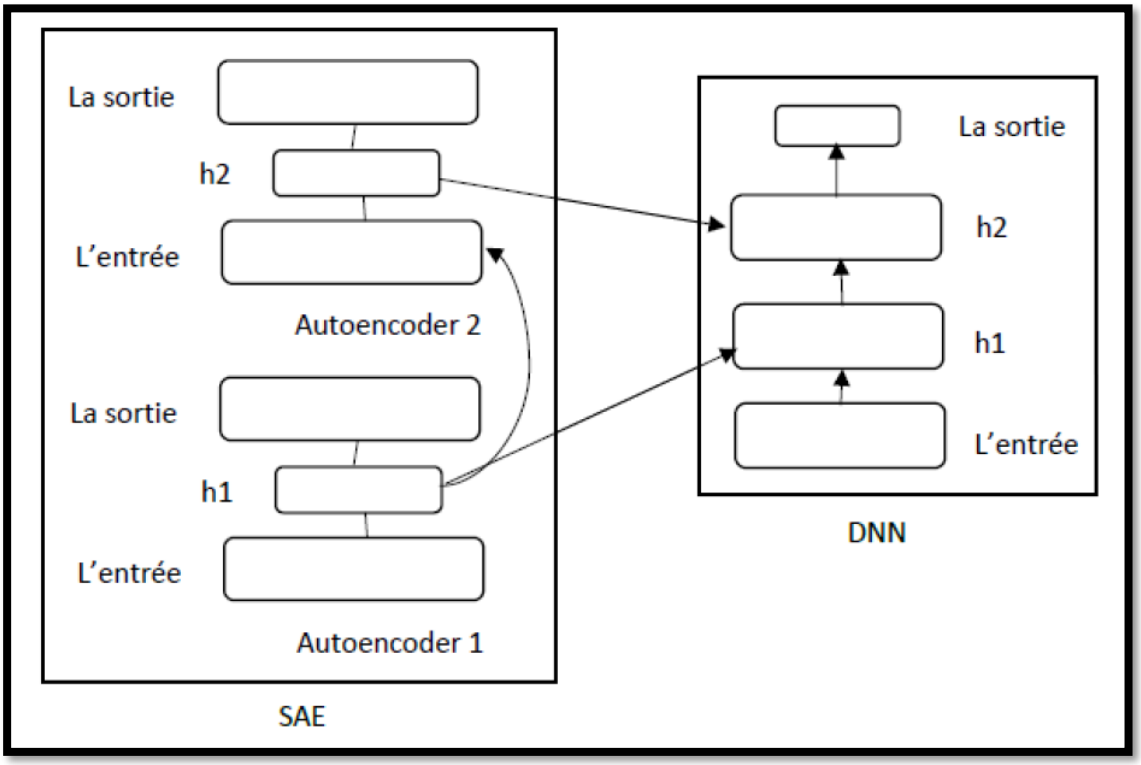


Figure.2.4 l'architecture de SAE-DNN[22].

3.3. Les principales techniques de classification de l'apprentissage profond

Nous décrirons deux architectures de Deep Learning dans ce document; cependant, il y a beaucoup d'autres conceptions qui sont utilisées dans diverses industries; nous nous concentrerons sur les réseaux de neurones convolution (CNN) et (LSTM).[25].

a. L'architecture de réseaux neuronaux convolutifs CNN

Les réseaux neuronaux convolutifs [Hubel & Wiesel 1962] sont des réseaux d'apprentissage inspirés du cortex visuel. Les systèmes de référence [Ying et al. 2018], le traitement du langage naturel [Kim 2014], et la vision via ordina- [Krizhevsky et al. 2012] ont tous utilisé ces réseaux.
En raison de ses qualités inspirées par des images naturelles, leur utilisation dans la vision par ordinateur a été un énorme succès[25].
Le processus de catégorisation à l'aide des méthodes d'apprentissage traditionnelles dans la vision par ordinateur est divisé en deux étapes : ex- caractéristiques et l'apprentissage.
En raison de l'effort manuel, ces qualités sont appelées caractéristiques faites à la main et sont nécessaires dans l'étude des traits discriminants. Ces propriétés sont extraites sans supervision à l'aide des méthodes utilisées. Si certaines caractéristiques de différenciation ont été négligées pendant la phase d'extraction, cette séparation entre les modules d'extraction et de classification peut nuire à la catégorisation des tâches[25].

- **L’architecture générale d’un réseau de neurones convolutif :**

Il existe trois types de couches dans un réseau de neurones convolutifs : convolution, mise en commun et pleine couche connectée. L’architecture d’un CNN est illustrée à la Figure 2.5 [23].

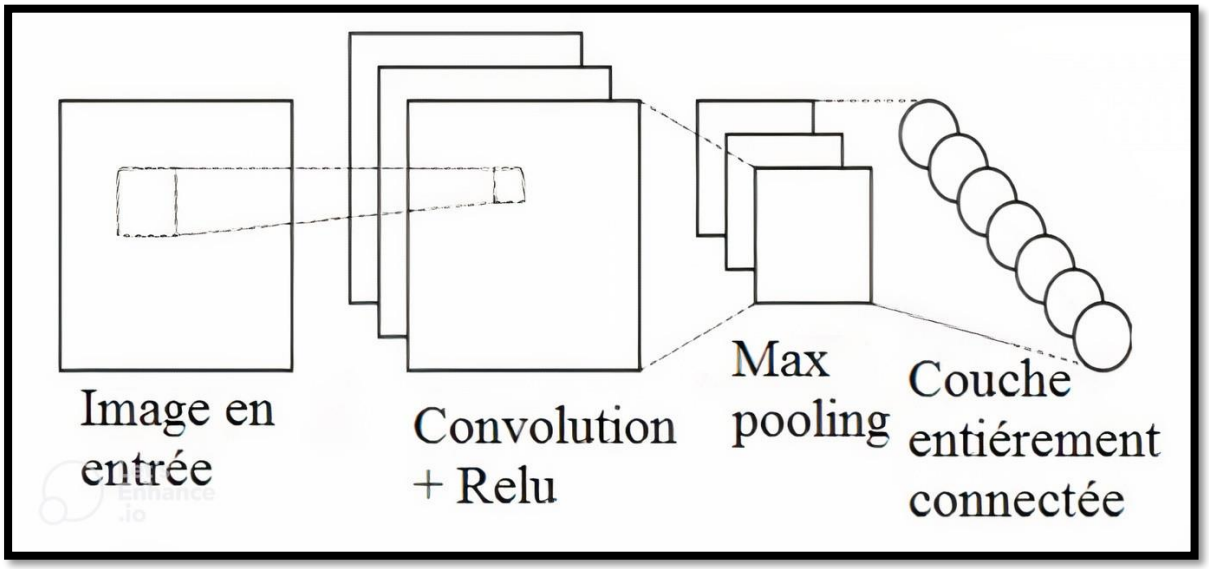


Figure.2.5 L’architecture générale d’un réseau de neurones convolutif [23].

- ✓ **Couche de convolution**

La partie de construction fondamentale d’un CNN définit la couche de convolution.
Pendant l’apprentissage, le but de cette couche est d’extraire implicitement les caractéristiques importantes des images entrantes.Cette couche effectue une opération de convolution entre deux matrices, dont la première représente un sous-ensemble des données d’entrée (re-field ceptif), et la seconde représente un filtre qui contient l’ap- une opération de convolution génère une troisième matrice, qui est référencé par la carte des caractéristiques. La Figure 2.6 représente un produit scalaire entre un filtre et un champ réceptif effectuant une convolution[23].
Les résultats du produit sont ensuite combinés pour former un seul résultat qui apparaît comme une boîte sur la carte des caractéristiques.
Enfin, sur les champs réceptifs restants de la matrice d’entrée, le filtre de poids est glissé par une étape S, et cette procédure est répétée pour tous les autres champs.
La taille de la nouvelle carte des caractéristiques $N^{(t+1)}$ dans une convolution est déterminée par quatre hyperparamètres : la taille de la carte des caractéristiques de l’année ou de la matrice d’entrée $N^{(t)}$, la taille du filtre F, la valeur de l’étape S et la valeur de la marge P (équation 2.1).

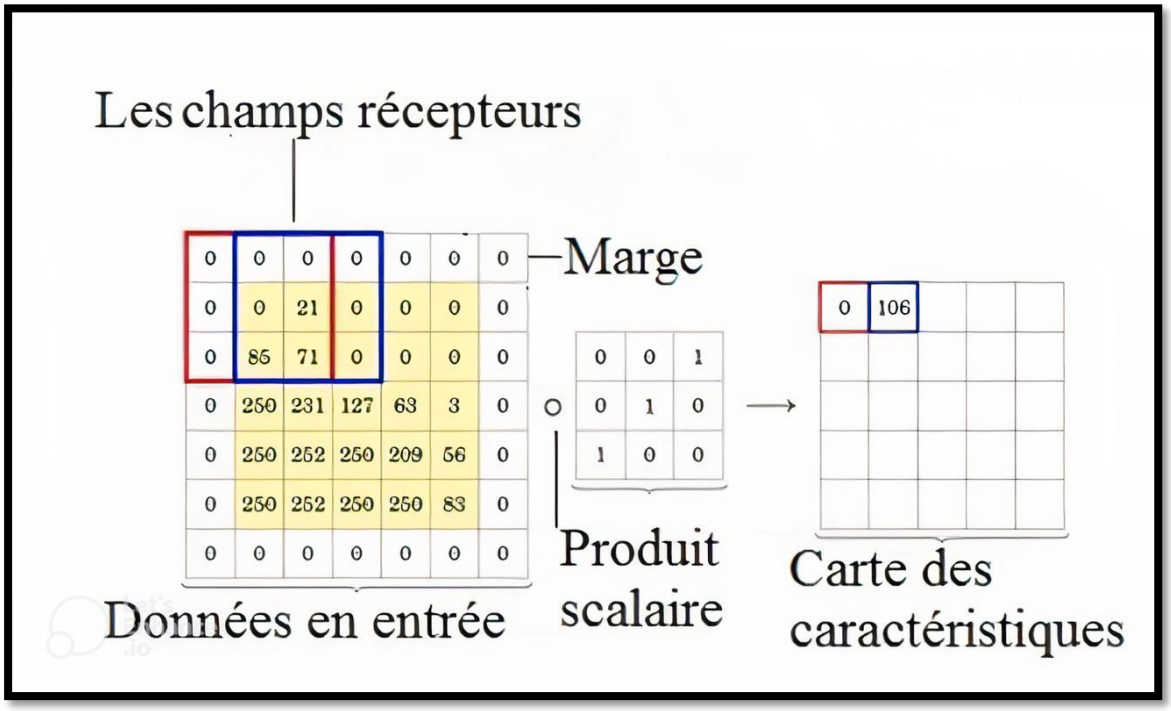


Figure2.6 Une opération de convolution[23].

Les valeurs nulles qui entourent la matrice d'entrée sont représentées par la marge. Cet espace empêche le filtre de sortir des limites de la matrice. Lorsque les filtres Nc sont appliqués aux données d'entrée, ils produisent une carte de caractéristiques avec une taille de $N^{(t+1)} \times N^{(t-1)} \times$Nc, où Nc est la profondeur-deur. Les cartes caractéristiques sont concaténées pour générer une couche convolutionnelle [23].

$$N^{(t+1)} = \frac{N^{(t)} - f + ep}{S} - 1$$

Le processus de convolution illustre le dimen- CNN, dans lequel chaque cellule (neurone) de la tique commune n'est liée qu'à un sous-ensemble de champs d'entrée de neurones.
En outre, en appliquant le même filtre à toutes les cartes de caractéristiques, il est en mesure d'identifier les caractéristiques qui ont été précédemment découverts dans diverses sections de l'image.La fonction d'activation ReLu est appliquée à la couche de convolution générée à la fin de chaque opération de convolution pour l'améliorer. Les États membres de l'Union européenne continueront à généraliser[23].

- ✓ **Couche de pooling**

Le travail de la couche de mise en commun est d'abaisser la dimensionnalité des couches de convolution qui apparaissent. Le but de cette réduction est d'augmenter la précision en se concentrant sur les caractéristiques les plus importantes. De plus, la réduction du nombre de paramètres réduit la taille du modèle et optimise la complexité temporelle. L'équation 2.1 avec
P = 0 calcule la taille de la matrice de sortie de l'opération de mutualisation. La mise en commun maximale et la mise en commun moyenne sont les deux formes de mise en commun. La valeur maximale du champ réceptif est retournée par la procédure de Max-pooling, alors que la

moyenne des valeurs est retournée par le processus de Avg-pooling. Le type d'architecture CNN le plus courant est la mise en commun maximale Figure 2.7 [23].

- ✓ **Couche entièrement connectée**

Les couches entièrement connectées (FC) dans un CNN ont la même structure que les couches MLP. Ces couches sont utilisées pour apprendre les combinaisons non linéaires entre les attributs récupérés par les couches convolutionnelles. Le résultat de la dernière couche de convolution [N, N, Nc] est aplati en un vecteur de taille [N N Nc].Ce vecteur relie tous les niveaux complètement liés à la couche d'entrée.La couche finale est utilisée pour la prédiction dans la classification supervisée, et elle est basée sur la fonction d'activation Softmax[23].

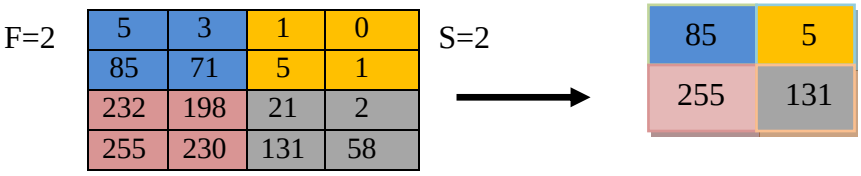


Figure 2.7 – Une opération de Max-pooling[25].

4. Conclusion

Dans ce chapitre, nous avons abordé le domaine de l'apprentissage profond dans son ensemble, ainsi que ses principes de base, ainsi que les différentes formes d'apprentissage profond et les architectures les plus couramment utilisées, dans ce chapitre.

Le chapitre sur la conception, qui comprend la modélisation de la réponse à notre sujet de fin d'étude, commencera dans le chapitre suivant.

1. Introduction :

Ce chapitre est consacré à la présentation de la solution proposée et de sa mise en œuvre; nous décrirons d'abord le but du projet, puis un aperçu général de la méthode de travail que nous avons suivie .Ensuite, avant de prendre une décision finale sur l'ensemble de données final, y compris les nombreuses étapes de son prétraitement, nous citerons et présenterons les différents ensembles de données sur lesquels nous avons appliqué nos modèles (partie image et texte).

Après la présentation de l'ensemble de données choisi, nous discuterons des différents modèles utilisés pour reconnaître les émotions dans notre ensemble d'images, ainsi que les différents modèles utilisés pour analyser les sentiments dans notre ensemble de texte, et nous conclurons par une discussion sur la technique de fusion utilisée pour améliorer les résultats obtenus.

2. Objectif du projet :

Notre projet est divisé en quatre parties, dont la première est la procédure de découverte des sentiments à partir d'une phrase et d'un texte spécifiques que vous avez préalablement identifiés. Et la seconde est d'analyser les émotions dans une image spécifique en utilisant une base de données pour les images et enfin d'analyser les sentiments dans une simple vidéo ou une vidéo liée à un texte spécifique.
Le but de notre projet est d'essayer d'améliorer la prédiction des émotions dans un texte, une image ou une vidéo..

3. Conception

3.1. L'architecture générale proposée :

Dans la fusion multimodal besoin de différent base de donnée, le premier est du texte et l'autre est spécifique à l'image, soit pour analyser les sentiments dans une vidéo que nous utilisons les deux. Parce que dans l'analyse des émotions vidéo est divisé en un ensemble d'imageset pour La vidéo attachée au dialogue est divisé en deux phases la première consiste à diviser la vidéo en images la seconde est d'entrer dans le dialogue et il devient un ensemble de phrases et se décompose en utilisant le base de donné du texte.

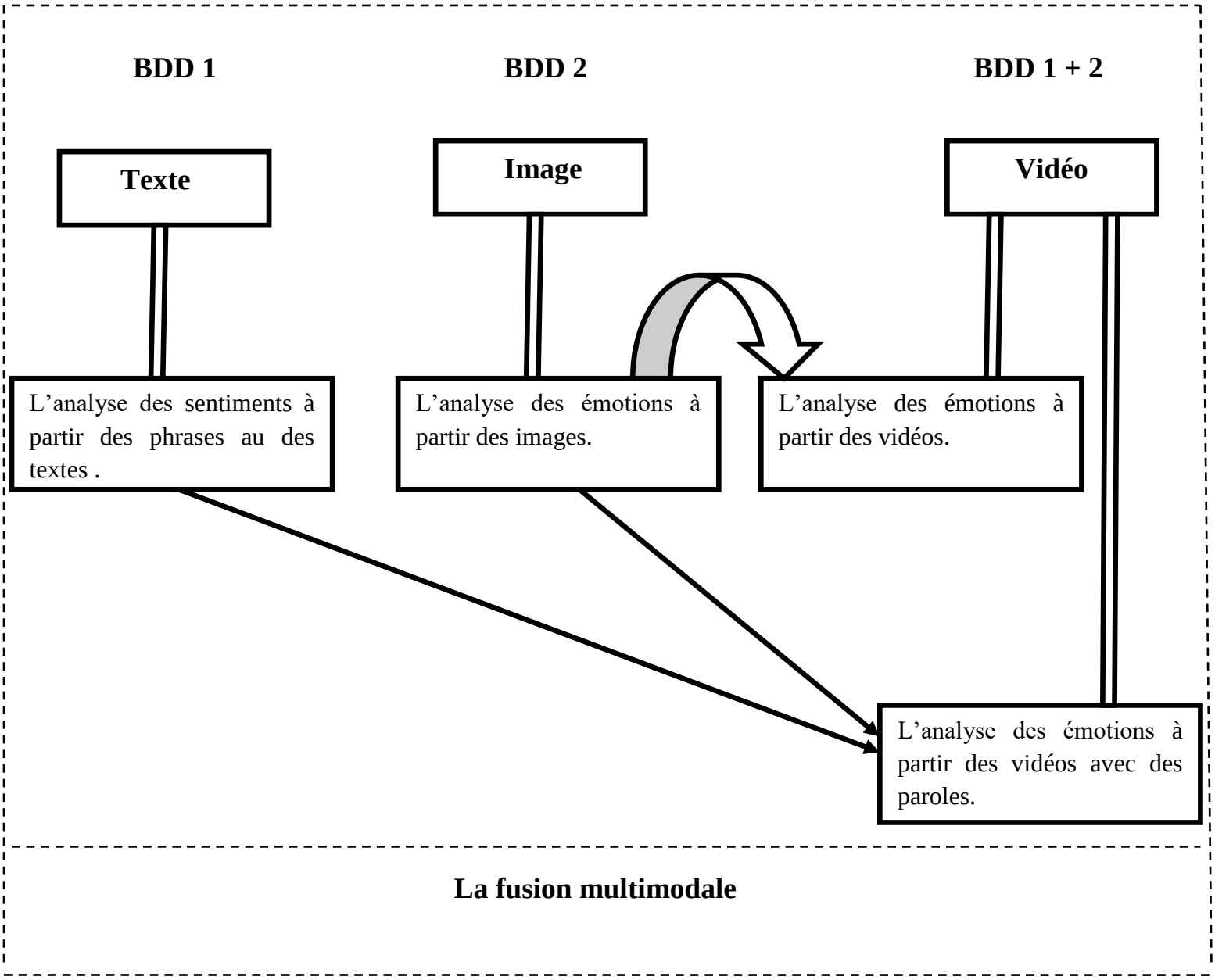


Figure.3.1 l'architecture de la fusion multimodale.

3.2. Acquisition des données

Nous allons vous montrer ici les ensembles de données que nous avons utilisés pour tester nos modèles avant de décider de notre ensemble de données final. Notre point de départ était un ensemble de données de modalité unique, comme la base de données « Emotions dataset for NLP » pour le texte et la base de données « FER 2013 » pour les images.

a. B ase de donnée du texte

Le jeu de données« Emotions dataset forNLP » comprend du TEXTE qui décrit de nombreux sentiments, y compris la joie, la colère, le dégoût, la peur, la tristesse, l'étonnement et la neutralité.

Source de la DataSet : https://www.kaggle.com/datasets/praveengovi/emotions-dataset
Licence : Domaine publique

- **Contexte**

Cette collection de documents et leurs sentiments est très utile pour les problèmes de classification de traitement de langage naturel

- **Contenu**

L'ensemble de données est divisé en apprentissage, test et validation pour créer le modèle d'apprentissage profond à partir d'une liste de documents **Example**

Tableau.3.2 exemple de classement.

Example de texte	**Classe**
I feel like I am still looking at a blank canvas blank pieces of paper; sadness	sadness

Figure.3.2 La structure générale de la base de données Emotions dataset for NLP.

- **Inconvénients de la dataset :**

- Base de données petite d'une taille de 738kb.
- Les textes n'expriment pas toujours les sentiments dans leurs sens apparent.

b. Base de données des images « fer2013 »

- **Description de la base de données :**

fer2013 est une base de données contenant un ensemble d'images représentants différentes expressions faciales à savoir : Heureux, fâché, dégouté, peur, triste,
surpris et neutre.
Source de la DataSet : https://www.kaggle.com/msambare/fer2013
Licence : Domaine publique

- **La représentation des données :**

Les images de visage en niveau de gris mesurant 48x48 pixels composent les données.
Les visages ont été automatiquement capturés de sorte qu'ils sont plus ou moins centrés dans chaque plan et occupent presque le même espace.
La tâche consiste à catégoriser chaque visage en fonction de l'émotion montrée dans l'expression du visage dans l'une des sept catégories (0=Fâché, 1=Dégoût, 2=Peur, 3=Heureux, 4=Triste, 5=Surprise, 6=Neutre).
L'ensemble d'apprentissage se compose de 28 709 exemples et l'ensemble de test se compose de 3 589 exemples .

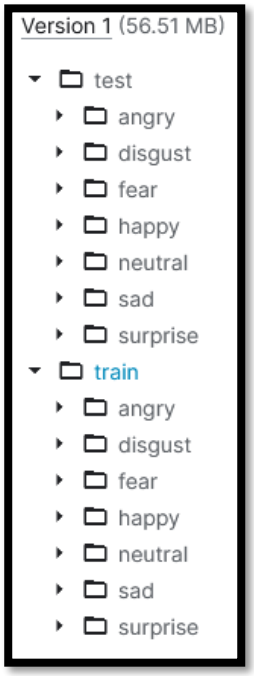


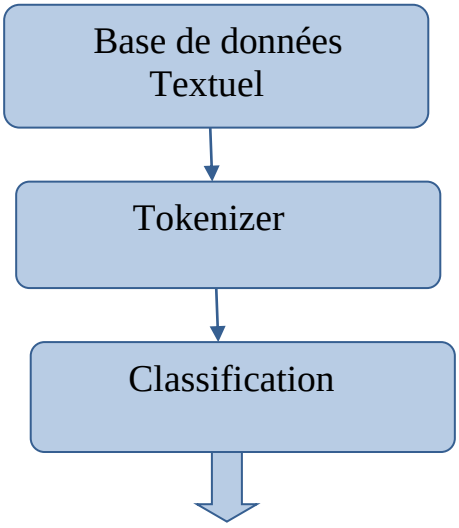
Figure.3.3 La structure générale de la base de données FER2013.

- **Inconvénients de la dataset :**

- Base de données petite (53,89Mb)
- Données d'une seule modalité, ce qui signifie que la base de données contient uniquement des images d'expressions faciales.

3.3. Partie 1 : l'analyse des sentiments des textes

Le schéma ci-dessous représente le processus adapté lors de l'analyse des sentiments :



- **Les bibliothèques utilisées :**

Le tableau ci-dessous englobe les différentes bibliothèques et packages utilisés lors de l'implémentation de nos différents modèles d'analyse des sentiments ainsi que leurs utilités, tout

en utilisant le langage python.

Tableau.3.2 Les différentes bibliothèques utilisées

Bibliothèque	**Les principales utilités**
Matplotlib et ScitPlot	Visualisation et de création de statistiques.
Keras	Prétraitement des images et Création des différentes couches des modèles.
Numpy	Utilisation des différentes fonctions mathématiques.
Pandas	Utilisation de DataFrame ou matrice de données.

- Le premier modèle LSTM :

```
Model: "sequential"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 embedding (Embedding)       (None, 300, 150)          1650000

 dropout (Dropout)           (None, 300, 150)          0

 lstm (LSTM)                 (None, 128)               142848

 dropout_1 (Dropout)         (None, 128)               0

 dense (Dense)               (None, 64)                8256

 dropout_2 (Dropout)         (None, 64)                0

 dense_1 (Dense)             (None, 6)                 390

=================================================================
Total params: 1,801,494
Trainable params: 1,801,494
Non-trainable params: 0
```

Figure.3.4 L'architecture du premier modèle LSTM pour l'analyse des sentiments.

Tableau.3.3 Les paramètres du premier modèle de l'analyse des sentiments.

L'optimisateur	Adam
Le nombre d'epochs	10
La taille du batch	64

Suite à l'exécution de ce modèle nous avons constaté le tableau suivant :

Tableau3.4 Les taux de précision et perte pour modèle de classification des sentiments.

Taux	**Données d'entraînement**	**Données de validation**
Précision (Accuracy)	0.97	0.42
Perte (Loss)	0.06	1.48

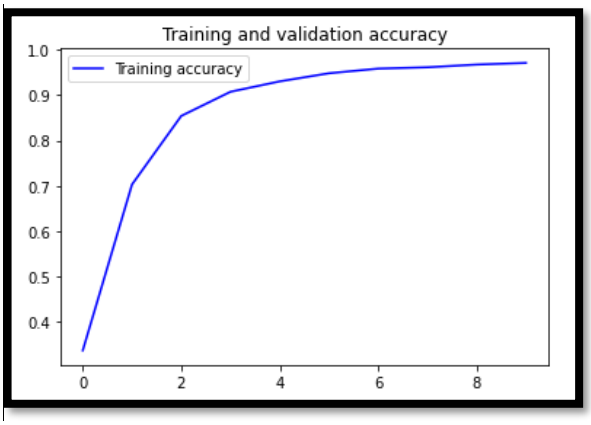
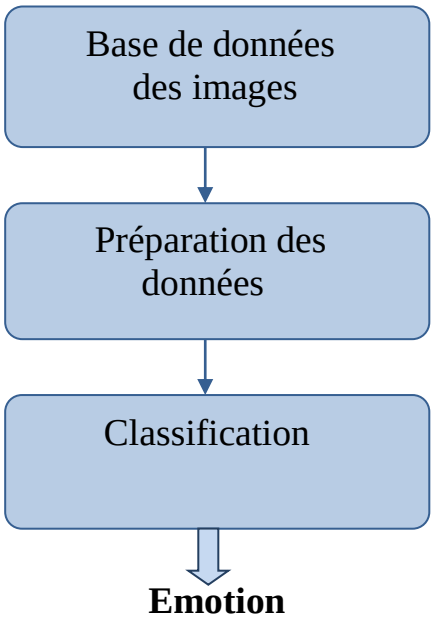


Figure.3.5 Le taux de perte

Figure.3.6 Le taux de Précision

3.4. Partie 2 : l'analyse des émotions dans les images



- **Les bibliothèques utilisées :**

Le tableau ci-dessous englobe les différentes bibliothèques et packages utilisés lors de L'implémentation de nos différents modèles de reconnaissance d'émotions ainsi que leurs Utilités, tout en utilisant le langage python.

Tableau3.5. les différentes bibliothèques utilisées

Bibliothèque	**Les principales utilités**
Matplotlib et ScitPlot	Visualisation et créations de statistiques.
Keras	Prétraitement des images et création des différentes couches des modèles.
Numpy	Utilisation des différentes fonctions mathématiques.
Pandas	Utilisation de DataFrame ou matrice de données.
Opencv	Traitement des images et détection d'élément précis.

- Le premier modèle CNN :

Le premier modèle est un modèle basé sur l'architecture CNN qui comprends l'implémentation des différentes couches de cette architecture.
La figure ci-dessous décrit les différentes couches de notre premier modèle.

```
dropout_33 (Dropout)         (None, 24, 24, 64)        0

conv2d_47 (Conv2D)           (None, 24, 24, 128)       73856

conv2d_48 (Conv2D)           (None, 22, 22, 256)       295168

batch_normalization_23 (Bat  (None, 22, 22, 256)       1024
chNormalization)

max_pooling2d_23 (MaxPoolin  (None, 11, 11, 256)       0
g2D)

dropout_34 (Dropout)         (None, 11, 11, 256)       0

flatten_11 (Flatten)         (None, 30976)             0

dense_22 (Dense)             (None, 1024)              31720448

dropout_35 (Dropout)         (None, 1024)              0

dense_23 (Dense)             (None, 7)                 7175

=================================================================
Total params: 32,116,743
Trainable params: 32,116,103
Non-trainable params: 640
```

Figure.3.6 L'architecture du deuxième modèle CNN pour la reconnaissance des émotions.

Tableau.3.6 Les paramètres du deuxième modèle de l'analyse d'émotions.

L'optimisateur	Adam
Le nombre d'epochs	10
La taille du batch	64

Suite à l'exécution de ce modèle nous avons constaté le tableau suivant :

Tableau.3.7 Les taux de précision et perte pour le modèle de reconnaissance d'émotions.

Taux	**Données d'entraînement**	**Données de validation**
Précision (Accuracy)	0.99	0.61
Perte (Loss)	0.07	1.84

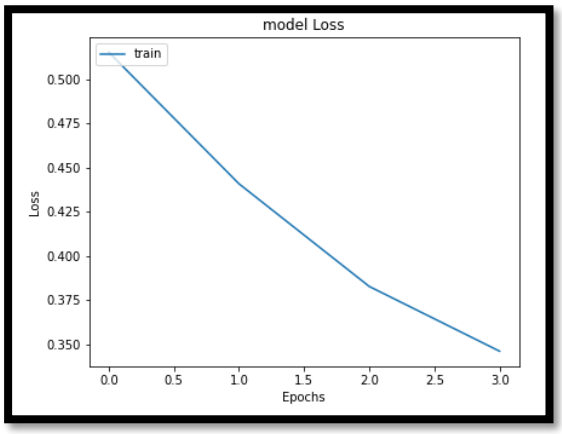


Figure.3.8 Le taux de perte.

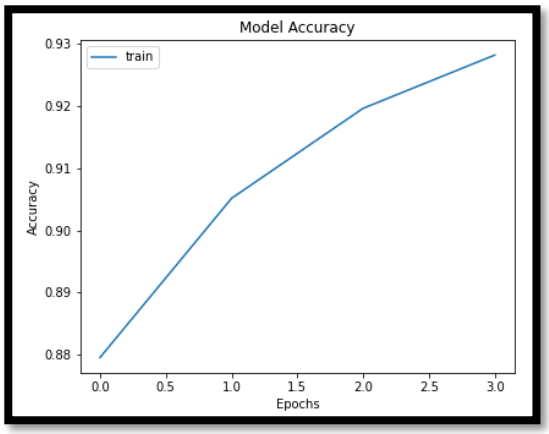


figure.3.9 Le taux de précision .

4. Implémentation et résultats

4.1. Technologie utilisé :

- Nous avons utilisé un ordinateur dell de système exploitation 64 bites , la RAM 2Go .
- Pour le stockage des données, nous avons utilisé et exploité l'un des services gratuits qu'offre Google appelé Google drive .

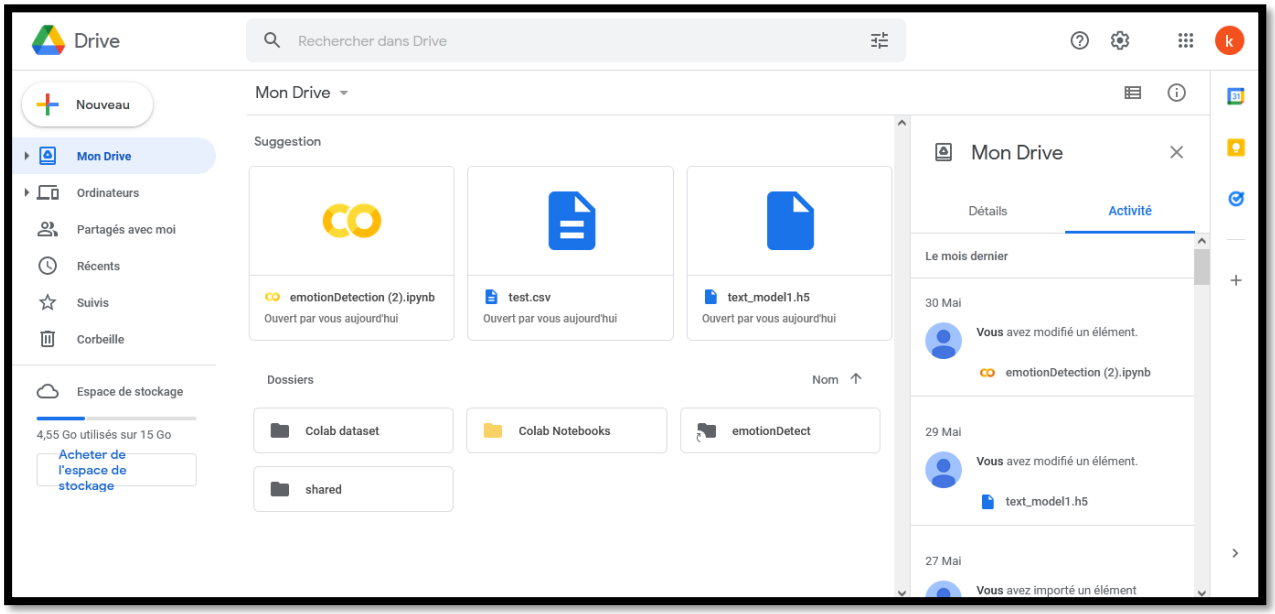


Figure.3.10. L'interface des données stockées dans Google drive .

- Afin de développer nos nombreux modèles, nous avons utilisé et géré l'un des services gratuits de Google appelé Colaboratory, qui est un environnement dédié à l'apprentissage automatique et profond qui vous permet de créer et d'exécuter n'importe quel code Python.
  Ceci est dû à ses nombreux avantages, y compris :
  - Aucune configuration n'est nécessaire.
  - L'accès au GPU est illimité.
  - Facile à partager.
- Langage utilisée Python : est un langage de programmation de haut niveau créé par 'Guido Van Rossumet' un grand nombre de contributeurs volontaires depuis 1989.Python est un langage de programmation portable, dynamique, extensible et gratuit qui offre une approche modulaire et orientée objet. Il dispose d'une bibliothèque standard complète.
- Concernant l'exécution des modèles nous aurons besoin de la connexion internet.

4.2. Expérimentations

➢ Télécharger le code cloud

```
[ ]  from google.colab import drive
     drive.mount('/content/drive')
```

Figure.3.11 téléchargement du code cloud .

➢ Dans cette partie nous chargeons la base de données du texte

```
#%reset
import tensorflow as tf
import numpy as np
import pandas as pd
dataset_train = pd.read_csv('/content/drive/MyDrive/Colab dataset/emotionDetect/training.csv')
dataset_test = pd.read_csv('/content/drive/MyDrive/Colab dataset/emotionDetect/test.csv')
dataset_train['length'] = [len(x) for x in dataset_train['text']]
labels = { 0: "sadness"  , 1: "anger"  , 2: 'love', 3: 'surprize' , 4: 'fear' , 5: 'joy'}
dataset_train["label"]=dataset_train["label"].replace(labels)
dataset_test["label"]=dataset_test["label"].replace(labels)
dataset_train.head()
display(dataset_train)
text = ' '.join(dataset_train["text"])
text
```

Figure.3.12 code chargement de la base de données du texte.

- Le code suivant permet d'importer les bibliothèques utilisées .

```
import matplotlib.pyplot as plt
import seaborn as sns
import numpy as np
import pandas as pd

from wordcloud import WordCloud
import torch
from sklearn.model_selection import train_test_split
from sklearn.metrics import confusion_matrix,classification_report,accuracy_score
from sklearn.preprocessing import LabelEncoder

import tensorflow as tf
from tensorflow.keras.utils import to_categorical
from tensorflow.keras.preprocessing.text import Tokenizer
from tensorflow.keras.preprocessing.sequence import pad_sequences

from tensorflow.keras.optimizers import Adam
from tensorflow.keras.models import Sequential
from tensorflow.keras.callbacks import EarlyStopping
from tensorflow.keras.layers import Dense, LSTM, Embedding, Bidirectional,Dropout, Conv2D, MaxPool2D

import re
import nltk
from nltk.corpus import stopwords
from nltk.stem import PorterStemmer
from transformers import BertTokenizer, BertConfig,AdamW, BertForSequenceClassification,get_linear_schedule_with_warmup
```

Figure.3.13 importation des bibliothèques.

.

- Dans cette partie nous préparons le base de donnée du texte .

```
lb = LabelEncoder()

dataset_train['label'] = lb.fit_transform(dataset_train['label'])
dataset_test['label'] = lb.fit_transform(dataset_test['label'])
lb.classes_ = np.load('/content/drive/MyDrive/Colab dataset/emotionDetect/lb_classes.npy',allow_pickle=True)
display(dataset_train.head(3))
nltk.download('stopwords')
stopwords = set(nltk.corpus.stopwords.words('english'))
max_len=dataset_train['length'].max()
tokenizer = Tokenizer()
tokenizer.fit_on_texts(" ".join(dataset_train['text']))
```

Figure.3.14 Code de préparation des bases de donnée du texte .

- cette partie nettoie le texte virgules, des points et des emojis.

```
from tensorflow.keras.preprocessing.text import one_hot
vocabSize = 11000
def text_cleaning(df, column):
    stemmer = PorterStemmer()
    corpus = []

    for text in df[column]:
        text = re.sub("[^a-zA-Z]", " ", text)
        text = text.lower()
        text = text.split()
        text = [stemmer.stem(word) for word in text if word not in stopwords]
        text = " ".join(text)
        corpus.append(text)
    one_hot_word = [one_hot(input_text=word, n=vocabSize) for word in corpus]
    pad = pad_sequences(sequences=one_hot_word,maxlen=max_len,padding='pre')
    print(pad.shape)
    return pad
x_train_txt = text_cleaning(dataset_train, "text")
x_test_txt = text_cleaning(dataset_test, "text")
y_train_txt = dataset_train["label"]
y_test_txt = dataset_test["label"]
y_train_txt = to_categorical(y_train_txt)
y_test_txt = to_categorical(y_test_txt)
```

```
(16000, 300)
(2000, 300)
```

Figure.3.15 nettoyage du texte.

- Cette partie pour l'entrainement du modèle .

```
model = Sequential()
model.add(Embedding(input_dim=vocabSize,output_dim=150,input_length=300))
model.add(Dropout(0.2))
model.add(LSTM(128))
model.add(Dropout(0.2))
model.add(Dense(64,activation='sigmoid'))
model.add(Dropout(0.2))
model.add(Dense(6,activation='softmax'))
model.compile(optimizer='adam',loss='categorical_crossentropy',metrics=['accuracy'])
model.summary()
callback = EarlyStopping(monitor="val_loss", patience=2, restore_best_weights=True)
hist = model.fit(x_train_txt,y_train_txt,epochs=10,batch_size=64, verbose=1, callbacks=[callback])
model.save('/content/drive/MyDrive/Colab dataset/emotionDetect/text_model1.h5')
```

Figure.3.16 l'apprentissage du modèle.

- Cette étape est sauvegarde du modèle.

```
from tensorflow import keras
model = keras.models.load_model('/content/drive/MyDrive/Colab dataset/emotionDetect/text_model1.h5')
callback = EarlyStopping(monitor="val_loss", patience=2, restore_best_weights=True)
hist = model.fit(x_train_txt,y_train_txt,epochs=20,batch_size=64, verbose=1, callbacks=[callback])
model.save('/content/drive/MyDrive/Colab dataset/emotionDetect/text_model1.h5')
```

Figure.3.17 sauvegarde du modèle

- Ce code représente la partie de classification du CNN pour la reconnaissance des émotions dans une image.

```
#!unzip "/content/drive/MyDrive/Colab dataset/emotionDetect/imageDataset/fer.zip" -d "/content/drive/MyDrive/Colab dataset/emotionDetect/imageDataset/"
from keras.preprocessing.image import ImageDataGenerator
from keras.layers import Conv2D,Dense, MaxPooling2D,Flatten,Dropout,Activation, BatchNormalization
from tensorflow.keras.optimizers import Adam,RMSprop,SGD
from keras import regularizers
from keras.callbacks import ModelCheckpoint, CSVLogger,TensorBoard,EarlyStopping,ReduceLROnPlateau
from sklearn.model_selection import train_test_split
import datetime
from tensorflow.keras.utils import plot_model

def get_model(input_size, classes=7):
    #BUILDING the CNN
    model = tf.keras.models.Sequential()

    model.add(Conv2D(32, kernel_size=(3, 3), padding='same', activation='relu', input_shape =input_size))
    model.add(Conv2D(64, kernel_size=(3, 3), activation='relu', padding='same'))
    model.add(BatchNormalization())
    model.add(MaxPooling2D(2, 2))
    model.add(Dropout(0.25))

    model.add(Conv2D(128, kernel_size=(3, 3), activation='relu', padding='same', kernel_regularizer=regularizers.l2(0.01)))
    model.add(Conv2D(256, kernel_size=(3, 3), activation='relu', kernel_regularizer=regularizers.l2(0.01)))
    model.add(BatchNormalization())
    model.add(MaxPooling2D(pool_size=(2, 2)))
    model.add(Dropout(0.25))
```

Figure.3.18. l'architecture du CNN pour la reconnaissance des émotions dans une image.

- Dans cette partie nous chargeons la base de données d'images.

```
#!unzip "/content/drive/MyDrive/Colab dataset/emotionDetect/imageDataset/fer.zip" -d "/content/drive/MyDrive/Colab dataset/emotionDetect/imageDataset/"
from keras.preprocessing.image import ImageDataGenerator
from keras.layers import Conv2D,Dense, MaxPooling2D,Flatten,Dropout,Activation, BatchNormalization
from tensorflow.keras.optimizers import Adam,RMSprop,SGD
from keras import regularizers
from keras.callbacks import ModelCheckpoint, CSVLogger,TensorBoard,EarlyStopping,ReduceLROnPlateau
from sklearn.model_selection import train_test_split
import datetime
from tensorflow.keras.utils import plot_model

def get_model(input_size, classes=7):
    #BUILDING the CNN
    model = tf.keras.models.Sequential()

    model.add(Conv2D(32, kernel_size=(3, 3), padding='same', activation='relu', input_shape =input_size))
    model.add(Conv2D(64, kernel_size=(3, 3), activation='relu', padding='same'))
    model.add(BatchNormalization())
    model.add(MaxPooling2D(2, 2))
    model.add(Dropout(0.25))

    model.add(Conv2D(128, kernel_size=(3, 3), activation='relu', padding='same', kernel_regularizer=regularizers.l2(0.01)))
    model.add(Conv2D(256, kernel_size=(3, 3), activation='relu', kernel_regularizer=regularizers.l2(0.01)))
    model.add(BatchNormalization())
    model.add(MaxPooling2D(pool_size=(2, 2)))
    model.add(Dropout(0.25))
```

Figure.3.19 chargement la base de données d'images.

- Cette partie est concernent l'entrainement et reconversion du modèle.

```
from tensorflow import keras
model = get_model((48,48,1),classes)
#model = keras.models.load_model('/content/drive/MyDrive/Colab dataset/emotionDetect/my_model.h5')
model.summary()
callback = EarlyStopping(monitor="val_loss", patience=2, restore_best_weights=True)
checkpoint_filepath = '/content/drive/MyDrive/Colab dataset/emotionDetect/tmp/checkpoint'
model_checkpoint_callback = tf.keras.callbacks.ModelCheckpoint(
    filepath=checkpoint_filepath,
    save_weights_only=True,
    monitor='val_accuracy',
    mode='max',
    save_best_only=True)
hist = model.fit(x_train, y_train, epochs=3, callbacks=[callback,model_checkpoint_callback])
model.save('/content/drive/MyDrive/Colab dataset/emotionDetect/my_model1.h5')
```

Figure.3.20. Entrainement et reconversion du modèle.

- Cette partie du code est pour prédire l'émotion dans la vidéo.

```
def predict_video(path,start=0,end=0,skip_frame=0):
  cap = cv2.VideoCapture(path)
  frame_index = 0
  fps = cap.get(cv2.CAP_PROP_FPS)
  print(fps)
  start1 = start*round(fps)
  end1 = end*round(fps)
  print(start1)
  print(end1)
  if (cap.isOpened()== False):
    print("Error opening video stream or file")
  predictions = np.array([[.0,.0,.0,.0,.0,.0,.0]],dtype='float32')
  while(cap.isOpened()):
    ret, frame = cap.read()
    frame_index += 1
    if start1>frame_index or (skip_frame!=0 and frame_index%skip_frame!=0) :
      continue
    if end1<frame_index and end1!=0 :
      break
    if ret == True:
      frame_prediction = predict_sigle_frame(frame,showFrame=True)
      if frame_prediction.size > 0 and np.max(frame_prediction,axis=0)>0.55:
        predictions = np.insert(predictions,0,frame_prediction,axis=0)
```

Figure.3.21 Prédiction de l'émotion dans la vidéo..

- Cette partie du code détermine l'émotion la plus dominante dans la vidéo .

```
pre_sum = np.sum(predictions,axis=0)/np.sum(predictions)
msg = '\n'.join([class_mapping[i] + "{:10.2f}%".format(pre_sum[i]*100) for i in range(0,7)])
index = np.argmax(pre_sum, axis=0)
print('Video prediction')
print(msg)
print('final result : '+class_mapping[index]+' video')
```

Figure.3.22 Détermination de l'émotion dans la vidéo.

- Dans Cette partie nous donnons le plus grand pourcentage dans le choix des émotions pour chacun des textes ou de l'image .

```
text_importence = 0.3
pre_mean = pre_sum*(1-text_importence) + txt_prediction*text_importence
msg = '\n'.join([class_mapping[i] + "{:10.2f}%".format(pre_mean[i]*100) for i in range(0,7)])
index = np.argmax(pre_mean, axis=0)
print('Report : Text and Video prediction')
print(msg)
print('final result : '+class_mapping[index]+' video')
```

Figure.3.23 Le choix de la classe.

4.3. Un cas d'utilisation de notre système

a. Texte

La première chose que nous faisons est de sélectionner le texte que nous voulons traiter , nous avons un endroit dédié à l'écriture de la phrase ou du texte sur lequel vous souhaitez travailler et un endroit spécial pour montrer le résultat.

```
sentences = [
            "don't ever let somebody tell you that you can not do something",
            "This is outrageous, how can you talk like that?",
            "I feel like im all alone in this world",
            "He is really sweet and caring",
            ]
```

Figure.3.24 Exemple du texte exécuté .

```
don't ever let somebody tell you that you can not do something
anger : 0.9597151279449463


This is outrageous, how can you talk like that?
surprize : 0.9245413541793823


I feel like im all alone in this world
sadness : 0.9972366094589233


He is really sweet and caring
anger : 0.8765789270401001
```

Figure.3.25 Résultat de classification d'émotion.

b. Image :

La première chose que nous faisons est de sélectionner l'image que nous voulons traiter , et nous avons un endroit pour montrer le résultat de l'mage sélectionner .

```
image = cv2.imread('/content/drive/MyDrive/Colab dataset/emotionDetect/sadBoy.jpg')
predict_sigle_frame(image)
```

Figure.3.26 l'image sélectionnée.

- La figure dessous présente le résultat d'émotion dans l'image sélectionnée, il a choisi une photo triste parce qu'il est triste a 64,12%, ce qui est le plus grand.

Figure3.27 Résultat de classification de l'émotion.

c. vidéo :

La première chose que nous faisons est de sélectionner la vidéo que nous voulons traiter, et nous avons un endroit pour montrer le résultat de la vidéo sélectionnée.

```
predictions = predict_video(path='/content/drive/MyDrive/Colab dataset/emotionDetect/vid3.mp4'
```

Figure.3.28 Sélectionner une vidéo.

Deuxième chose est de splitter la vidéo en images et traiter chaque image à la fois et enfin choisir l'émotion qui a le plus grand pourcentage entre toutes les images.

Figure.3.29 Résultat de classification des émotions des deux premières parties de la vidéo.

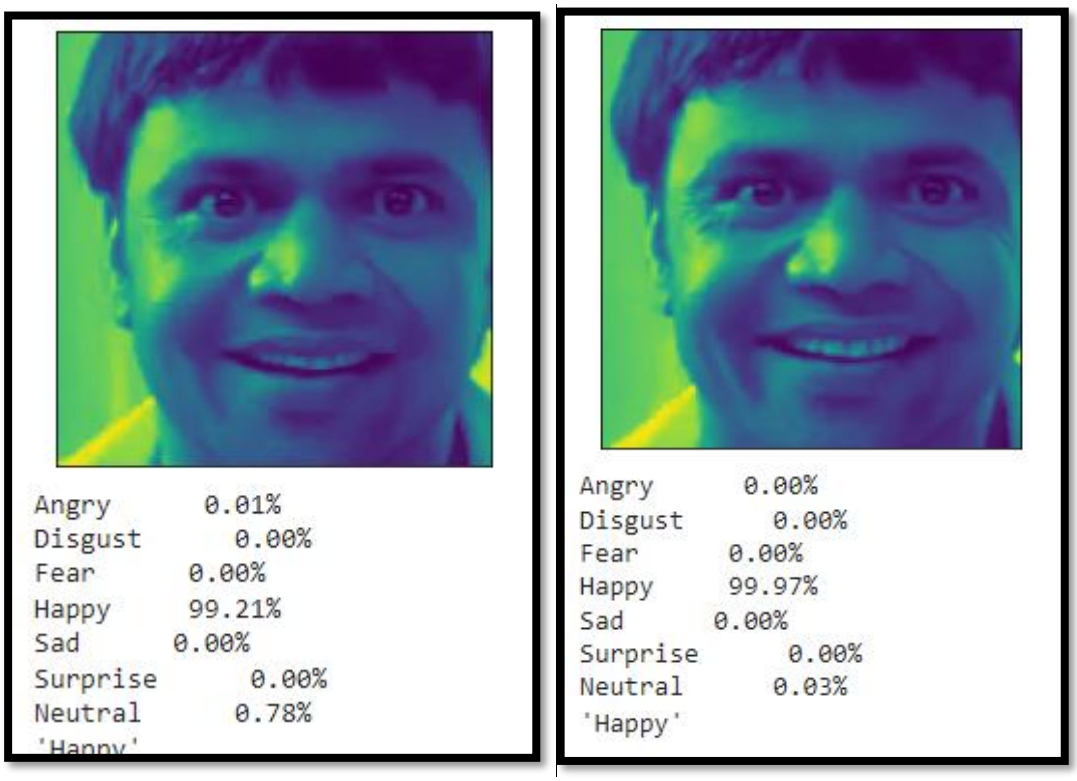


Figure.3.30 Résultat de classification des émotions de la troisième et quatrième partie de la vidéo.

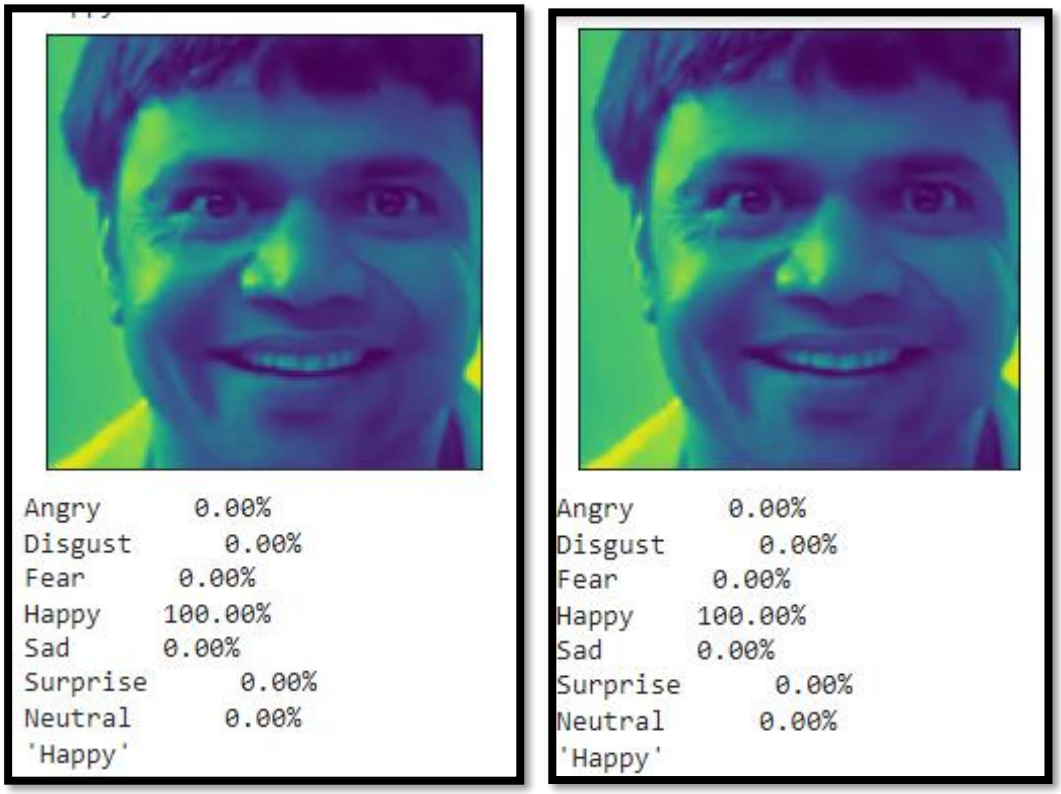


Figure.3.31 Résultat de classification des émotions des deux dernières parties de la vidéo.

d. Vidéo avec texte :

La première chose que nous faisons est de sélectionner le texte de la vidéo et la vidéo que nous voulons traiter, et nous avons un endroit pour montrer le résultat de la vidéo sélectionné.

```
[ ]  sentence = get_sub('/content/drive/MyDrive/Colab dataset/emotionDetect/vid/scene.csv',start='1:31')
     print(sentence)

     dont ever let somebody tell you that you can not do something not even me
```

Figure.3.32. Sélectionner le texte de la vidéo

```
predictions = predict_video(path='/content/drive/MyDrive/Colab dataset/emotionDetect/vid3.mp4'
```

Figure.3.33. Sélectionner la vidéo.

La deuxième chose est de diviser la vidéo en des images et traiter image par image. La troisième étape est de sélectionner et traiter le texte de la vidéo. En fin, nous choisissons l'image ou le texte de cette vidéo qui dépend davantage du processus de l'analyse des sentiments.

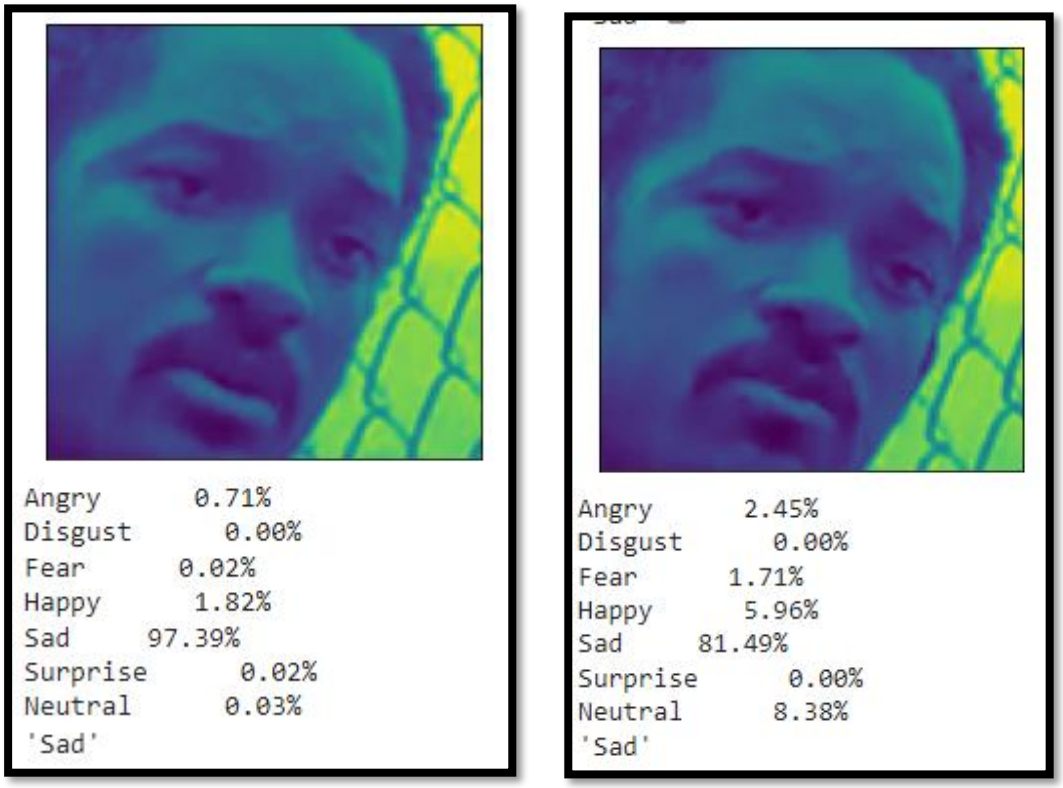


Figure.3.34 résultat de classification des émotions des deux dernières parties de la vidéo.

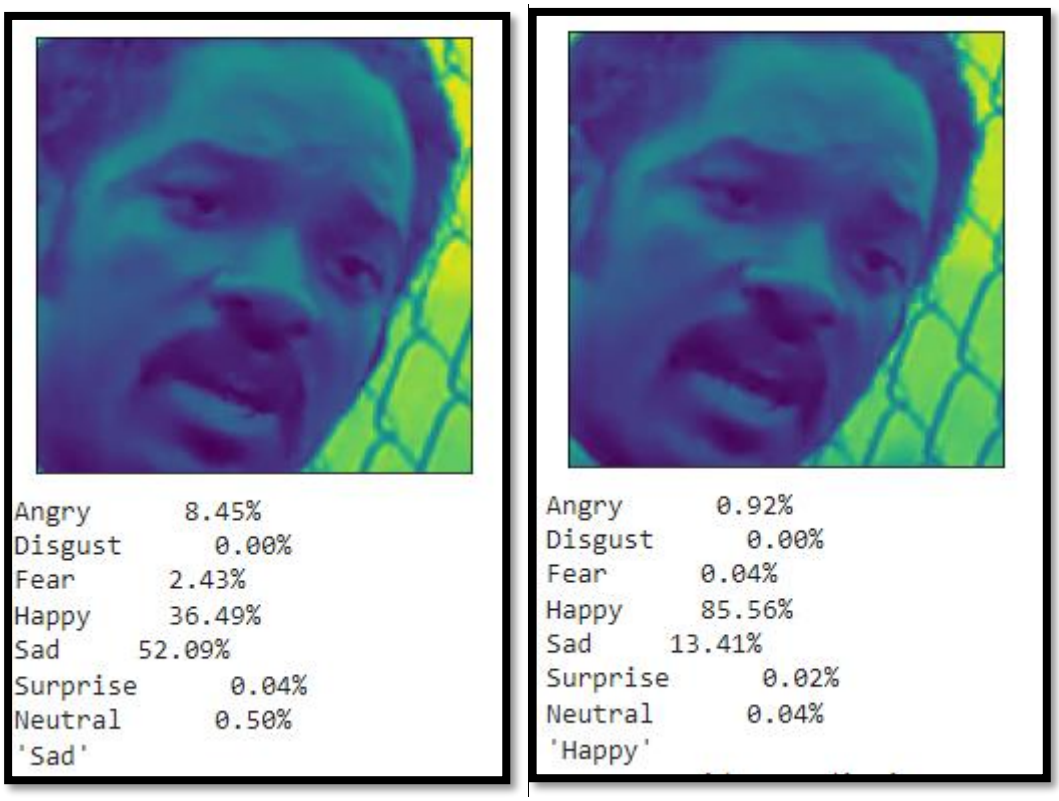


Figure3.35. résulta de classement d'émotion les deux dernier partie du vidéo.

- La figure dessous présente la prédiction résultat des image du vidéo

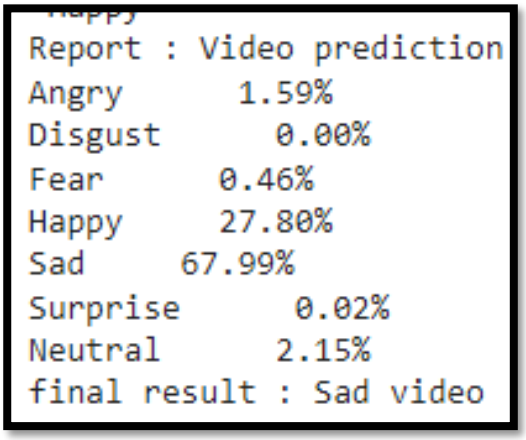


Figure.3.36 Résultat de la prédiction de la vidéo.

- La figure dessous présente le résultat des émotions du texte associé à la vidéo.

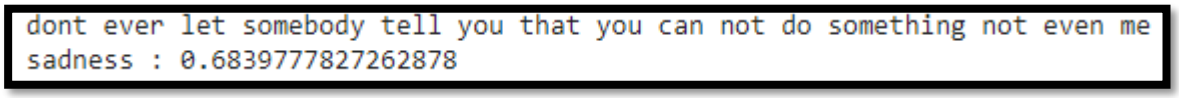


Figure.3.37 Résultat des émotions du texte de la vidéo.

- La figure dessous présente le résultat final de la vidéo.

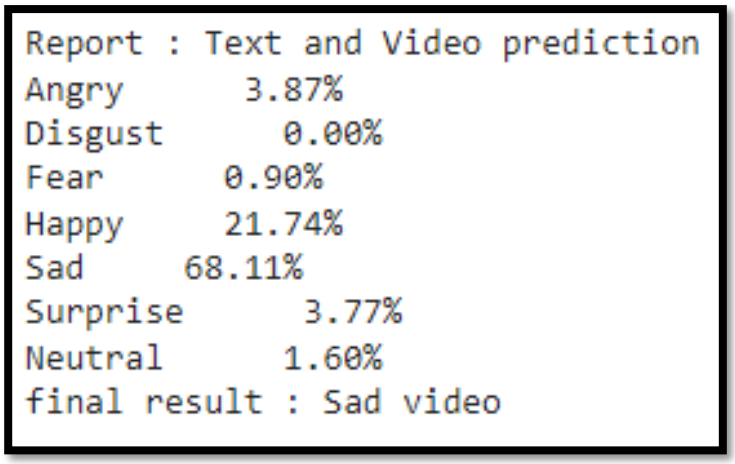


Figure.3.38 Résultat final de classification de la vidéo.

4. Conclusion

Dans ce chapitre, nous avons regroupé les parties conception et implémentation ,commenceront par l'architecture générale proposée pour notre projet. Ensuite, nous avons déterminé les bases de données utilisées et détaillé les parties d'analyse des émotions (texte, image). Dans la partie implémentation nous avons expliqué les parties importantes du code source et ensuit enous avons présenté des captures d'écran décrivant quelques scénarios d'exécution de notre travail.

La mise en œuvre des modèles d'analyse des sentiments dans les textes et la reconnaissance des émotions dans les images et les vidéos, sont les parties majeures de ce projet de fin d'étude.
Nous avons travaillé à chaque étape sur une modalité particulière et dans la phase vidéo nous l'avons divisé en images et nous avons travaillé sur chaque image seule et dans la dernière nous avons sélectionné les sentiments les plus fréquents et les plus relatifs. Quant à la vidéo avec le dialogue, nous avons écrit le dialogue contenu dans la vidéo et analysé les émotions dans le dialogue sous forme d'une analyse de texte, et dans le dernier nous avons choisi laquelle des proportions sur lesquelles nous nous concentrons le texte ou les images d'une vidéo.
Durant notre travail, nous avons rencontré certaines difficultés lors des différentes étapes citées précédemment, et parmi ces difficultés :
- Les images de l'ensemble de données ne conviennent pas à la reconnaissance d'émotions faciales car certaines images ne contiennent pas de visages.
- Le temps passé dans la phase d'apprentissage sur les images utilisées pour identifier les émotions.

[1] N. M. Hakak, M. Mohd, M. Kirmani And M. Mohd, "Emotion Analysis : A Survey," India, 2017.
[2] S. Kaur And N. Kulkarni, "Emotion Recognition - A Review," India, 2021.
[3] Khadoudja Ghanem. "Reconnaissance Des Expressions Faciales A Base D'informations Video Estimation De L'intensité Des Expressions Faciales". 2010.
[4] E. Rabot-Creusvaux. [Online]. Available: Http://Www.Psycho-Emdrsophro. Com/Pages/Les-Emotions.Html. [Accessed Mai 2021].
[5] M. Soleymani, D. Garcia, B. Jou, B. Schuller, S.-F. Chang And M. Pantic, "Image And Vision Computing - A Survey Of Multimodal Sentiment Analysis," 2017.
[6] Thibaut Thonet "Modèles Thématiques Pour La Découverte Non Supervisée De Points De Vue Sur Le Web" L'université Toulouse 3 Paul Sabatier, 2017.
[7] Q. Yao, "Multi-Sensory Emotion Recognition With Speech And Facial Expressions," 2014.
[8] Franck Davoine, Bouchra Abboud, And Mo Dang "Analyse De Visages Et D'expressions Faciales" Par Modèle Actif D'apparence. 2004.
[9] Catherine Soladie " Représentation Invariante Des Expressions Faciales. : Application En Analyse Multimodale Des Emotions. Phd Thesis" Supélec, 2013.
[10] Hugo Mercier "Outils Informatiques D'analyse Des Expressions Faciales En Langue Des Signes" Paul Sabatier Toulouze Iii, 2007.
[11] Mina Navraan, Nasrollah Moghadam Charkari, And Muharram Mansoorizadeh . " Automati Facial Emotion Recognition Method Based On Eye Region Changes. Information Systems & Telecommunication"Page 221, 2016.
[12] Nouioua Et Naouel .Mezzah Et Samia Et Mr Madi Moussa "Reconnaissance Automatique Des Expressions Faciales " Mémoire De Fin D'études En Vue De L'obtention Du Diplôme De Master Université Abderrahmane Mira De Bejaia.
[13] T. Luo And G. Xu "Sentiment Analysis" New York, 2013.
[14] M. Soleymani, D. Garcia, B. Jou, B. Schuller, S.-F. Chang And M. Pantic, "Image And Vision Computing - A Survey Of Multimodal Sentiment Analysis," 2017.
[15] F. Rahdari, E. Rashedi And M. Eftekhari, "A Multimodal Emotion Recognition System Using Facial Landmark Analysis," 2018.
[16] N. Shoumy, L.-M. Ang, K. P. Seng, D. Rahaman And T. Zia, "Multimodal Big Data Affective Analytics: A Comprehensive Survey Using Text, Audio, Visual And Physiological Signals," 2019.
[17] Jie Geng, Zhenjiang Miao, Member, Ieee, And Xiao-Ping Zhang, Senior Member, Ieee " Efficient Heuristic Methods For Multimodal Fusion And Concept Fusion In Video Concept Detection".
[18] Kamel Mohamed, "Reconnaissance De Formes Appliquée A L'écriture Arabe Manuscrite Par Des Multiclassifieurs", Thèse De Magister En Informatique, Université Mohamed Khider, Biskra, 2010.
[19] Andrei Doncescu, "Les Réseaux De Neurones Artificiels", Support De Cours.
[20] Boughaba Mohammed Et Boukhris Brahim "L`Apprentissage Profond (Deep Learning) Pour La Classification Et La Recherche D'images Par Le Contenu" Mémoire Master Professionnel Universite Kasdi Merbah Ouargla.
[21] L. Alzubaidi, J. Zhang, A. J. Humaidi, A. Al-Dujaili, Y. Duan, O. Al-Shamma, J.

Santamar.A, M. A. Fadhel, M. Al-Amidie And L. Farhan, "Review Of Deep Learning: Concepts, Cnn Architectures, Challenges, Applications, Future Directions," 2021.
[22] Labiad Ali "Sélection Des Mots Clés Basée Sur La Classification Et L'extraction Des Règles D'association" Mémoire Présenté A L'université Du Québec A Trois-Rivières Comme Exigence Partielle De La Maîtrise En Mathématiques Et Informatique Appliquées.
[23] Dif Nassima 'L'apprentissage Profond Pour Le Traitement D'images' Thèse De Doctorat Université Djillali Liabès De Sidi Bel Abbès 2020.

Références Web (Technique)

[W1] Https://Www.Cdrin.Com/Blog/Assistant-Virtuel-Emotions-Base-Animation-3d.
[W2] Https://Www.Math.Univ-Toulouse.Fr/~Besse/Wikistat/Pdf/St-M-App-Rn.Pdf.