

# Chapitre 3 : Intelligence Artificielle

## Introduction

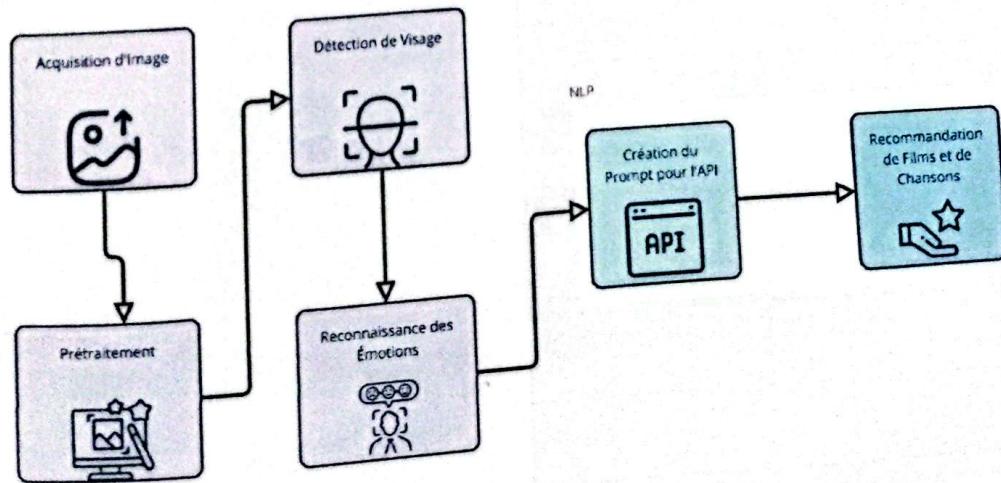
L'intelligence artificielle (IA) vise à créer des systèmes capables de simuler l'intelligence humaine, incluant l'apprentissage, le raisonnement et la perception visuelle. Parmi les applications prometteuses de l'IA, la reconnaissance des émotions permet de détecter et d'interpréter les états émotionnels humains à partir d'images faciales.

Les technologies clés dans ce domaine incluent les Convolutional Neural Networks (CNN) pour l'extraction de caractéristiques visuelles, OpenCV pour le traitement d'images, et Keras pour la construction de modèles de deep learning. En combinant ces outils avec le traitement du langage naturel (NLP), auquel on fournit des prompts générés, il est possible de développer des systèmes capables de fournir des recommandations personnalisées, telles que des films et des chansons, en fonction de l'humeur de l'utilisateur, enrichissant ainsi son expérience globale.

## 1 Flux de travail (Workflow) :

Le diagramme suivant présente le flux de travail du processus de reconnaissance des émotions. Ce flux de travail fournit une vue d'ensemble claire des étapes clés impliquées dans la détection et la classification des émotions à partir des images (ou vidéos, selon le cas). Il met en évidence le déroulement séquentiel des tâches, les points de décision et les interactions entre les différentes composantes du système.

CNS



*Figure 25: Diagramme du flux de travail pour la reconnaissance des émotions*

## 2 Datasets

### 2.1 Importance des Datasets de Qualité

Un dataset de haute qualité est essentiels pour entraîner des modèles performants. Il doit être bien étiqueté et diversifié pour représenter fidèlement les variations des émotions humaines.

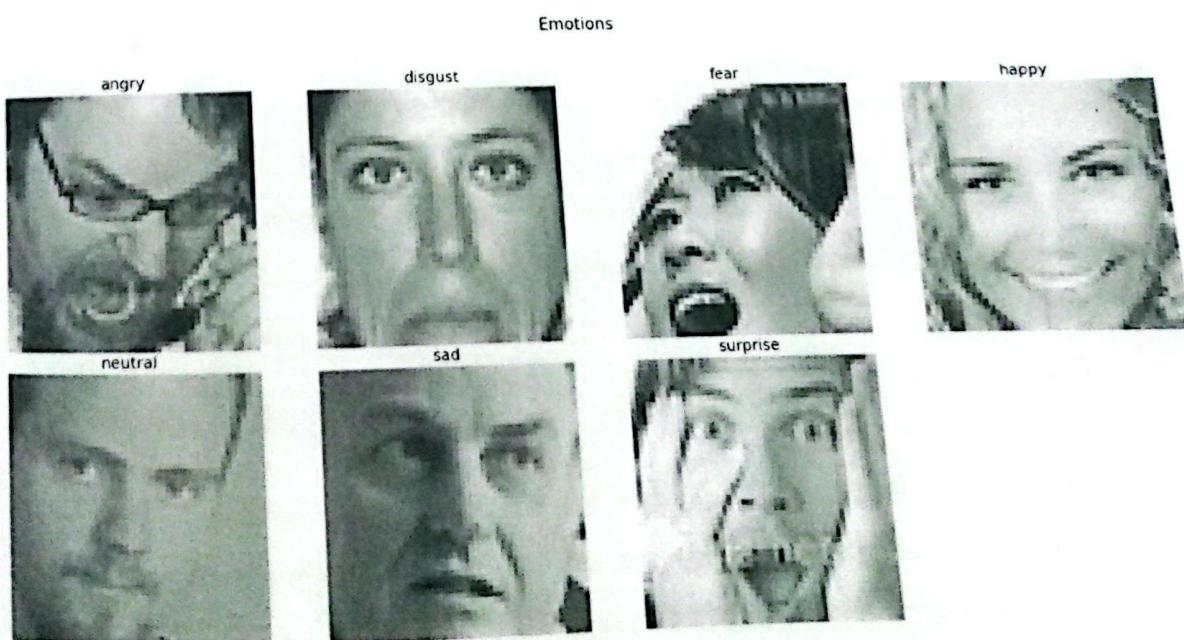
### 2.2 Dataset Utilisé (FER-2013 Dataset)

Le dataset FER-2013 (Facial Expression Recognition 2013) est largement utilisé pour l'entraînement et l'évaluation des modèles de reconnaissance des émotions. Introduit lors des ICML 2013 Challenges in Representation Learning, il contient des milliers d'images faciales annotées.

#### Contenu

Images : Images en niveaux de gris de 48x48 pixels de visages.

Labels : Chaque image est étiquetée avec l'une des sept catégories d'émotions : Angry, Disgust, Fear, Happy, Sad, Surprise, Neutral.



*Figure 26: Les Emotions dans la base de données [17]*

## Structure

Training Set : 28,709 images

Public Test Set : 3,589 images

Private Test Set : 3,589 images

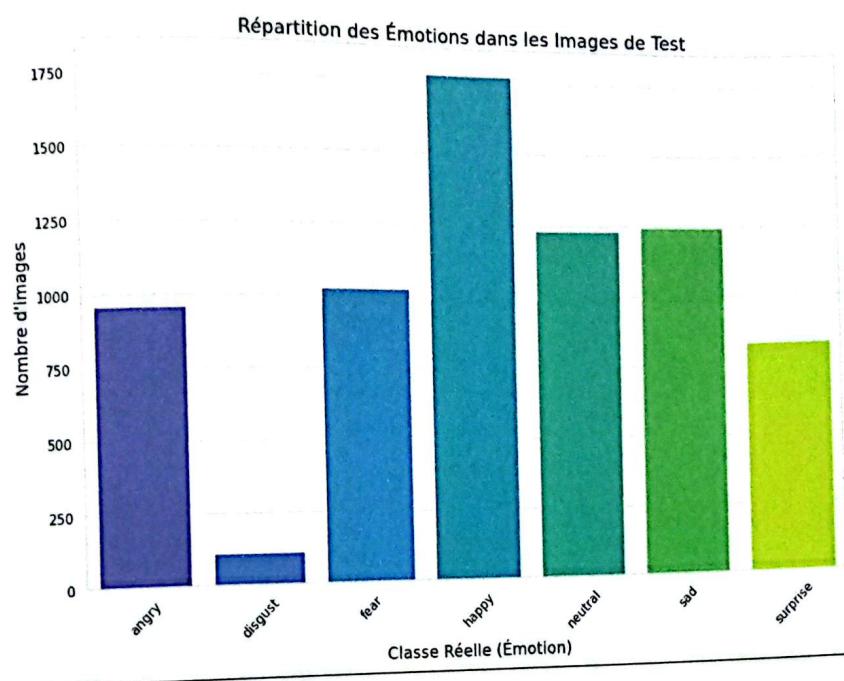


Figure 27: Répartition des émotions dans la base de données de Test

## Utilisation

Le dataset est utilisé pour entraîner des Convolutional Neural Networks (CNNs) afin de détecter les émotions dans les images faciales. Les images, bien que de basse résolution, présentent une variété d'expressions faciales, ce qui en fait un ensemble de données complet et stimulant pour le développement de systèmes robustes de reconnaissance des émotions.

## 3 Preprocessing

Le prétraitement des données est une étape cruciale pour garantir que les images utilisées pour l'entraînement des modèles sont de haute qualité et prêtes à être analysées. Cela comprend plusieurs techniques et étapes visant à améliorer la performance et la robustesse du modèle de reconnaissance des émotions.

### 3.1 Normalisation des Données

La normalisation assure que toutes les valeurs de pixels des images sont dans une gamme similaire, typiquement [0, 1] ou [-1, 1]. Cela aide à accélérer la convergence pendant l'entraînement et à améliorer la performance du modèle.

### 3.2 Étapes et Techniques de Prétraitement

**Redimensionnement des Images :** Toutes les images sont redimensionnées à une taille uniforme (par exemple, 48x48 pixels) pour maintenir la cohérence et réduire la charge computationnelle.

**Conversion en Niveaux de Gris :** Étant donné que les images FER-2013 sont en niveaux de gris, s'assurer que toutes les images d'entrée sont converties en niveaux de gris si ce n'est pas déjà fait.

**Réduction du Bruit :** Appliquer des filtres pour réduire le bruit et améliorer la clarté des caractéristiques faciales.

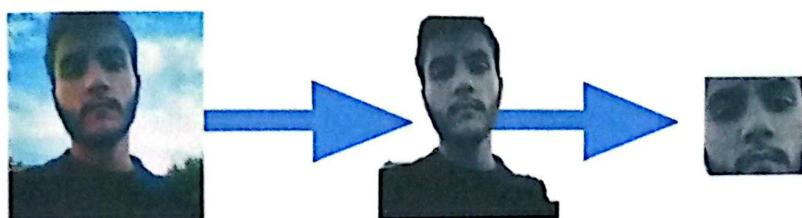


Figure 28: Traitement d'image

### 3.3 Détection de Visage et Extraction des Caractéristiques Faciales

**Détection de Visage :** Utiliser des bibliothèques comme OpenCV pour détecter les visages dans les images, permettant ainsi de se concentrer sur la région d'intérêt (ROI).

**Extraction des Caractéristiques Faciales :** Extraire les principales caractéristiques faciales (yeux, bouche, sourcils) essentielles pour la reconnaissance des émotions. Cela peut impliquer des techniques telles que l'histogramme des gradients orientés (HOG) ou l'utilisation de modèles pré-entraînés pour la détection de points de repère faciaux.

## 4 Modèles Utilisés

### 4.1 Réseaux de Neurones Convolutifs (CNNs)

#### 4.1.1 Comprendre l'Architecture des CNNs

Les CNNs sont composés de plusieurs couches organisées de manière hiérarchique, avec des couches initiales se concentrant sur l'extraction de caractéristiques de bas niveau et des couches plus profondes combinant ces caractéristiques pour détecter des formes plus complexes.

#### 4.1.2 Les Couches du Réseau CNNs

Les CNNs passent une image à travers une série de couches qui appliquent des filtres pour détecter différentes caractéristiques. Les filtres convolutifs glissent sur l'image et produisent des cartes de caractéristiques qui mettent en évidence les aspects pertinents de l'image.

- **Couches Convolutives** : Appliquent des filtres pour produire des cartes de caractéristiques.
- **Couches de Pooling** : Réduisent les dimensions des cartes de caractéristiques en combinant les valeurs dans une région locale.
- **Couches Personnalisées** : Convertissent les cartes de caractéristiques 2D en vecteur 1D (Flatten) et effectuent la classification (Dense).

### 4.2 Le modèle MobileNet

MobileNet est une architecture de réseau de neurones convolutifs (CNN) conçue pour les appareils mobiles et embarqués où les ressources de calcul sont limitées. MobileNet utilise des convolutions séparables en profondeur pour réduire le nombre de paramètres et les calculs nécessaires, tout en maintenant de bonnes performances. L'architecture se compose de :

**Convolutions Séparables en Profondeur** : Une convolution en profondeur suivie d'une convolution ponctuelle.

**Convolutions Standard** : Utilisées au début et à la fin du réseau.

**Couches de Pooling** : Réduisent la dimensionnalité des cartes de caractéristiques.

**Couches Dense** : Utilisées à la fin pour la classification.

### 4.3 Bibliothèques Python Utilisées

Les bibliothèques Python sont essentielles pour la manipulation des données, la construction de modèles de deep learning, le traitement des images et la visualisation des résultats. Voici les bibliothèques les plus utilisées dans ce projet :

Keras : Une bibliothèque de réseaux de neurones de haut niveau, intégrée avec TensorFlow, qui simplifie la création et l'entraînement de modèles de deep learning.

OpenCV : Une bibliothèque open-source pour le traitement des images et la vision par ordinateur. Elle est couramment utilisée pour la détection des visages et l'extraction des caractéristiques faciales.

NumPy : Une bibliothèque fondamentale pour le calcul scientifique avec Python. Elle prend en charge des tableaux et matrices de grande taille, ainsi qu'une large collection de fonctions mathématiques de haut niveau.

Matplotlib : Une bibliothèque de visualisation 2D qui permet de générer des graphiques et des visualisations interactives pour comprendre et interpréter les résultats de l'entraînement du modèle.

## 5 Reconnaissance des Émotions

### 5.1 Evaluer les Caractéristiques Faciales :

Les caractéristiques spécifiques du visage, comme les yeux, la bouche et les sourcils, sont cruciales pour la reconnaissance des émotions. Voici comment ces éléments sont analysés :

- Yeux :

Ouverture et Fermeture : La taille des yeux peut indiquer des émotions comme la surprise (yeux grands ouverts) ou la tristesse (yeux partiellement fermés).

Direction du Regard : Le mouvement et la direction du regard peuvent fournir des indices supplémentaires sur l'état émotionnel.

- Bouche :

Forme et Position : Une bouche souriante indique généralement la joie, tandis qu'une bouche frownée indique la tristesse ou la colère.

Ouverture : Une bouche ouverte peut suggérer la surprise ou le rire.

- Sourceils :

Position et Forme : Des sourcils levés peuvent indiquer la surprise ou l'inquiétude, tandis que des sourcils froncés sont souvent associés à la colère ou la concentration.

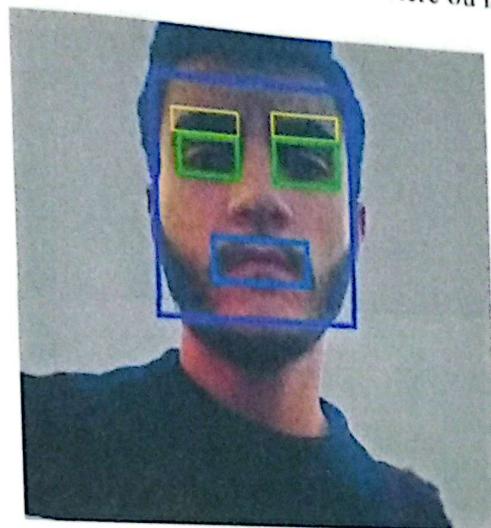


Figure 29 : Extraction des Caractéristiques Faciales

## 5.2 Classification des Émotions :

MobileNet, un modèle de CNN optimisé, utilise les caractéristiques extraites pour classer les émotions. Les couches finales appliquent une activation softmax pour déterminer l'émotion présente dans l'image.

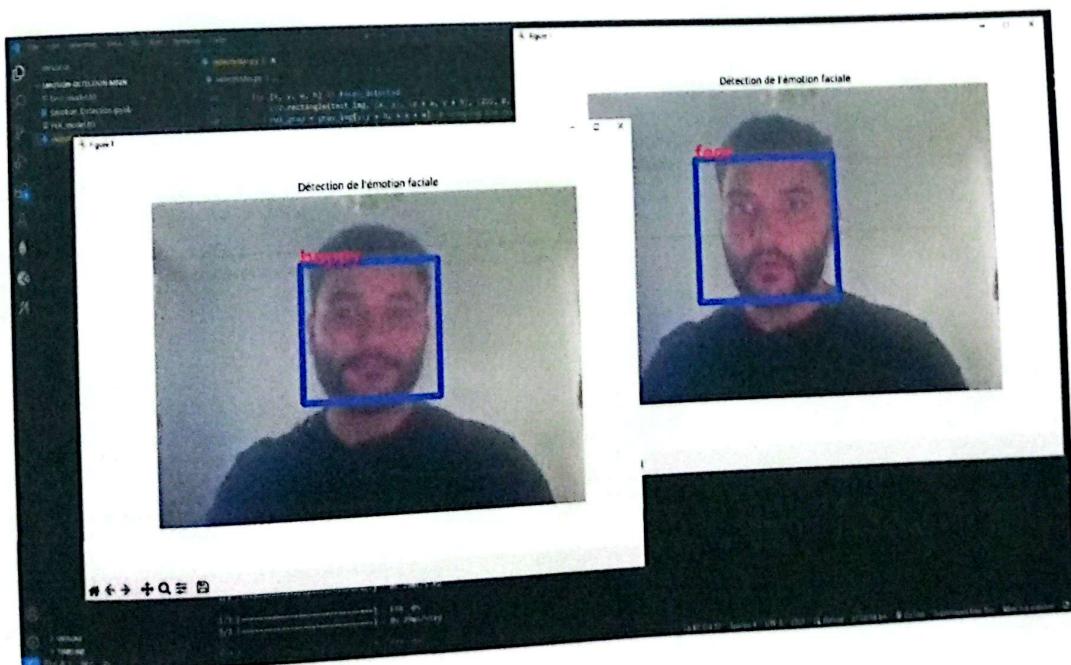


Figure 30: Capture du visage en temps réel via webcam et détection des émotions faciales.

## 6 Evaluation

L'évaluation du modèle est une étape cruciale pour s'assurer que le modèle de reconnaissance des émotions fonctionne efficacement et peut généraliser à de nouvelles données. Voici les principales composantes de l'évaluation :

### 6.1 Matrice de confusion :

Cette matrice nous permet de visualiser les performances du modèle en montrant le nombre de prédictions correctes et incorrectes pour chaque classe d'émotions. Elle aide à identifier les types d'erreurs que le modèle fait le plus souvent.

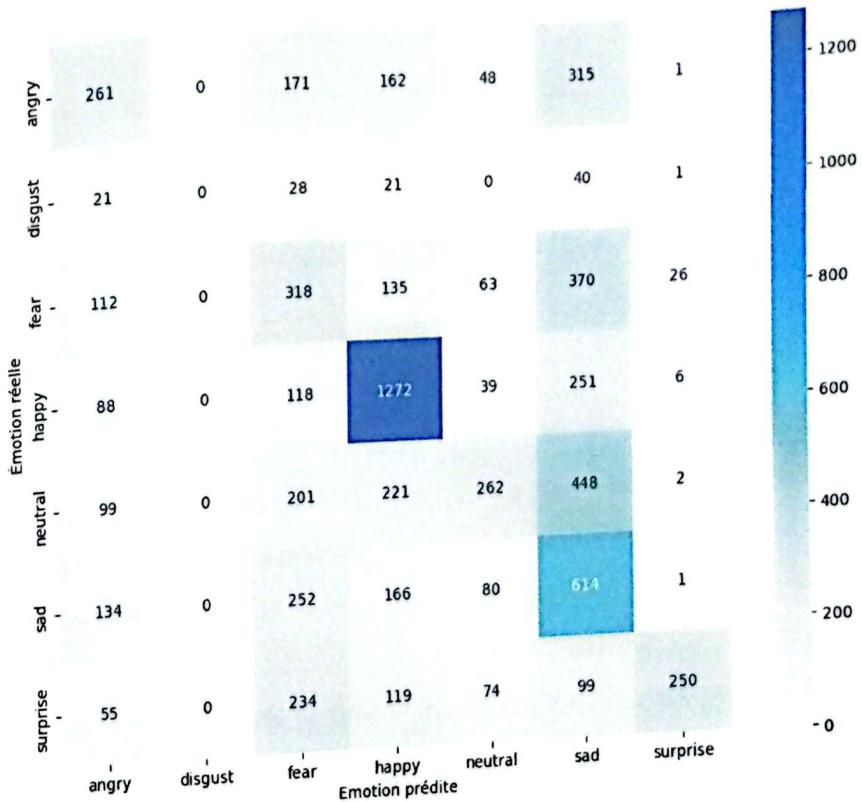


Figure 31: Matrice de confusion

Cette matrice montre comment le modèle a prédit chaque émotion par rapport aux émotions réelles. Par exemple, pour les images d'émotion "heureux", le modèle a correctement prédit cette émotion 1272 fois, mais a également confondu cette émotion avec "colère", "peur", "neutre", etc., un certain nombre de fois. Cela nous permet de voir les tendances de confusion et d'améliorer le modèle en conséquence.

*Tableau 7 : Évaluation des Performances du Modèle par Classe d'Émotion*

Emotion	Precision	Recall	F1-Score	Support
Angry	0.34	0.27	0.30	958
Disgust	0.00	0.00	0.00	111
Fear	0.24	0.31	0.27	1024
Happy	0.61	0.72	0.66	1774
Neutral	0.46	0.21	0.29	1233
Sad	0.29	0.49	0.36	1247
Surprise	0.87	0.30	0.45	831

- Precision : Proportion de prédictions correctes parmi celles faites.
- Recall : Proportion de véritables éléments détectés.
- F1-score : Équilibre entre précision et rappel.
- Support : Nombre d'exemples réels pour chaque classe.

## 7 Recommandations des médias

### 7.1 Génération du Prompt :

Dans cette section, nous abordons la manière dont le système génère des prompts pour l'API en fonction des émotions détectées et des préférences de l'utilisateur. Une fois l'émotion identifiée, le système prend en compte divers paramètres, tels que la langue préférée de l'utilisateur et l'émotion recherchée dans le contenu recommandé. Le prompt généré, par exemple, pourrait être "Je souhaite un film en anglais qui suscite une émotion de bonheur."

*Tableau 8 : Processus de génération du prompt*

Langue	Média	Genre	Émotion	Prompt
Français	Film	Comédie	Bonheur	"Recommande un film comédie en français qui inspire le bonheur."
Anglais	Film	Drame	Tristesse	" Recommande un film drame en anglais qui évoque la tristesse."
Espagnol	Film	Documentaire	Peur	" Recommande un film documentaire en espagnol qui suscite de la peur."

Anglais	Film	Thriller	Neutre	" Recommande un film en anglais qui reste neutre en termes d'émotion, peut-être un thriller."
Espagnol	Film	Drame	Colère	" Recommande un film en espagnol qui exprime de la colère."
Français	Musique	Pop	Bonheur	" Recommande une chanson pop en français qui apporte du bonheur."
Anglais	Musique	Rock	Tristesse	" Recommande une chanson rock en anglais qui évoque la tristesse."

## 7.2 Intégration du NLP :

Le traitement du langage naturel (NLP) joue un rôle essentiel à ce stade. Une fois le prompt créé, le NLP intervient pour analyser et comprendre le langage naturel utilisé dans le prompt, garantissant qu'il est correctement interprété par l'API de recommandation, comme Gemini Flash 1.5. Le NLP permet également de s'assurer que les requêtes sont formulées de manière claire et précise, ce qui augmente la pertinence des recommandations fournies.



Figure 32: Gemini (NLP) [19]

## 7.3 Recommandation de Films et de Chansons :

Après la génération du prompt et son traitement par le NLP, l'API, en l'occurrence Gemini Flash 1.5, fournit une recommandation de médias (films ou chansons) en fonction des critères spécifiés. Cette recommandation est ensuite présentée à l'utilisateur, contribuant à une expérience personnalisée et enrichie. Cette section détaille également la manière dont les résultats sont filtrés et présentés pour maximiser la satisfaction de l'utilisateur.

Tableau 9: Prompt et Recommandation

Prompt	Résultat
"Je souhaite un <b>film</b> en français qui inspire le bonheur."	<i>Amélie</i>
"Je veux un <b>film</b> en anglais qui évoque la tristesse."	<i>The Pursuit of Happyness</i>
"Je cherche un <b>film</b> d'animation en espagnol qui suscite de la peur."	<i>El Laberinto del Fauno</i>

"Je veux un <b>film</b> en anglais qui reste neutre en termes d'émotion, peut-être un thriller."	<i>The Social Network</i>
"Je veux une <b>chanson</b> pop en français qui apporte du bonheur."	<i>Aux Champs-Élysées</i>
"Je recherche une <b>chanson</b> rock en anglais qui évoque la tristesse."	<i>Tears in Heaven</i>

## Conclusion :

En conclusion, l'intelligence artificielle rend possible une expérience utilisateur plus humaine et personnalisée, en recommandant des films et des chansons adaptés à l'humeur du moment. Grâce à des technologies comme les réseaux de neurones, OpenCV pour l'analyse des images, et Keras pour le deep learning, il est maintenant possible de reconnaître et comprendre les émotions à partir de simples images. En associant cela au traitement du langage naturel (NLP), on peut créer des recommandations sur mesure, rendant l'interaction avec les contenus numériques plus naturelle et agréable.

saisis. Si le compte est activé et les informations sont correctes, un jeton d'authentification « Token » lui sera attribué et sera redirigé vers l'interface principale. Sinon, un message d'erreur lui sera affiché.

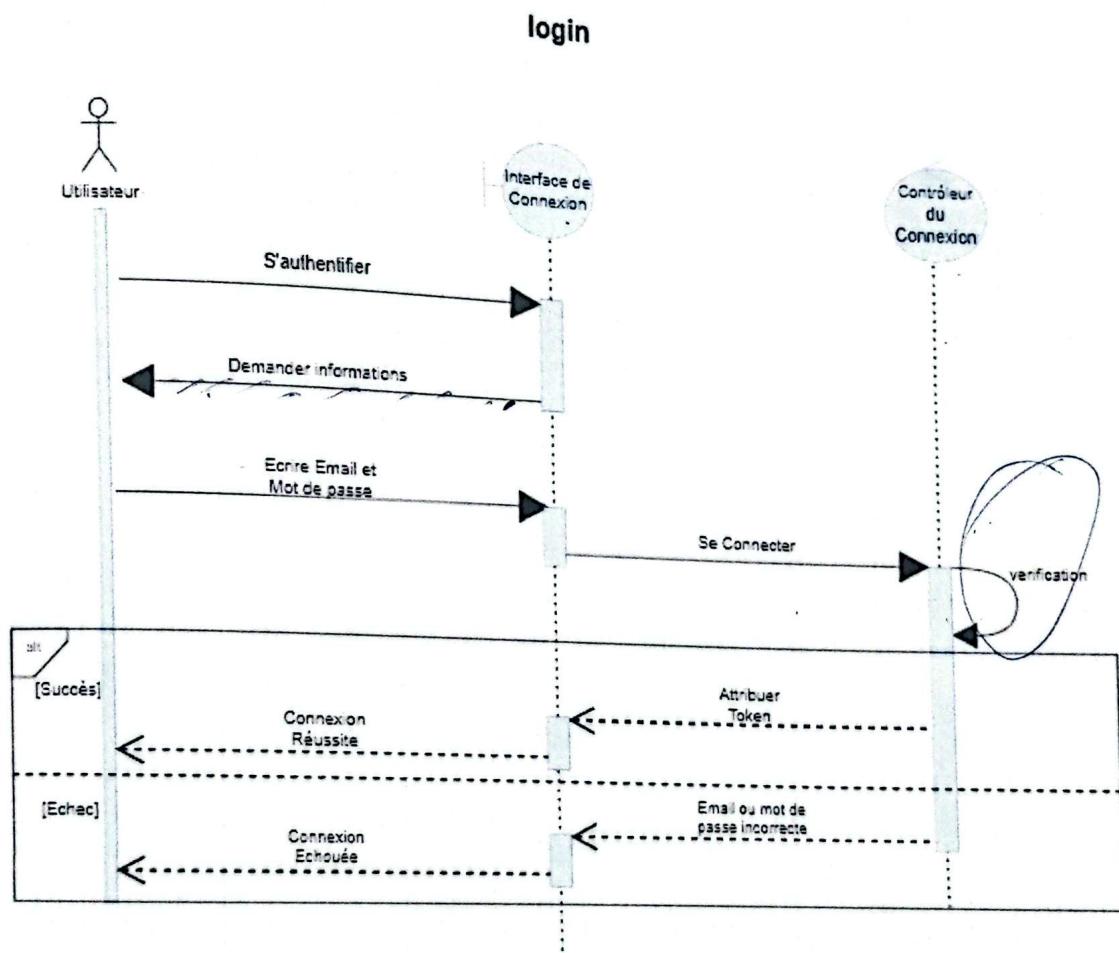


Figure 22 : Diagramme de séquence de « S'authentifier »

## 4.2 Diagramme de séquence de cas d'utilisation «Avoir Recommandation»

L'utilisateur a la possibilité d'avoir des recommandations de média qu'il a choisi, il doit demander l'interface de visualisation média, le system affiche l'interface et demande à l'utilisateur de choisir ses préférences du média.