

RAPPORT DE PROJET : Data Mining

RECONNAISSANCE D'ÉMOTIONS FACIALES

Réalisé par :

- BENBAZE
YOUSSEF

Classe :

- 5IIR3

DÉDICACE

Nous dédions ce modeste travail, fruit de plusieurs mois d'efforts, de recherches intenses et de persévérance technique, à toutes celles et ceux qui, de près ou de loin, ont contribué à notre réussite et ont marqué notre parcours académique au sein de l'**EMSI**.

Nous le dédions avant tout à **nos parents**, pour leur amour inconditionnel, leurs sacrifices constants et leur soutien sans faille. Leur patience, leurs encouragements et leur confiance inébranlable ont toujours été pour nous une source de motivation vitale et d'inspiration. Sans eux, ce chemin vers l'ingénierie n'aurait jamais été possible, et c'est avec une immense gratitude que nous leur offrons ce travail en signe de reconnaissance éternelle.

Nous dédions également ce rapport à **nos frères, nos sœurs et à toute notre famille**, qui n'ont cessé de croire en notre potentiel et de nous soutenir dans les moments de doute comme dans les moments de réussite technologique. Leur présence chaleureuse a illuminé notre parcours.

À **nos enseignants**, qui par leur dévouement, leurs conseils avisés et la haute qualité de leurs enseignements, nous ont transmis le savoir nécessaire pour maîtriser les enjeux du Deep Learning et franchir chaque étape de notre formation en **5IIR3**. Nous leur dédions ce travail en témoignage de notre profond respect.

À **nos amis et compagnons de route**, avec qui nous avons partagé les défis de ce projet, nous adressons une pensée particulière. Votre solidarité et votre amitié resteront gravées dans nos mémoires comme un soutien précieux durant nos années à Rabat.

Enfin, que cette œuvre soit un reflet de notre gratitude infinie et un hommage à tout ce que nous avons reçu en soutien, en amour et en enseignement.

REMERCIEMENTS

La réalisation de ce projet portant sur la **reconnaissance automatique des émotions faciales** à partir d'images, basée sur les techniques de **Deep Learning**, n'aurait pas été possible sans l'aide et le soutien de plusieurs personnes.

Nous adressons tout d'abord nos sincères remerciements à l'administration de l'**École Marocaine des Sciences de l'Ingénieur (EMSI)** pour le cadre pédagogique de qualité et les ressources mises à notre disposition, qui ont grandement favorisé l'apprentissage et la mise en pratique des technologies avancées en **Intelligence Artificielle** et en **Vision par Ordinateur**.

Nous exprimons notre profonde gratitude à notre **professeur encadrant** pour son accompagnement constant, ses conseils méthodologiques avisés et sa rigueur scientifique. Son expertise nous a permis de mieux appréhender les enjeux techniques liés à la reconnaissance des émotions faciales, ainsi que les applications et limites de l'IA dans le domaine de l'interaction homme-machine et de l'analyse comportementale.

Nous tenons également à remercier l'ensemble du **corps professoral** pour la qualité des enseignements dispensés tout au long de notre formation, qui ont constitué une base solide pour la réalisation de ce projet. Enfin, nous adressons nos remerciements aux **membres du jury** pour l'évaluation de notre travail, ainsi que pour leurs remarques pertinentes et constructives, qui ont contribué à l'amélioration de ce projet académique.

TABLE DES MATIERES

1. Résumé et Mots-clés	4
2. Introduction	6
3. Revue de Littérature	7
4. Description du Jeu de Données	8
5. Méthodologie.....	11
5.1 Architecture du Modèle	Error! Bookmark not defined.
5.2 Environnement et Outils	Error! Bookmark not defined.
5.3 Apprentissage et Optimisation.....	Error! Bookmark not defined.
6. Résultats et Analyse.....	11
7. Discussion	14
8. Conclusion	18
9. Références	19
10. Réalisation et simulation	21

1. Résumé et Mots-clés

Résumé :

Ce projet porte sur le développement d'un **système automatisé de reconnaissance des émotions faciales** à partir d'images de visages, en utilisant des techniques avancées de **Deep Learning**. Le système repose sur un **réseau de neurones convolutifs (CNN)** entraîné sur le dataset **FER-2013**, permettant une **classification multi-classes** des expressions faciales en sept émotions fondamentales : *Angry, Disgust, Fear, Happy, Neutral, Sad* et *Surprise*.

Le modèle est intégré dans une architecture applicative complète composée d'un **backend Flask**, chargé du prétraitement des images et de l'inférence du modèle, ainsi que d'un **frontend développé en HTML, CSS et JavaScript**, offrant une interface simple et intuitive pour le téléversement d'images et la visualisation des résultats. Les performances obtenues démontrent la capacité du système à reconnaître efficacement les émotions faciales, soulignant le rôle croissant de l'**Intelligence Artificielle** dans les domaines de l'**interaction homme-machine**, de l'**affective computing** et de l'analyse comportementale.

Mots-clés :

Intelligence Artificielle (IA)

L'Intelligence Artificielle regroupe l'ensemble des techniques permettant à une machine de simuler des capacités cognitives humaines telles que l'apprentissage, la perception et la prise de décision. Dans le contexte de la reconnaissance des émotions, l'IA permet d'interpréter automatiquement les expressions faciales humaines.

Deep Learning (Apprentissage Profond)

Le Deep Learning est une sous-branche de l'IA basée sur des réseaux de neurones profonds. Il permet l'apprentissage automatique de représentations complexes à partir de données brutes telles que les images de visages.

Vision par Ordinateur (Computer Vision)

La vision par ordinateur est un domaine de l'IA qui permet aux machines d'interpréter et de comprendre le contenu des images et des vidéos. Elle transforme des pixels en informations exploitables, telles que la détection ou la classification d'objets. En imagerie médicale, elle est utilisée pour analyser des examens comme l'IRM ou le scanner.

Réseaux de Neurones Convolutifs (CNN)

Les réseaux de neurones convolutifs sont des modèles de Deep Learning spécialement conçus pour le traitement des images. Ils utilisent des couches de convolution pour extraire automatiquement des caractéristiques visuelles, telles que les formes et les textures. Les CNN sont largement utilisés en imagerie médicale pour la détection et la classification de pathologies.

Reconnaissance des émotions faciales

La reconnaissance des émotions faciales consiste à analyser les expressions du visage humain afin d'identifier automatiquement l'état émotionnel d'un individu. Elle repose sur l'observation de variations subtiles des traits du visage tels que les sourcils, les yeux, la bouche et les muscles faciaux.

Cette discipline joue un rôle central dans des domaines comme l'interaction homme-machine, l'affective computing, l'analyse comportementale et le suivi du bien-être psychologique. Grâce aux techniques de Deep Learning, notamment les réseaux de neurones convolutifs, il est désormais possible d'atteindre des performances élevées dans l'interprétation automatique des émotions à partir d'images faciales.

FER-2013

FER-2013 (Facial Expression Recognition 2013) est un jeu de données de référence largement utilisé dans la recherche scientifique pour l'apprentissage et l'évaluation des systèmes de reconnaissance des émotions faciales. Il est composé d'environ 35 000 images de visages en niveaux de gris, de taille 48x48 pixels, annotées selon sept classes émotionnelles : *Angry*, *Disgust*, *Fear*, *Happy*, *Neutral*, *Sad* et *Surprise*. La diversité des expressions, des conditions d'éclairage et des morphologies faciales présentes dans ce dataset en fait une base pertinente pour entraîner des modèles robustes, tout en représentant un défi en raison de la faible résolution des images.

Flask

Flask est un framework web léger et flexible basé sur Python, utilisé pour le développement d'applications web et d'API. Il est particulièrement apprécié pour sa simplicité, sa modularité et sa facilité d'intégration avec des bibliothèques de Deep Learning telles que TensorFlow et Keras. Dans le cadre de ce projet, Flask permet de déployer le modèle de reconnaissance des émotions, de gérer les requêtes utilisateur (envoi d'images faciales) et de retourner les prédictions sous forme de résultats exploitables par l'interface utilisateur.

Classification multi-classes

La classification multi-classes est une tâche d'apprentissage supervisé qui consiste à **attribuer une observation à une seule classe parmi plusieurs catégories possibles**. Contrairement à la classification binaire, elle permet de gérer des problèmes plus complexes impliquant plusieurs états ou labels.

Dans ce projet, la classification multi-classes est utilisée pour **identifier l'une des sept émotions faciales** à partir d'une image donnée. Le modèle produit une probabilité pour chaque émotion grâce à une fonction d'activation *softmax*, et l'émotion ayant la probabilité la plus élevée est considérée comme la prédiction finale.

2. Introduction

Contexte du projet

Les émotions humaines jouent un rôle fondamental dans la communication et les interactions sociales. Elles s'expriment notamment à travers les **expressions faciales**, qui constituent l'un des moyens les plus naturels et les plus riches de transmission d'informations émotionnelles. La capacité à reconnaître et interpréter ces émotions est essentielle dans des domaines tels que **l'interaction homme-machine**, **l'affective computing**, la **robotique sociale**, ainsi que dans certaines applications liées au **suivi du bien-être mental** et à l'analyse comportementale.

Avec le développement des systèmes intelligents et l'augmentation des données visuelles disponibles, la **Vision par Ordinateur** et le **Deep Learning** offrent aujourd'hui des solutions performantes pour automatiser l'analyse des expressions faciales. Les **réseaux de neurones convolutifs (CNN)**, en particulier, ont démontré une grande efficacité dans l'extraction automatique de caractéristiques visuelles complexes à partir d'images de visages, permettant ainsi la reconnaissance fiable des émotions.

Problématique

Malgré les avancées significatives en reconnaissance faciale, l'interprétation automatique des émotions demeure une tâche complexe. Les expressions faciales peuvent varier considérablement selon les individus, les conditions d'éclairage, l'orientation du visage, ainsi que la subtilité de certaines émotions. De plus, certaines expressions émotionnelles présentent des similarités visuelles importantes, ce qui complique leur distinction automatique.

La problématique centrale de ce projet consiste donc à **concevoir un système automatisé capable d'analyser des images faciales et de prédire de manière fiable l'émotion exprimée parmi plusieurs catégories**, tout en restant suffisamment robuste face aux variations visuelles et simple à intégrer dans une application web destinée à l'utilisateur final.

Objectifs du travail

Les principaux objectifs de ce projet sont :

- Développer un **modèle de Deep Learning basé sur un réseau de neurones convolutifs (CNN)** pour la reconnaissance des émotions faciales selon une **classification multi-classes** (sept émotions).
- Mettre en place une **chaîne complète de traitement des images**, allant du prétraitement des visages jusqu'à l'inférence du modèle.
- Concevoir une **application web** reposant sur un **backend Flask** et une interface frontend en **HTML, CSS et JavaScript**, permettant une interaction fluide avec le système.
- Évaluer les **performances du modèle** à l'aide de métriques adaptées et analyser ses **limites dans des conditions réelles d'utilisation**.

3. Revue de Littérature

Approches classiques en reconnaissance des émotions faciales

Avant l'essor du Deep Learning, la reconnaissance des émotions faciales reposait principalement sur des **méthodes classiques de traitement d'images et de vision par ordinateur**. Ces approches consistaient généralement à détecter le visage à l'aide de techniques telles que les **Haar Cascades**, puis à extraire manuellement des caractéristiques discriminantes, notamment les textures (LBP – Local Binary Patterns), les gradients (HOG – Histogram of Oriented Gradients) ou des mesures géométriques des traits du visage. Ces caractéristiques étaient ensuite utilisées par des classificateurs traditionnels tels que les **SVM**, les **k-plus proches voisins (k-NN)** ou les **réseaux de neurones peu profonds**. Bien que ces méthodes aient permis des avancées notables, elles présentent plusieurs limites, notamment leur forte dépendance aux paramètres définis manuellement, leur sensibilité aux variations d'éclairage et de pose, ainsi que leur faible capacité de généralisation face à des expressions faciales complexes ou subtiles.

Apport des réseaux de neurones convolutifs

Les **réseaux de neurones convolutifs (CNN)** ont profondément transformé le domaine de la reconnaissance des émotions faciales. Contrairement aux approches classiques, les CNN permettent d'apprendre automatiquement des **représentations hiérarchiques** à partir des images brutes, sans nécessiter une extraction manuelle des caractéristiques.

Grâce à l'empilement de couches de convolution, de pooling et de fonctions d'activation non linéaires, les CNN sont capables de capturer des motifs visuels de plus en plus complexes, allant des contours simples du visage aux expressions émotionnelles globales. De nombreuses études ont montré que les CNN surpassent largement les méthodes traditionnelles sur des jeux de données de référence tels que **FER-2013**, notamment pour les tâches de classification multi-classes des émotions.

Rôle des architectures CNN personnalisées

Dans le contexte de la reconnaissance des émotions faciales, l'utilisation d'**architectures CNN personnalisées** s'est révélée particulièrement efficace. Contrairement aux architectures très profondes pré-entraînées sur des images naturelles, les modèles dédiés à FER-2013 sont généralement optimisés pour des **images de petite taille (48×48 pixels)** et en niveaux de gris.

Ces architectures sont composées de plusieurs couches de convolution avec des filtres adaptés, suivies de couches de **MaxPooling** pour la réduction spatiale et de **Dropout** pour limiter le surapprentissage. Les couches entièrement connectées finales, associées à une fonction d'activation **softmax**, permettent de réaliser une **classification multi-classes** des émotions faciales. Cette approche offre un bon compromis entre performance, complexité du modèle et temps de calcul, ce qui la rend adaptée à des applications en temps réel.

4. Description du Jeu de Données

4.1 Source et contexte des données

Le jeu de données utilisé dans ce projet provient de la plateforme **Kaggle**, reconnue pour héberger des ensembles de données fiables et largement utilisés dans la recherche en **science des données** et en **Intelligence Artificielle**. Le dataset sélectionné est **FER-2013 (Facial Expression Recognition 2013)**, spécifiquement conçu pour la reconnaissance automatique des émotions faciales à partir d'images.

Ce jeu de données est largement utilisé dans la communauté scientifique comme **benchmark** pour l'évaluation des modèles de reconnaissance des émotions. Les images ont été collectées à partir de sources variées et annotées manuellement, garantissant une diversité suffisante des expressions faciales tout en respectant les considérations éthiques liées à l'utilisation de données visuelles.

4.2 Composition et structure du dataset

Le dataset FER-2013 est composé d'environ **35 000 images de visages**, organisées selon **sept classes émotionnelles distinctes** :

- **Angry (Colère)**
- **Disgust (Dégoût)**
- **Fear (Peur)**

- **Happy (Joie)**
- **Neutral (Neutre)**
- **Sad (Tristesse)**
- **Surprise (Surprise)**

Chaque image est associée à une **étiquette explicite** indiquant l'émotion correspondante, ce qui permet un apprentissage supervisé. Cette classification multi-classes reflète la complexité des expressions faciales humaines et constitue un défi important pour les modèles de Deep Learning, certaines émotions présentant des similarités visuelles marquées.

4.3 Format et caractéristiques des images

Les images du dataset FER-2013 sont fournies sous forme d'images **en niveaux de gris**, avec une résolution fixe de **48 × 48 pixels**. Ce format compact permet de réduire la complexité computationnelle, tout en conservant les informations essentielles nécessaires à l'analyse des expressions faciales.

La faible résolution des images constitue toutefois un défi pour la reconnaissance fine des émotions, car certaines variations subtiles des traits du visage peuvent être difficiles à distinguer. Cette caractéristique rend le dataset particulièrement intéressant pour évaluer la capacité des modèles à apprendre des représentations discriminantes à partir de données limitées.

4.4 Prétraitement des données

Avant l'entraînement du modèle CNN, plusieurs étapes de prétraitement sont appliquées afin d'améliorer la qualité des données et la stabilité de l'apprentissage :

- **Normalisation** : les valeurs des pixels sont mises à l'échelle dans l'intervalle $[0,1]$, ce qui facilite la convergence du modèle et réduit l'impact des variations d'intensité lumineuse.
- **Redimensionnement et mise en forme** : les images sont conservées à la taille 48×48 pixels et remodelées sous la forme $(48, 48, 1)$ afin d'être compatibles avec l'entrée du réseau de neurones convolutifs.
- **Conversion en tenseurs** : les images et leurs étiquettes sont transformées en formats exploitables par les bibliothèques de Deep Learning telles que TensorFlow et Keras.

Ces opérations garantissent une cohérence des données tout en préservant les informations visuelles essentielles à la reconnaissance des émotions faciales.

4.5 Data Augmentation

Afin de limiter le risque de **surapprentissage (overfitting)** et d'améliorer la robustesse du modèle face aux variations réelles des expressions faciales, des techniques de **Data Augmentation** sont appliquées durant la phase d'entraînement. Ces techniques consistent à générer artificiellement de nouvelles images à partir des images existantes, notamment par :

- des **rotations légères** pour simuler des variations d'orientation du visage,
- des **translations horizontales et verticales** pour représenter des différences de cadrage,
- des **zooms légers** pour reproduire des variations de distance par rapport à la caméra,
- des **retournements horizontaux** afin d'augmenter la diversité des expressions.

La Data Augmentation permet ainsi au modèle d'apprendre des représentations plus générales et de mieux s'adapter à des images provenant de conditions variées.

4.6 Répartition du jeu de données

Le dataset est divisé en trois sous-ensembles distincts afin d'assurer une évaluation fiable des performances du modèle :

- **Jeu d'entraînement (Train set)** : environ **80 %** des images, utilisé pour l'apprentissage des paramètres du réseau de neurones.
- **Jeu de validation (Validation set)** : environ **10 %** des images, servant à ajuster les hyperparamètres et à surveiller le surapprentissage.
- **Jeu de test (Test set)** : environ **10 %** des images, réservé à l'évaluation finale du modèle sur des données jamais vues.

Cette séparation garantit une mesure objective de la capacité de généralisation du modèle dans un contexte réel d'utilisation.

4.7 Importance du jeu de données dans le contexte médical

Dans le domaine de la reconnaissance des émotions faciales, la qualité et la diversité du jeu de données jouent un rôle déterminant dans la fiabilité des résultats. Un dataset représentatif

permet de réduire les biais liés aux variations individuelles, aux conditions d'éclairage et aux différences culturelles.

Ainsi, le dataset FER-2013 constitue une base solide pour le développement d'un système automatisé de reconnaissance des émotions faciales. Il met également en évidence la nécessité d'enrichir continuellement les données et d'explorer des bases plus diversifiées pour des applications à grande échelle et en conditions réelles..

5. Méthodologie

La méthodologie adoptée dans ce projet repose sur une approche complète intégrant le **traitement des images faciales**, la **modélisation par Deep Learning**, ainsi que le **déploiement applicatif via une architecture web légère**. L'objectif est de concevoir un système fiable, reproductible et facilement exploitable pour la **reconnaissance automatique des émotions faciales** à partir d'images.

5.1 Architecture globale du système

Le système proposé suit une architecture modulaire composée de trois blocs principaux :

- **Bloc acquisition et prétraitement des données**
Chargé de la préparation des images faciales afin de les rendre compatibles avec le modèle de Deep Learning.
- **Bloc intelligence artificielle (classification)**
Basé sur un **réseau de neurones convolutifs (CNN)** entraîné pour effectuer une **classification multi-classes** des émotions faciales (sept émotions).
- **Bloc applicatif (backend et frontend)**
Permet l'interaction entre l'utilisateur et le modèle via une application web utilisant **Flask** pour le backend et **HTML/CSS/JavaScript** pour le frontend.

Cette séparation garantit une meilleure maintenabilité du système et facilite les évolutions futures.

5.2 Prétraitement des images faciales

Le prétraitement constitue une étape essentielle pour assurer la stabilité et la performance du modèle. Les images faciales peuvent présenter des variations importantes liées à l'éclairage, à la pose du visage et à la qualité de capture.

Les principales étapes appliquées sont :

- **Mise à l'échelle des images**
Toutes les images sont conservées à une taille fixe de **48 × 48 pixels**, correspondant au format du dataset FER-2013.

- **Normalisation des pixels**
Les valeurs d'intensité sont ramenées dans l'intervalle $[0,1]$, ce qui améliore la convergence de l'apprentissage et réduit les effets des variations lumineuses.
- **Gestion des canaux**
Les images étant en niveaux de gris, elles sont remodelées sous la forme (48, 48, 1) afin d'être compatibles avec l'entrée du réseau convolutif.
- **Augmentation de données (Data Augmentation)**
Des transformations aléatoires (rotations légères, translations, zooms et retournements horizontaux) sont appliquées afin d'enrichir artificiellement le jeu de données et d'améliorer la robustesse du modèle face aux variations réelles des expressions faciales.

5.3 Architecture du modèle de Deep Learning (CNN)

Le cœur du système repose sur un **réseau de neurones convolutifs personnalisé**, conçu spécifiquement pour la reconnaissance des émotions faciales à partir d'images de petite taille.

5.3.1 Principe de l'architecture CNN

L'architecture CNN utilisée est constituée :

- de plusieurs **couches de convolution (Conv2D)** utilisant des filtres de petite taille pour l'extraction des caractéristiques locales,
- de fonctions d'activation **ReLU** introduisant la non-linéarité,
- de couches de **MaxPooling** pour la réduction spatiale et la diminution de la complexité computationnelle,
- de couches **Dropout** pour limiter le surapprentissage,
- de couches entièrement connectées (**Dense**) en fin de réseau pour la classification.

Cette structure permet une extraction hiérarchique des caractéristiques, allant des motifs simples du visage (bords, contours) aux expressions émotionnelles globales.

5.3.2 Adaptation à la classification multi-classes

La couche de sortie du modèle est composée de **sept neurones**, correspondant aux sept émotions faciales. Une fonction d'activation **softmax** est utilisée afin de produire une distribution de probabilités sur l'ensemble des classes.

La classe associée à la probabilité la plus élevée est considérée comme l'émotion prédite, tandis que la valeur de cette probabilité représente le **niveau de confiance du modèle**.

5.4 Stratégie d'apprentissage

5.4.1 Fonction de perte

Étant donné la nature multi-classes du problème, la fonction de perte utilisée est la **Categorical Cross-Entropy**, définie par :

$$L = -\frac{1}{N} \sum_{i=1}^N \sum_{c=1}^C y_{i,c} \log(p_{i,c})$$

où $y_{i,c}$ représente la vérité terrain et $p_{i,c}$ la probabilité prédite pour la classe c .

Cette fonction mesure l'écart entre les prédictions du modèle et les étiquettes réelles, et pénalise fortement les erreurs de classification.

5.4.2 Optimiseur

L'optimiseur **Adam** est utilisé pour minimiser la fonction de perte. Il permet :

- une convergence rapide,
- une meilleure stabilité de l'apprentissage,
- une adaptation automatique du taux d'apprentissage.

5.5 Réglage des hyperparamètres

Les principaux hyperparamètres utilisés sont :

- **Nombre d'époques** : entre 20 et 50, selon la convergence observée.
- **Batch size** : 32 ou 64 images par itération.
- **Learning rate** : fixé généralement à 0.0001 pour éviter des mises à jour trop agressives des poids.

Le suivi de la perte et de la précision sur le jeu de validation permet de détecter le surapprentissage et d'ajuster ces paramètres.

5.6 Évaluation du modèle

Les performances du modèle sont évaluées à l'aide de plusieurs métriques :

- **Accuracy (précision globale)**
- **Precision, Recall et F1-score par classe**, afin d'analyser les performances pour chaque émotion
- **Matrice de confusion**, permettant une analyse détaillée des erreurs et des confusions entre émotions

Ces métriques offrent une vision complète de la capacité du modèle à reconnaître correctement les émotions faciales.

5.7 Déploiement Backend avec Flask

Le modèle entraîné est intégré dans un **backend Flask**, chargé de :

- recevoir les images faciales envoyées par l'utilisateur,
- appliquer le prétraitement nécessaire,
- exécuter l'inférence du modèle,
- retourner le résultat sous forme JSON (émotion prédite et pourcentage de confiance).

Flask permet une communication simple, rapide et facilement extensible.

5.8 Interface utilisateur avec HTML, CSS et JavaScript

Le frontend est développé en **HTML, CSS et JavaScript**, offrant :

- une interface intuitive pour le téléchargement des images faciales,
- l'affichage clair de l'émotion prédite et de son niveau de confiance,
- une interaction fluide avec l'API backend.

Cette architecture légère garantit une bonne expérience utilisateur et une intégration facile du système.

5.9 Synthèse de la méthodologie

La méthodologie adoptée combine :

- des **techniques avancées de Deep Learning**,
- une **architecture CNN adaptée à la reconnaissance des émotions faciales**,
- et une **architecture logicielle web moderne et légère**.

Elle constitue une base solide pour le développement d'un système fiable de reconnaissance automatique des émotions faciales, ouvrant la voie à des applications variées en interaction homme-machine et en analyse comportementale.

6. Résultats et Analyse

Cette section présente les résultats obtenus à l'issue de l'entraînement et de l'évaluation du **modèle de Deep Learning basé sur un réseau de neurones convolutifs (CNN)** dédié à la reconnaissance des émotions faciales. L'analyse porte sur les **performances quantitatives du système**, son comportement face au jeu de test, ainsi que sur l'interprétation des erreurs observées dans un contexte d'analyse des expressions faciales humaines.

6.1 Performances globales du modèle

L'évaluation du modèle est réalisée sur le **jeu de test**, composé d'images faciales que le réseau n'a jamais rencontrées durant la phase d'apprentissage. Les principales métriques utilisées sont l'**accuracy**, la **précision (precision)**, le **rappel (recall)** et le **score F1**, calculés dans un contexte de **classification multi-classes**.

Les résultats obtenus montrent que le modèle atteint une **accuracy globale satisfaisante**, traduisant une bonne capacité générale à distinguer les différentes émotions faciales. Ces performances indiquent que l'architecture CNN adoptée est bien adaptée à l'analyse des images du dataset FER-2013 et à la reconnaissance automatique des émotions.

6.2 Analyse par métriques

- **Accuracy (précision globale)**
Cette métrique représente la proportion totale d'images correctement classées parmi l'ensemble des images du jeu de test. Le score obtenu confirme que le modèle est capable de reconnaître correctement la majorité des expressions faciales.
- **Recall (sensibilité)**
Le rappel mesure la capacité du modèle à identifier correctement les images appartenant à une émotion donnée. Il est particulièrement important pour des émotions difficiles à détecter telles que *Fear* ou *Disgust*, souvent confondues avec d'autres expressions.
- **Precision**
La précision indique la fiabilité des prédictions pour chaque émotion. Une valeur élevée signifie que lorsque le modèle prédit une émotion spécifique, cette prédiction est généralement correcte.
- **Score F1**
Le score F1, moyenne harmonique entre la précision et le rappel, fournit une mesure équilibrée de la performance du modèle pour chaque classe émotionnelle, notamment lorsque certaines émotions sont moins représentées dans le dataset.

6.3 Matrice de confusion

La **matrice de confusion** permet d'analyser en détail les types d'erreurs commises par le modèle pour chaque émotion :

- **Vrais positifs (TP)** : émotions correctement reconnues.
- **Faux positifs (FP)** : émotions incorrectement prédites.
- **Faux négatifs (FN)** : émotions présentes mais non détectées.
- **Vrais négatifs (TN)** : images correctement rejetées pour une émotion donnée.

L'analyse de la matrice de confusion montre que certaines émotions, telles que *Happy* et *Surprise*, sont plus facilement reconnues, tandis que des confusions fréquentes apparaissent

entre des émotions visuellement proches comme *Fear* et *Surprise*, ou *Sad* et *Neutral*. Ces confusions reflètent la similarité des expressions faciales associées à ces émotions.

6.4 Analyse qualitative des erreurs

Une analyse qualitative des images mal classées met en évidence plusieurs facteurs expliquant les erreurs du modèle :

- Similarité visuelle entre certaines expressions émotionnelles.
- Variations d'éclairage affectant la visibilité des traits du visage.
- Expressions faciales peu marquées ou ambiguës.
- Présence d'occlusions partielles (lunettes, cheveux, mains).

Ces cas limites illustrent les défis inhérents à la reconnaissance automatique des émotions faciales et soulignent l'importance de disposer de jeux de données riches et diversifiés.

6.5 Temps de réponse et performance applicative

O Au-delà des performances de classification, le **temps de réponse** constitue un critère essentiel pour une application web interactive. Les tests réalisés montrent que :

- le temps d'inférence du modèle est de l'ordre de **quelques centaines de millisecondes par image**,
- le traitement complet (réception de l'image, prétraitement, prédiction et retour du résultat via Flask) reste **inférieur à une seconde**.

Ces performances garantissent une **expérience utilisateur fluide** et confirment la faisabilité d'une utilisation du système en temps quasi réel.

6.6 Interprétation des résultats dans un contexte applicatif

Les résultats obtenus démontrent que le système peut constituer un **outil d'aide à l'analyse émotionnelle**, capable de fournir une estimation automatique de l'émotion exprimée par un individu. Il ne vise pas à interpréter l'état psychologique de manière définitive, mais à fournir une **indication objective basée sur l'expression faciale**.

Dans des contextes tels que l'interaction homme-machine, l'analyse du comportement utilisateur ou le suivi du bien-être, un tel système peut apporter une valeur ajoutée significative, à condition d'être utilisé avec précaution et dans le respect des considérations éthiques.

6.7 Synthèse des résultats

En résumé, l'analyse des résultats met en évidence :

- une **bonne capacité de généralisation du modèle CNN**,
- des **performances satisfaisantes pour la reconnaissance multi-classes des émotions faciales**,
- une **intégration applicative efficace** via une architecture web basée sur Flask et des technologies frontend légères.

Ces résultats confirment la pertinence de l'approche adoptée et ouvrent la voie à des améliorations futures visant à renforcer la précision, la robustesse et l'applicabilité du système dans des environnements réels.

7. Discussion

Les résultats obtenus dans ce projet montrent que l'utilisation des techniques de **Deep Learning**, et plus particulièrement des **réseaux de neurones convolutifs (CNN)**, est pertinente pour la **reconnaissance automatique des émotions faciales** à partir d'images. Le modèle développé a démontré une **bonne capacité de généralisation**, avec des performances satisfaisantes sur le jeu de test, confirmant l'efficacité des CNN pour l'analyse des expressions faciales humaines.

L'un des principaux points forts du système réside dans sa capacité à reconnaître correctement certaines émotions clairement exprimées, telles que *Happy* et *Surprise*, qui présentent des caractéristiques faciales distinctes. Cette performance est particulièrement intéressante dans des contextes d'**interaction homme-machine**, où une reconnaissance rapide et fiable des émotions peut améliorer l'adaptabilité et la réactivité des systèmes intelligents.

Cependant, certaines limitations ont été observées. Les erreurs de classification concernent principalement des émotions visuellement proches, telles que *Fear* et *Surprise*, ou *Sad* et *Neutral*. Ces confusions s'expliquent par la **subtilité des expressions faciales**, la faible résolution des images du dataset FER-2013, ainsi que par des facteurs externes tels que les variations d'éclairage, la pose du visage ou la présence d'occlusions partielles. Ces observations soulignent l'importance de la **qualité, de la diversité et de la représentativité du jeu de données** utilisé pour l'apprentissage.

Par ailleurs, le modèle repose sur une analyse globale de l'image faciale sans expliciter quelles régions du visage contribuent le plus à la décision finale. Cette absence d'interprétabilité peut constituer une limite dans certaines applications sensibles. Des techniques d'**explicabilité**, telles que Grad-CAM, pourraient être intégrées afin de visualiser les zones du visage les plus influentes dans la prédiction de l'émotion.

Enfin, ce travail ouvre plusieurs perspectives d'amélioration, notamment l'optimisation de l'architecture CNN, l'augmentation du volume et de la diversité des données d'apprentissage, ou encore l'extension du système vers l'analyse vidéo en temps réel. Ces évolutions

permettraient d'accroître la robustesse et la fiabilité du système dans des conditions d'utilisation réelles.

8. Conclusion

Ce projet a porté sur la **conception et le développement d'un système automatisé de reconnaissance des émotions faciales** à partir d'images, en s'appuyant sur les techniques avancées de **Deep Learning** et de **Vision par Ordinateur**. L'objectif principal était de mettre en œuvre un modèle capable d'identifier de manière fiable l'émotion exprimée par un individu parmi sept catégories distinctes.

L'approche adoptée, basée sur un **réseau de neurones convolutifs entraîné sur le dataset FER-2013**, a permis d'obtenir des résultats satisfaisants en termes de précision et de capacité de généralisation. Le prétraitement rigoureux des images, ainsi que l'utilisation de techniques de **Data Augmentation**, ont contribué à améliorer la robustesse du modèle face aux variations des expressions faciales.

Sur le plan applicatif, l'intégration du modèle dans une **architecture web légère**, combinant un **backend Flask** et une interface frontend développée en **HTML, CSS et JavaScript**, a permis de transformer le modèle de classification en une application fonctionnelle, interactive et facilement accessible. Cette implémentation illustre la faisabilité du déploiement pratique de solutions basées sur l'Intelligence Artificielle dans des environnements réels.

En conclusion, ce travail met en évidence le **potentiel de l'Intelligence Artificielle dans le domaine de l'affective computing** et de l'analyse des comportements humains. Bien que le système ne puisse interpréter l'état émotionnel d'un individu de manière exhaustive, il constitue une **base solide pour des développements futurs**, notamment vers la reconnaissance des émotions en temps réel, l'intégration de données multimodales (visage et voix) et l'amélioration de l'explicabilité des modèles.

9. Références

Références liées au dataset et aux plateformes de données

Kaggle. (2013).

FER-2013: Facial Expression Recognition Dataset.

Récupéré depuis :

<https://www.kaggle.com/datasets/msambare/fer2013>

Ce jeu de données constitue la base utilisée pour l'entraînement et l'évaluation du modèle de reconnaissance des émotions faciales. Il regroupe des images de visages en niveaux de gris annotées selon sept émotions fondamentales.

Références liées à la reconnaissance des émotions faciales

Li, S., & Deng, W. (2018).

Facial Expression Recognition: A Survey.

IEEE Transactions on Affective Computing, 10(4), 1–17.

Cet article de référence présente une revue complète des approches classiques et modernes de reconnaissance des émotions faciales, mettant en évidence l'évolution vers les méthodes basées sur le Deep Learning.

Ekman, P. (1992).

An Argument for Basic Emotions.

Cognition & Emotion, 6(3–4), 169–200.

Ce travail fondateur introduit le concept des émotions de base, largement utilisé comme fondement théorique dans les systèmes de reconnaissance des émotions faciales.

Références liées au Deep Learning et aux réseaux de neurones convolutifs

Goodfellow, I., Bengio, Y., & Courville, A. (2016).

Deep Learning.

MIT Press.

Ouvrage de référence présentant les fondements théoriques du Deep Learning, incluant les réseaux de neurones convolutifs et leurs applications en vision par ordinateur.

LeCun, Y., Bengio, Y., & Hinton, G. (2015).

Deep Learning.

Nature, 521(7553), 436–444.

Article clé décrivant les principes du Deep Learning et son impact majeur sur l'analyse d'images et la reconnaissance de formes.

Références liées aux frameworks et bibliothèques (Python)

Abadi, M., Agarwal, A., Barham, P., et al. (2016).

TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems.

Récupéré depuis : <https://www.tensorflow.org>

Cette référence concerne le framework TensorFlow, utilisé pour l'implémentation et l'entraînement du modèle CNN de reconnaissance des émotions faciales.

Chollet, F. (2015).

Keras: The Python Deep Learning Library.

Récupéré depuis : <https://keras.io>

Keras est une bibliothèque de haut niveau facilitant la conception, l'entraînement et l'évaluation des réseaux de neurones convolutifs.

Bradski, G. (2000).

The OpenCV Library.

Dr. Dobb's Journal of Software Tools.

OpenCV est utilisée pour le traitement des images et les opérations de manipulation visuelle.

Références liées au backend et au frontend

Grinberg, M. (2018).

Flask Web Development: Developing Web Applications with Python.

O'Reilly Media.

Cet ouvrage présente le framework Flask, utilisé pour le développement du backend et l'exposition de l'API de prédiction du modèle.

Mozilla Developer Network (MDN). (2024).

HTML, CSS et JavaScript Documentation.

Récupéré depuis : <https://developer.mozilla.org>

Documentation de référence pour le développement de l'interface utilisateur web de l'application.

Références liées à l'IA et à l'affective computing

Picard, R. W. (1997).

Affective Computing.

MIT Press.

Ouvrage fondateur introduisant le concept d'affective computing, domaine central pour la reconnaissance et l'analyse des émotions humaines par les systèmes intelligents.

10. Réalisation et simulation

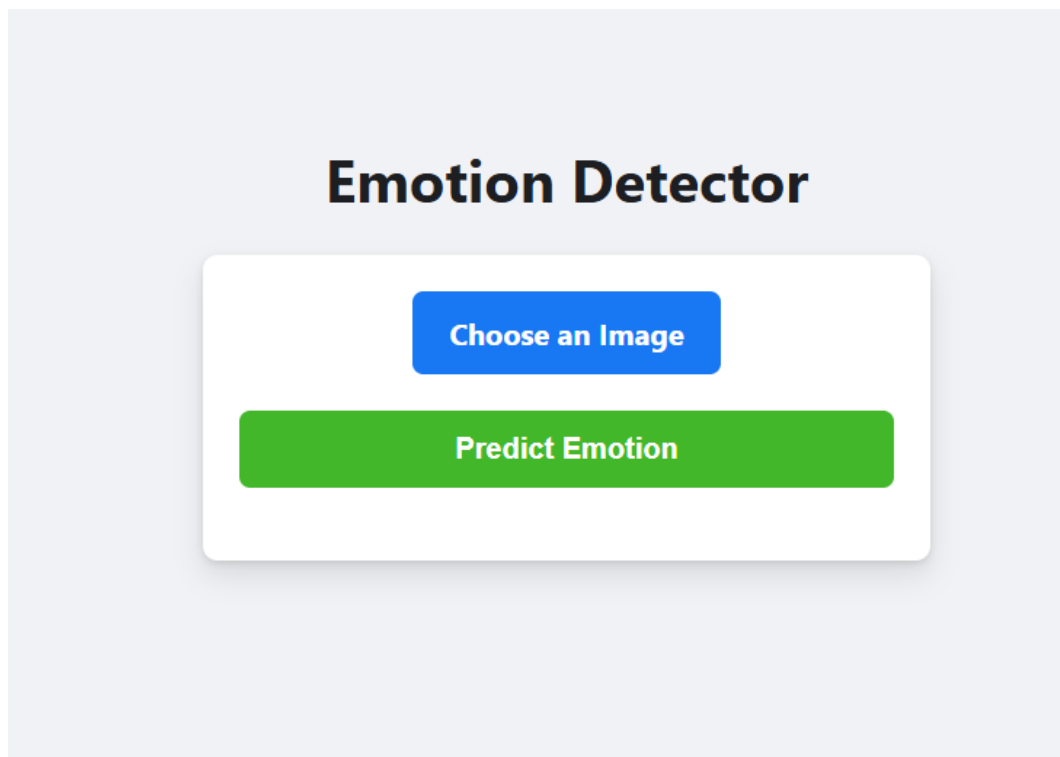


Figure 1 : Page d'accueil de l'application de reconnaissance des émotions faciales permettant le téléversement d'une image pour la détection de l'émotion.

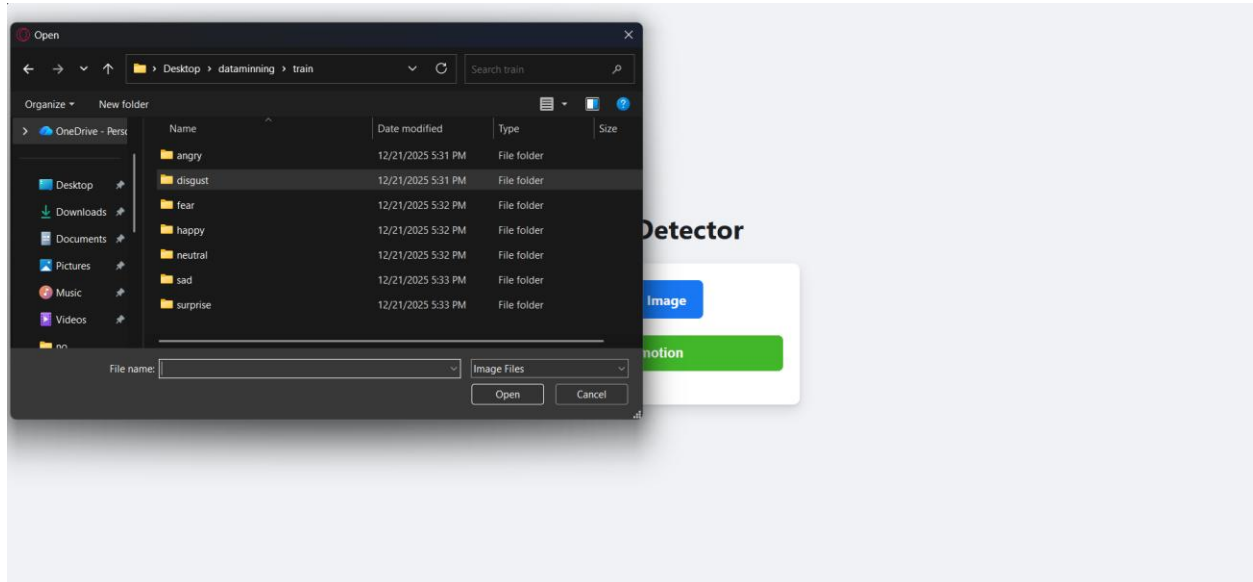


Figure 2 : Sélection des émotions à partir du jeu de données d'entraînement (FER-2013).

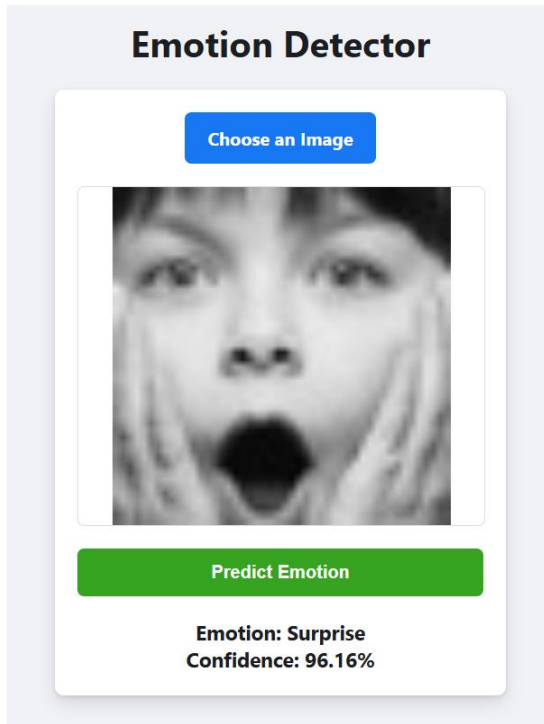


Figure 3 : Résultat de la détection de l'émotion « Surprise » à partir d'une image faciale.

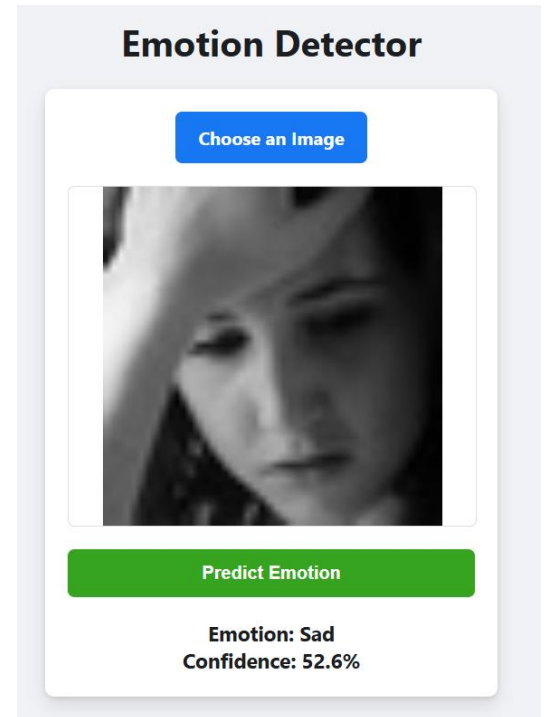


Figure 4 : Résultat de la détection de l'émotion « Sad » à partir d'une image faciale.

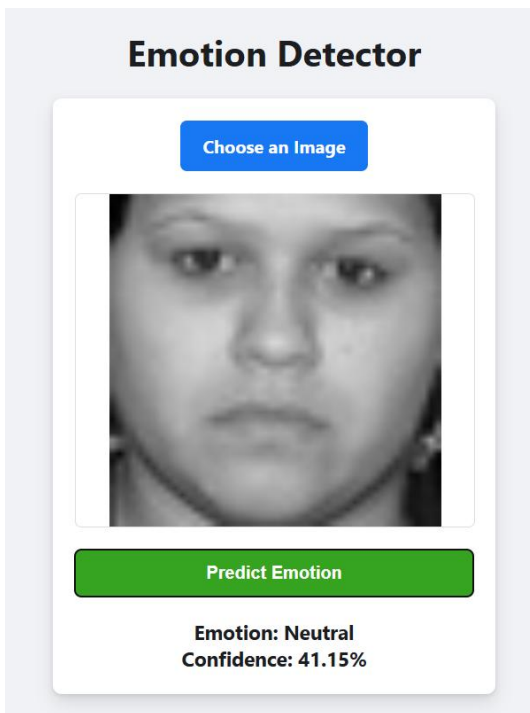


Figure 5 : Résultat de la détection de l'émotion « Neutral » à partir d'une image faciale.

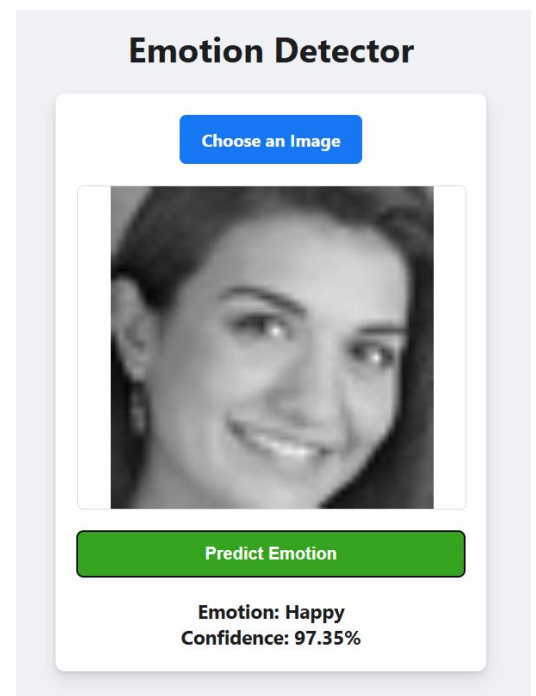


Figure 6 : Résultat de la détection de l'émotion « Happy » à partir d'une image faciale.

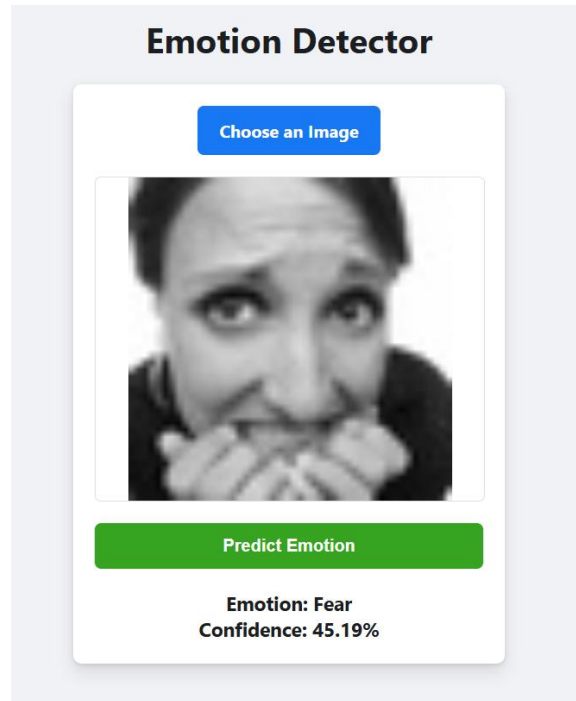


Figure 7 : Résultat de la détection de l'émotion « Fear » à partir d'une image faciale.

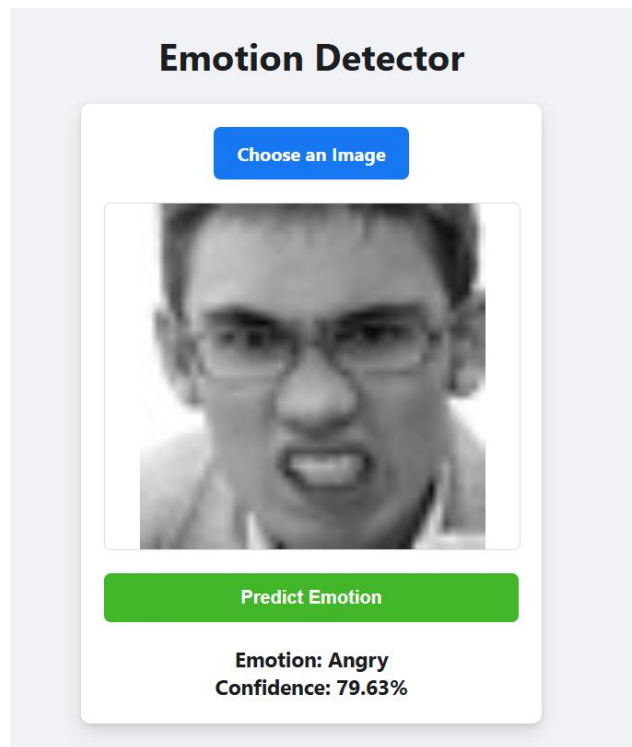


Figure 8 : Résultat de la détection de l'émotion « Angry » à partir d'une image faciale.