# DÉTECTION DES ÉMOTIONS FACIALES À L'AIDE DE RÉSEAUX NEURONAUX CONVOLUTIFS

Par

Alexandre Masse
Sciences Informatiques et Mathématiques,
Cégep de Trois-Rivières,
10 mars 2025

Une activité de réalisation soumise pour le Parcours
Scientifique du Cégep de Trois-Rivières

# COMITÉ DE SUPERVISION

Luc Morin, Superviseur
(Professeur de Mathématiques)

# RÉSUMÉ

Les émotions humaines sont l'état mental des sentiments. Il n'existe pas de lien clair entre les émotions et les expressions faciales et il y a une variabilité significative, ce qui rend la reconnaissance faciale un domaine de recherche complexe. Des caractéristiques telles que l'histogramme des gradients orientés (HOG) et la transformation de caractéristiques invariantes à l'échelle (SIFT) ont été envisagées pour la reconnaissance de motifs. Ces caractéristiques sont extraites des images selon des algorithmes prédéfinis manuellement. Ces dernières années, l'apprentissage automatique et les réseaux neuronaux ont été utilisés pour la reconnaissance des émotions. Dans ce rapport, un réseau neuronal convolutif (CNN) est utilisé pour extraire les caractéristiques des images afin de détecter les émotions.

# TABLE DES MATIÈRES

Comité de supervision....i
Résumé....ii
Table des matières.... iii
Chapitre 1: Introduction....1
Motivation....1
Objectifs....2
Enjeux éthiques et responsabilité de l'utilisation des CNN....2
Reconnaissance d'émotion faciale....3
Fonctionnement du logiciel....3
Chapitre 2 : Méthodologie....5
Choix technologiques....5
Préparation du jeu de données....5
Conversion des images en tableaux....6
Conversion des images en points de repère faciaux....6
Architecture du CNN....7
Chapitre 3 : Résultats....9
Mesures d'évaluation....9
Déterminer les meilleures valeurs de paramètres pour le modèle....10
Résultats du modèle....11
Discussion et améliorations possibles....13
Points forts et limites....13
Pistes d'amélioration....13
Conclusion....14
Bibliographie....15

# CHAPITRE 1: INTRODUCTION

Les émotions faciales jouent un rôle clé dans la communication humaine, aidant à comprendre les intentions des autres. En général, nous déduisons l'état émotionnel d'une personne—joie, tristesse, colère—grâce à ses expressions faciales et à son ton de voix. Les expressions faciales sont l'un des principaux canaux d'information dans les interactions sociales. Il est donc naturel que la recherche sur les émotions faciales ait pris de l'ampleur au cours de la dernière décennie, avec des applications en sciences perceptuelles et cognitives. L'intérêt pour la reconnaissance automatique des émotions faciales a également explosé avec l'essor des techniques d'Intelligence Artificielle. Ces technologies sont aujourd'hui largement utilisées et leur interaction avec les humains est en constante augmentation. Pour améliorer l'interaction homme-machine (*Human-Computer Interaction* ou HCI) et la rendre plus naturelle, il est essentiel que les machines puissent comprendre leur environnement, en particulier les intentions humaines. Grâce aux caméras et capteurs, elles peuvent analyser leur environnement en temps réel. Ces dernières années, les algorithmes d'Apprentissage Profond (*Deep Learning* ou DL) ont prouvé leur efficacité pour interpréter les informations captées. La détection des émotions est un élément clé permettant aux machines de mieux remplir leur rôle, car elle leur offre un aperçu de l'état interne des individus. En combinant des images faciales et des techniques de DL, une machine peut ainsi reconnaître et interpréter les émotions humaines.

## Motivation

L'Intelligence Artificielle (IA) et l'Apprentissage Automatique (*Machine Learning* ou ML) sont largement utilisés dans de nombreux domaines. En exploration de données (*data mining*), ils ont permis de détecter les fraudes dans le secteur de l'assurance. Des techniques de regroupement (*clustering*) ont été appliquées pour identifier des tendances dans les données boursières. Grâce à ces avancées, l'apprentissage automatique permet de développer des solutions de reconnaissance d'émotions faciales efficaces, peu coûteuses, fiables et avec un temps de calcul réduit. L'analyse des émotions faciales a de nombreuses applications concrètes. Dans le domaine de la santé mentale, elle permet de détecter des signes précoces de stress, d'anxiété ou de dépression, aidant ainsi les professionnels à mieux suivre l'état émotionnel des patients. En marketing et service client, elle est utilisée pour analyser les réactions des consommateurs face à une publicité ou mesurer la satisfaction en temps réel dans les interactions avec un assistant virtuel. Dans les véhicules intelligents, cette technologie peut détecter la fatigue ou la distraction des conducteurs et déclencher des alertes pour prévenir les accidents. Les jeux vidéo et la réalité virtuelle l'intègrent également pour adapter le gameplay en fonction des émotions du joueur. Finalement, dans la sécurité et la surveillance, elle peut aider à identifier des comportements suspects ou prévenir des actes violents en analysant les expressions faciales des individus dans des lieux publics.

## Objectifs

L'objectif principal de ce projet est de développer une base solide pour de futurs travaux en concevant et en implémentant un réseau de neurones convolutif (CNN) capable d'analyser les émotions faciales avec précision. Dans un premier temps, il s'agira de construire un modèle simple, permettant de comprendre les mécanismes fondamentaux des CNN et d'évaluer leurs performances sur cette tâche spécifique. Une fois cette première étape validée, l'objectif sera d'optimiser et de complexifier l'architecture du réseau afin d'améliorer sa capacité de généralisation et sa robustesse face aux variations des expressions faciales, des conditions d'éclairage et des différences individuelles. À terme, cette recherche vise à explorer des modèles avancés pouvant potentiellement surpasser les capacités du cortex visuel humain. En plus de l'amélioration des performances, ce projet ambitionne de poser les bases d'une solution adaptable à diverses applications, telles que l'interaction homme-machine, l'assistance aux personnes atteintes de troubles de la communication et l'amélioration des systèmes de surveillance émotionnelle dans des contextes professionnels ou médicaux.

## Enjeux éthiques et responsabilité de l'utilisation des CNN

Notre rencontre avec Réjean Trottier, chargé de projet en intelligence artificielle au Cégep de Trois-Rivières, nous a permis de mieux comprendre les enjeux éthiques importants liés au développement et à l'application des réseaux de neurones convolutifs pour la détection des émotions humaines. Bien que ces technologies ouvrent des perspectives importantes dans des domaines tels que la santé mentale, l'éducation ou le marketing, leur utilisation amène une réflexion rigoureuse sur leurs conséquences sociales, juridiques et morales. L'analyse des émotions humaines repose souvent sur l'acquisition et le traitement d'images contenant des données biométriques sensibles, notamment les expressions faciales. Cependant, la captation et l'exploitation de ces données peuvent se faire, dans certains cas, à l'insu des personnes concernées ou sans leur consentement. Il est impératif que les utilisateurs et concepteurs de ces systèmes respectent le cadre juridique en vigueur sur la protection des données personnelles et garantissent la transparence sur la collecte et l'usage des informations émotionnelles. Également, comme tout système d'apprentissage automatique, les CNN sont sensibles aux biais présents dans les données d'entraînement. Un jeu de données peu diversifié peut conduire à des erreurs de classification ou à des interprétations incorrectes des émotions, particulièrement pour certaines catégories de population selon le genre, l'âge ou l'origine ethnique. Ces biais peuvent induire des discriminations ou des injustices dans les décisions automatisées prises sur la base de ces analyses émotionnelles. Il est donc indispensable de concevoir des modèles justes et inclusifs. Finalement, la responsabilité de l'utilisation de ces technologies appartient autant aux concepteurs des modèles qu'aux acteurs qui les déploient. Les chercheurs et développeurs doivent anticiper les usages détournés de leurs projets et intégrer des dispositifs de contrôle et de limitation. Aussi, les organisations qui exploitent ces systèmes ont le devoir de mettre en place des politiques d'utilisation éthique et de veiller à leur conformité avec les principes de respect des droits humains.

# Reconnaissance d'émotion faciale

La reconnaissance des émotions faciales suit généralement quatre étapes. La première consiste à détecter un visage dans une image et à l'encadrer. Ensuite, il faut identifier les points de repère (*landmarks*) dans cette région faciale. La troisième étape consiste à extraire des caractéristiques spatiales et temporelles des différentes parties du visage. Enfin, un classificateur d'extraction de caractéristiques (*Feature Extraction* ou FE) utilise ces informations pour reconnaître l'émotion exprimée.
La Figure 1 illustre le processus ce processus pour une image donnée, où la région du visage et les points de repère sont détectés. Ces *landmarks* correspondent à des points visuellement significatifs comme le bout du nez, les extrémités des sourcils et de la bouche (voir Figure 2). Les relations entre deux points de repère ou encore la texture locale autour d'un *landmark* sont utilisées comme caractéristiques.

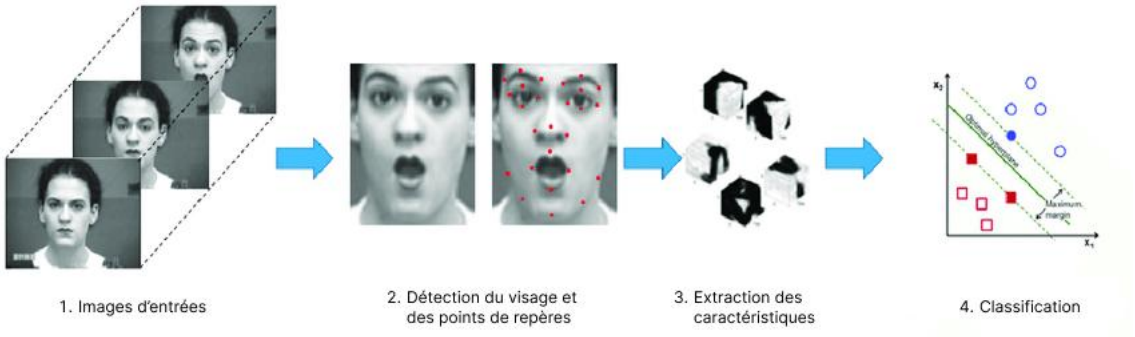


Figure 1 Processus de reconnaissance des émotions faciales pour une image.

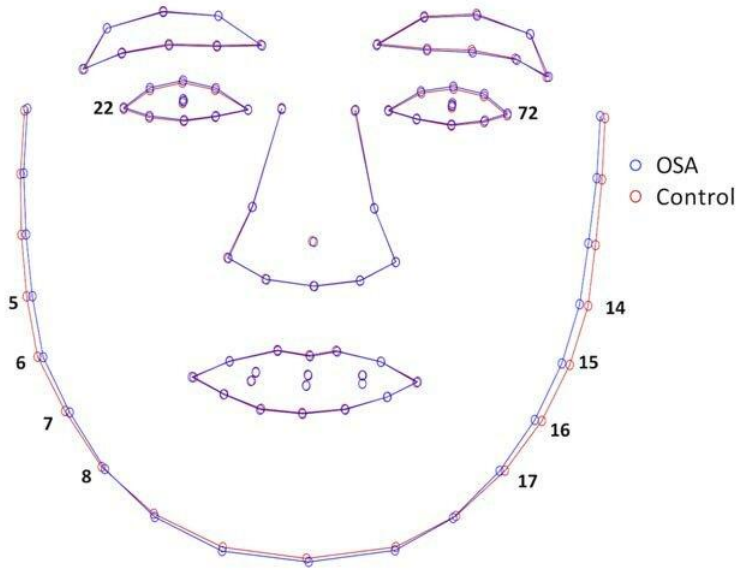


Figure 2 Points de repère faciaux à extraire d'un visage.

# Fonctionnement du logiciel

Le logiciel développé dans le cadre de ce projet est une application d'analyse d'émotions faciales basée sur un réseau de neurones convolutif (CNN). Son objectif principal est de

détecter et classifier les expressions faciales en temps réel ou à partir d'images statiques. L'application fonctionne en plusieurs étapes :

1. Acquisition des images: Les images sont capturées via une caméra en direct ou importées depuis une base de données.
2. Prétraitement des données: Les images sont redimensionnées, converties et normalisées pour garantir une meilleure efficacité du modèle.
3. Extraction des caractéristiques: Un CNN entraîné sur un jeu de données spécifique extrait les traits faciaux pertinents.
4. Classification des émotions: Le modèle attribue une catégorie émotionnelle à chaque visage détecté (ex. : joie, tristesse, colère, surprise, neutralité, etc.).
5. Affichage des résultats: L'interface utilisateur affiche l'émotion détectée en temps réel avec des indicateurs de confiance.

Le jeu de données utilisé pour ce modèle est le Face expression recognition dataset. Il s'agit d'un ensemble de données open source partagé publiquement sur Kaggle. Il comprend 28 821 images de visages en niveaux de gris de taille 48 × 48, annotées avec différentes émotions. Pour ce projet, sept émotions ont été retenues : joie, colère, neutre, tristesse, dégoût, surprise et peur.

# CHAPITRE 2 : MÉTHODOLOGIE

## Choix technologiques

J'ai choisi Python comme langage de programmation pour implémenter mon modèle, en raison de sa flexibilité, de son nombre élevé de bibliothèques dédiées à l'apprentissage automatique et de son adoption généralisée dans la recherche et l'industrie. Python est un standard dans le domaine de l'apprentissage automatique, utilisé aussi bien par les petits chercheurs que par les ingénieurs des grandes entreprises technologiques. Son écosystème inclut des frameworks puissants tels que TensorFlow, PyTorch et Keras, qui facilitent la conception, l'entraînement et l'optimisation des réseaux de neurones convolutifs. De plus, Python offre des bibliothèques comme NumPy, Pandas et OpenCV, essentielles pour le prétraitement des données, la manipulation des images et l'accélération des calculs. Voici les bibliothèques Python utilisées:

- **NumPy**: Numerical Python (NumPy) est une bibliothèque open source permettant de manipuler des tableaux et des matrices. Comme les entrées d'un réseau de neurones convolutif (CNN) sont des tableaux de nombres, NumPy est essentiel pour convertir des images en tableaux et effectuer facilement des multiplications matricielles et d'autres opérations CNN.
- **OpenCV**: OpenCV est une bibliothèque open source dédiée à la vision par ordinateur (*Computer Vision* ou CV), à l'apprentissage automatique et au traitement d'images. Elle permet d'analyser des images et des vidéos pour identifier des objets, des visages et même des écritures manuscrites. Associée à NumPy, elle peut traiter des structures de tableaux et réaliser des opérations mathématiques pour la reconnaissance de motifs.
- **Keras**: Keras est une bibliothèque open source qui facilite la construction et l'entraînement de modèles d'apprentissage profond (*deep learning* ou DL). Grâce à son API intuitive, Keras permet d'implémenter rapidement des réseaux de neuronnes convolutifs (CNN) pour des tâches comme la reconnaissance d'émotions faciales.
- **OneHot Encoder**: Partie de la bibliothèque Scikit-learn, cette technique est couramment utilisée pour convertir des catégories en un format exploitable par un modèle d'apprentissage automatique ou d'apprentissage profond.
- **Math**: Cette bibliothèque Python regroupe des fonctions mathématiques courantes, incluant des fonctions trigonométriques, logarithmiques et de conversion d'angles. Ces fonctions sont utilisées pour effectuer des calculs sur des tableaux et des matrices, indispensables pour les opérations des CNN.

Toutes ces bibliothèques sont soit pré-installées, soit installables via *pip* en Python. Elles jouent un rôle clé dans un réseau de neurones convolutif.

## Préparation du jeu de données

Avant d'être utilisées comme entrée dans un réseau de neurones convolutif, les données doivent être traitées. Le jeu de données présente certains défis qu'il faut résoudre pour améliorer les performances du modèle. Étant donné que le modèle prend des tableaux de nombres en entrée, les images doivent être converties en tableaux numériques. Voici quelques défis liés au jeu de données et comment ils sont traités:

i. **Déséquilibre des classes**: Un déséquilibre survient lorsqu'une classe d'émotions contient beaucoup plus d'images qu'une autre. Par exemple, si 2000 images représentent l'émotion "joie" et seulement 500 images représentent "peur", le modèle risque de favoriser l'émotion dominante. Pour résoudre ce problème, une augmentation de données est appliquée. Cette technique permet d'accroître artificiellement la quantité de données en utilisant des méthodes comme le recadrage, le remplissage et le retournement horizontal des images.
ii. **Variation de contraste**: Certaines images du jeu de données sont trop sombres ou trop claires. Comme les images contiennent des informations visuelles, un fort contraste améliore la détection des traits du visage. Étant donné que le CNN apprend automatiquement les caractéristiques des images pour classer les émotions, une forte variation du contraste peut nuire à ses performances. Une solution consiste à modifier les images pour centrer l'analyse uniquement sur le visage, ce qui réduit l'impact des variations de contraste.
iii. **Variation intra-classe**: Le jeu de données contient non seulement des visages humains, mais aussi des dessins et des visages animés. Cependant, les traits des visages réels et animés sont différents, ce qui peut perturber le modèle lorsqu'il extrait les points de repère faciaux. Pour améliorer la précision, seules les images de visages humains sont conservées et les autres sont supprimées.
iv. **Occlusion**: L'occlusion se produit lorsqu'une partie du visage est cachée, par exemple lorsqu'une main couvre un œil ou le nez ou lorsqu'une personne porte des lunettes de soleil ou un masque. Comme indiqué dans le tableau des points de repère faciaux, les yeux et le nez sont des caractéristiques primaires essentielles pour reconnaître les émotions. Les images contenant des occlusions sont donc supprimées, car elles risquent de fausser la classification des émotions.

## Conversion des images en tableaux

Une image est représentée par des valeurs numériques correspondant aux intensités des pixels. Pour traiter ces images, elles doivent être converties en tableaux numériques afin que le modèle puisse les exploiter. Le module NumPy permet de transformer une image en tableau et d'extraire ses attributs. Ainsi, une image de la classe "triste" du jeu de données peut être convertie en un tableau NumPy. Nos image sont constituées de 2 304 pixels (car 48 × 48 = 2 304), elles possèdent 2 dimensions (hauteur et largeur) et leur résolution est de 48 × 48 pixels.

## Conversion des images en points de repère faciaux

La détection des points de repère du visage se déroule en deux étapes, la localisation du visage dans l'image et la détection des points de repère faciaux. Un détecteur de visage frontal est utilisé pour détecter le visage dans une image. Une boîte rectangulaire est tracée autour du visage, définie par les coordonnées du coin supérieur gauche et du coin inférieur droit. Ensuite, un prédicteur de formes est appliqué pour extraire les caractéristiques faciales clés. Un objet appelé *landmarks* est utilisé, prenant deux arguments: l'image contenant le visage et la zone où les points de repère faciaux doivent être extraits (définie par les coordonnées du rectangle). La Figure 3 illustre les 64 points de repère détectés sur un visage.

Ces points incluent les yeux, sourcils, nez, bouche et contour du visage et sont essentiels pour l'analyse des expressions faciales.

Figure 3 Points de repère détectés sur un visage.

## Architecture du CNN

Les modèles d'apprentissage automatique peuvent être construits et entraînés grâce à un API de haut niveau comme Keras. Dans ce projet, un modèle de réseau de neurones convolutif (CNN) séquentiel est développé en utilisant TensorFlow avec l'API Keras. Cette approche est choisie, car Keras permet de construire un modèle couche par couche, facilitant ainsi l'expérimentation et l'optimisation. TensorFlow est une plateforme open source pour l'apprentissage automatique. Elle offre une large collection d'outils, de bibliothèques et de ressources communautaires pour concevoir et déployer des applications d'apprentissage automatique. L'architecture de mon CNN suit une structure classique de couches convolutives suivies de couches de pooling et de couches entièrement connectées pour l'apprentissage des caractéristiques et la classification. Le réseau commence par une première couche de convolution appliquant 32 filtres (*kernels*) de taille 3×3 avec un *stride* de 1, suivie d'une activation *ReLU* et d'une couche de *MaxPooling* pour réduire la dimension spatiale tout en conservant les informations essentielles. Cette séquence se répète avec un nombre croissant de filtres: 64, 96 et 128 filtres, chacun accompagné d'une activation *ReLU* et d'une opération de *MaxPooling*. L'augmentation progressive du nombre de filtres permet d'extraire des caractéristiques de plus en plus complexes de l'image en entrée. Après l'extraction des caractéristiques, la partie dense du réseau entre en jeu. Une première couche entièrement connectée (*Fully Connected* ou FC) avec 200 neurones cachés transforme les caractéristiques extraites en une représentation plus compacte et abstraite. Finalement, une dernière couche entièrement connectée avec 7 neurones applique une fonction d'activation *Softmax* pour produire une probabilité d'appartenance à chacune des 7 classes. Cette architecture suit les bonnes pratiques des CNN modernes: l'utilisation de petits filtres (3×3), de *ReLU* pour éviter le problème de gradient, de *MaxPooling* pour réduire la dimension et de couches entièrement connectées en fin de réseau pour la classification. Elle est bien adaptée à des tâches de classification d'images où une montée en complexité progressive des filtres est nécessaire pour capturer des motifs de plus en plus abstraits.

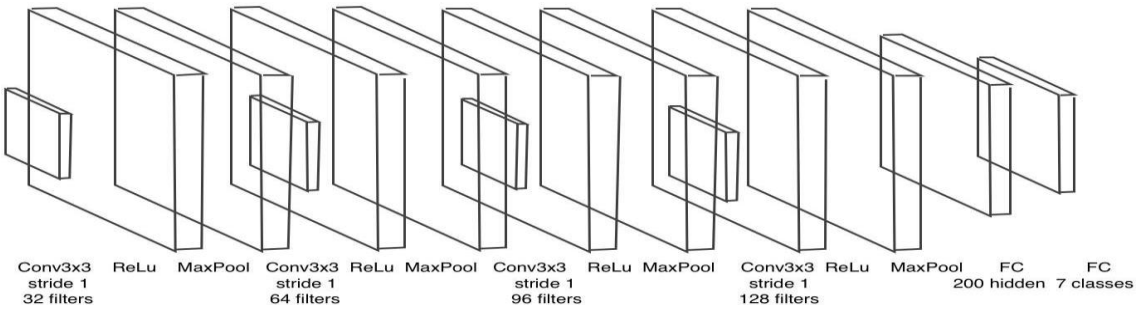


Figure 4 Architecture du CNN.

Les résultats du réseau convolutif sont par la suite affichés sous forme de probabilités associées à chaque émotion. Après avoir traité une image, le modèle retourne un vecteur où chaque valeur représente la probabilité que le visage exprime une émotion spécifique (joie, tristesse, colère, surprise, etc.). L'émotion ayant la probabilité la plus élevée est considérée comme la plus probable.

# CHAPITRE 3 : RÉSULTATS

Dans ce chapitre, les mesures utilisées pour évaluer la performance du modèle sont définies. Ensuite, les meilleures valeurs de paramètres pour chaque modèle sont déterminées à partir des résultats d'entraînement. Ces valeurs sont ensuite utilisées pour évaluer la précision (*accuracy*) et la perte (*loss*) du modèle. Finalement, les résultats obtenus pour ce CNN sont analysés.

## Mesures d'évaluation

La précision (*accuracy*), la perte (*loss*), la précision spécifique (*precision*), le rappel (*recall*) et le F-score sont les mesures utilisées pour évaluer la performance du modèle. Ces métriques sont définies ci-dessous.

**Précision**: La précision est donnée par

$$\text{Précision} = \frac{\text{Nombre de prédictions correctes}}{Nombre\ total\ de\ prédictions}$$

**Perte**: L'entropie croisée catégorielle (*categorical cross-entropy*) est utilisée comme fonction de perte et est donnée par:

$$Loss = -\sum_{i=1}^{m} y_i \log(p_i)$$

où *y* est un indicateur binaire (0 ou 1), *p* est la probabilité prédite et *m* est le nombre de classes (heureux, triste, peur, colère, etc.)

**Matrice de confusion**: La matrice de confusion fournit les valeurs des quatre combinaisons possibles entre les prédictions du modèle et les valeurs réelles: Vrai Positif (VP), Vrai Négatif (VN), Faux Positif (FP) et Faux Négatif (FN). Les mesures de précision, rappel et score F sont calculées à partir de ces valeurs.

- VP (Vrai Positif): l'émotion est correctement prédite.
- FP (Faux Positif): une émotion incorrecte est prédite.
- VN (Vrai Négatif): une émotion incorrecte est correctement écartée.
- FN (Faux Négatif): une émotion correcte est incorrectement écartée.

La Figure 5 illustre une matrice de confusion.

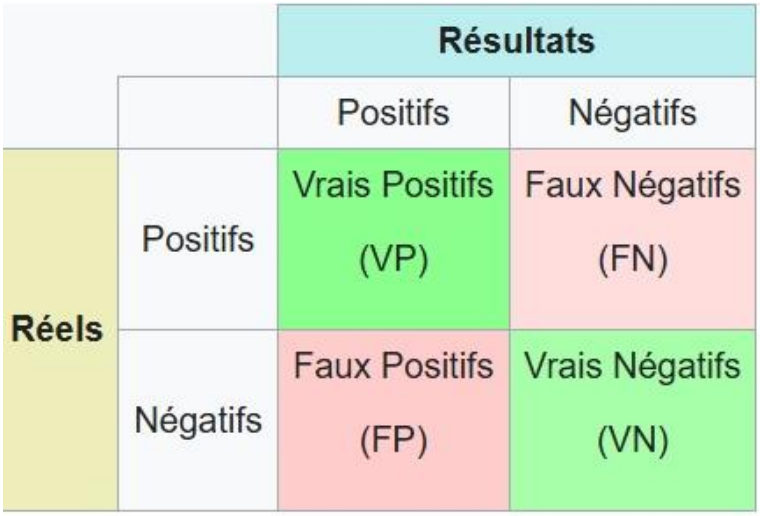


Figure 5 Une matrice de confusion.

**Rappel**: Le rappel est donné par

$$\text{Rappel} = \frac{VP}{VP + FN}$$

**Précision**: La précision est donnée par

$$\text{Précision} = \frac{VP}{VP + FP}$$

**F-score:** Le F-score est la moyenne harmonique du rappel (*recall*) et de la précision (*precision*). Il est calculé à l'aide de la formule

$$F\text{-score} = \frac{2 \times \text{Précision} \times \text{Rappel}}{\text{Précision} + \text{Rappel}}$$

## Déterminer les meilleures valeurs de paramètres pour le modèle

Les paramètres comme le taux d'apprentissage (*learning rate* ou LR), la taille des lots (*batch size*), le nombre d'époques (*epochs*) et la méthode de prétraitement des images. Le tableau 1 présente les valeurs de paramètres considérées pour ces modèles. Ces valeurs ont été choisies car elles sont couramment utilisées dans le milieu. Les trois méthodes de prétraitement d'images étudiées sont indiquées dans le tableau 2.

| Paramètre | Valeurs |
|---|---|
| Taux d'apprentissage | 0.1, 0.01, 0.001 |
| Taille des lots | 16, 32 |
| Méthode de pré-traitement des images | Méthode 1, Méthode 2, Méthode 3 |

Tableau 1 Les valeurs des paramètres considérées.

| Paramètre | Méthode 1 | Méthode 2 | Méthode 3 |
|---|---|---|---|
| Rotation (0-180°) | 10 | 8 | 5 |
| Décalage en largeur (fraction de la largeur de l'image) | 0.10 | 0.08 | 0.05 |
| Décalage en hauteur (fraction de la hauteur de l'image) | 0.10 | 0.08 | 0.05 |
| Zoom (0-1) | 0.10 | 0.08 | 0.05 |
| Retournement horizontal | Vrai | Vrai | Vrai |

Tableau 2 Les méthodes de pré-traitement des images considérées.

Un premier modèle, entraîné avec un taux d'apprentissage de 0.01, un *batch size* de 16 et l'optimiseur Adam, a atteint une précision maximale de 0.72. L'ajout d'un arrêt anticipé (*early stopping*) a permis d'éviter une augmentation de la perte sans gain de précision. Parmi les configurations testées, celle avec un taux d'apprentissage de 0.01, un *batch size* de 16, l'optimiseur Adam et la méthode 1 de pré-traitement des images a obtenu la meilleure précision (0.72) avec une perte moyenne de 0.90. Le deuxième modèle, entraîné en évitant les valeurs sous-performantes du premier modèle, a été testé avec différentes configurations. Celle avec un taux d'apprentissage de 0.001, un *batch size* de 16, l'optimiseur Nadam et la méthode 1 de pré-traitement des images a atteint une précision similaire (0.72) mais avec une perte plus faible (0.88). Ainsi, bien que les deux modèles offrent une précision comparable, le deuxième semble légèrement plus performant grâce à une meilleure minimisation de la perte.

## Résultats du modèle

Afin de comparer la performance des deux modèles, j'ai tracé l'historique de la fonction de perte pour ces deux architectures. Le résultat est illustré à la figure 6. Comme on peut l'observer, le modèle 2 a permis d'améliorer la précision de validation de 18,46%. J'ai également calculé la matrice de confusion pour du modèle le plus performant. La figure 7 illustre les résultats. Un point intéressant est que le modèle parvient bien à identifier l'émotion "heureux", ce qui suggère que les traits distinctifs d'un visage souriant sont plus facilement appréhendés par le réseau que ceux d'autres expressions. Cette matrice révèle également les classes les plus fréquemment confondues. Par exemple, l'étiquette "colère" est souvent classifiée comme "peur" ou "tristesse". En examinant les échantillons du jeu de données, on constate que cette confusion est cohérente: même pour un observateur humain, il n'est pas toujours évident de différencier une expression de colère d'une expression de tristesse. Cela souligne la variabilité interindividuelle dans la manifestation des émotions. En complément de cette matrice de confusion, j'ai évalué la précision de classification par classe pour chaque architecture. Le tableau 3 présente ces résultats. Comme on peut l'observer, la reconnaissance de l'émotion "heureux" affiche le score le plus élevé dans les deux modèles.

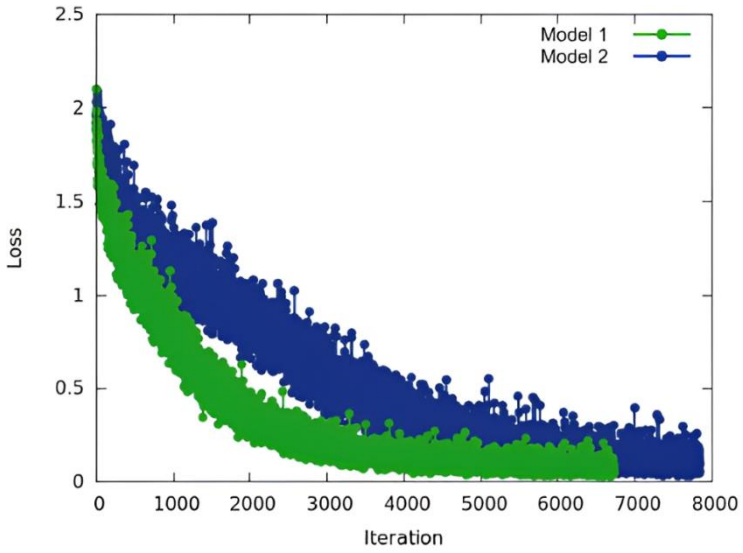


Figure 6 L'historique des fonctions de pertes des deux modèles.

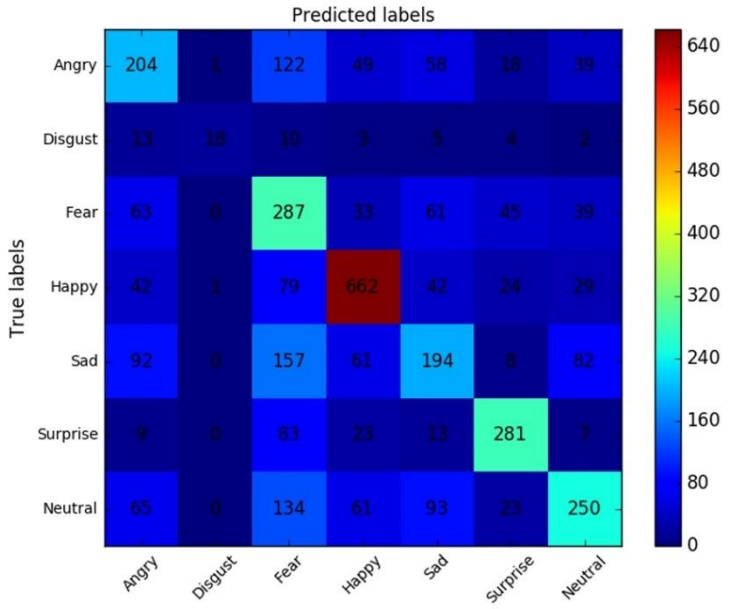


Figure 7 La matrice de confusion du modèle 2.

| Expression | Modèle 1 | Modèle 2 |
|---|---|---|
| Colère | 41% | 53% |
| Dégoût | 32% | 70% |
| Peur | 54% | 46% |
| Joie | 75% | 80.5% |
| Tristesse | 32% | 63% |
| Surprise | 67.5% | 62.5% |
| Neutre | 39.9% | 51.5% |

Tableau 3 La précision de la prédiction pour chaque expression dans les modèles 1 et 2.

# DISCUSSION ET AMÉLIORATIONS POSSIBLES

L’analyse des résultats obtenus met en évidence plusieurs aspects positifs de mon approche, mais aussi certaines limitations qui mériteraient d’être adressées pour améliorer les performances du modèle.

## Points forts et limites

L’un des principaux points forts du modèle est sa capacité à bien classifier certaines émotions, en particulier celles présentant des traits distinctifs marqués, comme l’émotion "heureux". Cette performance suggère que les caractéristiques extraites par les couches convolutives sont bien adaptées pour distinguer des expressions faciales claires. Cependant, des limitations persistent, notamment dans la classification des émotions plus subtiles ou proches les unes des autres, comme "colère", "tristesse" et "peur". Les erreurs de classification observées dans la matrice de confusion montrent que ces émotions sont souvent confondues, ce qui peut s’expliquer par une variabilité interindividuelle importante dans leur expression. Une autre difficulté est la qualité et la diversité des données d’entraînement: certaines émotions sont sous-représentées dans la base de données, ce qui entraîne un déséquilibre affectant la précision du modèle.

## Pistes d’amélioration

1. **Meilleur jeu de données**: Une solution prometteuse serait d’enrichir le jeu de données en intégrant un ensemble de données plus équilibré et diversifié. Actuellement, la distribution inégale des classes crée un biais qui favorise les émotions les plus représentées. L’intégration d’un *dataset* plus vaste, contenant une répartition homogène des émotions et des variations dans l’expression faciale, pourrait améliorer la robustesse du modèle. L’utilisation de techniques d’augmentation des données, comme la génération d’images synthétiques à l’aide de modèles génératifs adverses (GAN), permettrait également d’augmenter la quantité et la diversité des échantillons.
2. **Architecture plus performante**: Un autre axe d’amélioration concerne l’architecture du réseau de neurones. Des modèles plus avancés, tels que les réseaux à attention (Transformers appliqués aux images) ou des variantes optimisées des CNNs comme EfficientNet, pourraient permettre d’extraire des caractéristiques plus fines et mieux adaptées à la reconnaissance des expressions faciales.
3. **Ajout d’un filtre contre les biais**: Finalement, la mise en place d’un mécanisme de réduction des biais serait une amélioration essentielle. Comme mentionné précédemment, les données d’entraînement actuelles présentent des déséquilibres qui peuvent induire des biais dans les prédictions du modèle. Des méthodes comme *l’adversarial debiasing* (apprentissage d’un modèle en minimisant les corrélations avec des variables sensibles) pourraient être envisagées. Une autre approche complémentaire consisterait à évaluer systématiquement la performance du modèle sur différents sous-groupes démographiques afin de détecter et corriger d’éventuels biais.

# CONCLUSION

Dans ce travail de réalisation, j'ai développé un réseau de neurones convolutif (CNN) pour résoudre un problème de reconnaissance des expressions faciales et j'ai évalué sa performance en utilisant différentes techniques de post-traitement et de visualisation. Les résultats ont démontré que les CNN sont capables d'apprendre les caractéristiques faciales et d'améliorer la détection des émotions. Cela suggère que ces réseaux peuvent, à eux seuls, apprendre intrinsèquement les caractéristiques faciales clés en se basant uniquement sur les données brutes des pixels. Cette expérience m'a permis d'approfondir mes connaissances en apprentissage profond, notamment sur la conception et l'optimisation de modèles CNN. J'ai appris à manipuler des jeux de données complexes, à ajuster les hyperparamètres pour améliorer la précision et à utiliser des techniques de visualisation pour mieux interpréter les résultats du modèle. De plus, j'ai pris conscience de l'importance du choix des architectures et de la qualité des données d'entraînement pour obtenir des performances optimales.
Les applications potentielles de ce projet sont nombreuses. En psychologie, ce type de modèle pourrait être utilisé pour analyser les expressions faciales et mieux comprendre les émotions humaines. Il pourrait également être intégré à des assistants virtuels ou des robots sociaux afin de rendre leurs interactions plus naturelles et adaptées aux émotions des utilisateurs. Finalement, ce modèle pourrait contribuer à améliorer l'accessibilité, en aidant les personnes atteintes de troubles de la communication à mieux interagir avec leur environnement. Si je devais poursuivre ce projet, plusieurs améliorations pourraient être envisagées. Premièrement, il serait intéressant d'expérimenter d'autres architectures de réseaux neuronaux, comme les réseaux récurrents ou les *transformers*, afin d'évaluer leur impact sur la précision du modèle. Aussi, une optimisation des hyperparamètres plus poussée pourrait permettre d'augmenter la robustesse du modèle. Ainsi, ce projet m'a permis d'explorer en profondeur le potentiel des CNN pour la reconnaissance des émotions et ouvre la porte à de nombreuses perspectives d'amélioration et d'application dans divers domaines.

# BIBLIOGRAPHIE


1. https://www.kaggle.com/datasets/jonathanoheix/face-expression-recognition-dataset/data
2. https://github.com/keras-team/keras
3. https://scikit-learn.org/stable/
4. Bettadapura, Vinay (2012). Face expression recognition and analysis: the state of the art. arXiv preprint arXiv:1203.6722
5. G.E. Hinton et al., Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups, IEEE Signal Processing Magazine, vol. 29, no. 6, pp. 82-97, (2012).
6. Nicu Sebe, Michael S. Lew, Ira Cohen, Yafei Sun, Theo Gevers, Thomas S. Huang (2007) Authen tic Facial Expression Analysis. Image and Vision Computing 25.12: 1856-1863
7. Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. ”Imagenet classification with deep con volutional neural networks.” Advances in neural information processing systems. 2012.
8. Simonyan, Karen, and Andrew Zisserman. ”Very deep convolutional networks for large-scale im age recognition.” arXiv preprint arXiv:1409.1556 (2014).
9. A. Sinha, R.P. Aneesh, Real time facial emotion recognition using deep learning, International Journal of Innovations & Implementations in Engineering, vol. 1, pp. 1-5, (2019).
10. A. Mollahosseini, D. Chan, M.H. Mahoor, Going deeper in facial expression recognition using deep neural networks, IEEE Winter Conference on Applications of Computer Vision, (2016).
11. S. Shah, A comprehensive guide to convolutional neural networks, available online: https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the eli5-way-3bd2b1164a53, (2018).
12. A. Amini, A. Soleimany, MIT deep learning open access course 6.S191, available online: http://introtodeeplearning.com/ (2023).
13. M.O. Parkhi, A. Vedaldi, A. Zisserman, Deep face recognition, British Machine Vision Conference, pp. 41.1-41.12, (2015).
14. D.K. Jain, Z. Zhang, K. Huang, Multi angle optimal pattern based deep learning for automatic facial expression recognition, Pattern Recognition Letters, vol. 139, pp. 157-165, (2020).