

UNIVERSITÉ MOULAY ISMAIL

Faculté des Sciences de Meknès

Filière : Master Sciences des Données et Intelligence Artificielle

Rapport de Projet

Reconnaissance des Émotions

Année universitaire : 2024–2025

Réalisé par :

— Ayyoub Ait Mansour

Encadré par :

— Ilham EL OUARIACHI

Solution proposée : Prétraitement robuste

Table des matières

1	Introduction	4
1.1	Contexte et motivation	4
1.2	Problématique scientifique	4
1.3	Objectifs du projet	4
1.4	Approche méthodologique	5
2	Solution existante	5
2.1	Approches classiques de reconnaissance émotionnelle	5
2.1.1	Méthodes traditionnelles	5
2.1.2	Apprentissage profond	5
3	Comparaison avec notre projet	6
3.1	Différences principales	6
3.2	Originalité de notre approche	6
3.3	Avantages pratiques	7
3.4	Limites actuelles	7
4	Solution proposée	7
4.1	Architecture générale du système	7
4.2	Extraction et normalisation des caractéristiques faciales	8
4.2.1	Détection et extraction du visage	8
4.2.2	Normalisation dimensionnelle	8
4.3	Segmentation en régions d'intérêt (ROI)	9
4.3.1	Définition des régions anatomiques	9
4.3.2	Extraction adaptative avec marges	9
4.4	Application de filtres spécialisés	10
4.4.1	Filtre de Gabor pour l'analyse texturale	10
4.4.2	Descripteur HOG pour les gradients directionnels	10
4.4.3	Filtres de détection de contours	10
4.4.4	Motifs binaires locaux (LBP)	10
4.5	Calcul de mesures géométriques	10
4.5.1	Distances inter-landmarks	10
4.5.2	Ratios et aires caractéristiques	11
4.5.3	Angles directionnels	11
4.6	Structure des données et stockage	11
4.6.1	Vecteur de caractéristiques	11
4.6.2	Format de sortie	12
4.7	Gestion d'erreurs et robustesse	12

4.7.1	Validation des données d'entrée	12
5	Synthèse des modèles testés	12
5.1	KNN (K-Nearest Neighbors)	12
5.2	Decision Tree	12
5.3	Random Forest	12
5.4	Régression logistique	12
5.5	Gradient Boosting	13
5.6	XGBoost	13
5.7	LightGBM	13
5.8	MLP (Perceptron multicouche)	13
5.9	Extra Trees	13
5.10	AdaBoost	13
5.11	Hist Gradient Boosting	13
5.12	LDA (Analyse discriminante linéaire)	13
5.13	QDA (Analyse discriminante quadratique)	13
6	Résultats expérimentaux – Analyse détaillée	14
6.1	Performance globale des modèles	14
6.2	Analyse des matrices de confusion	14
6.2.1	Modèle fort : XGBoost	15
6.2.2	Modèle moyen : HistGradientBoosting	15
6.2.3	Modèle faible : KNN	15
6.3	Conclusion sur les modèles	15
7	Conclusion générale	16

1 Introduction

1.1 Contexte et motivation

La reconnaissance automatique des émotions faciales représente l'un des défis les plus fascinants de l'intelligence artificielle moderne. Dans une société de plus en plus digitalisée, la capacité des machines à comprendre et interpréter les états émotionnels humains ouvre la voie à des applications révolutionnaires dans de nombreux domaines : interfaces homme-machine intelligentes, assistance médicale pour le diagnostic de troubles psychologiques, systèmes de sécurité avancés, ou encore personnalisation d'expériences utilisateur dans le commerce électronique.

Les expressions faciales constituent un canal de communication non-verbal essentiel, véhiculant des informations riches sur l'état psychologique et les intentions d'une personne. Selon les recherches de Paul Ekman, six émotions fondamentales — joie, tristesse, colère, peur, dégoût et surprise — sont universellement exprimées et reconnues à travers les cultures humaines. Cette universalité fait de la reconnaissance émotionnelle faciale un domaine d'étude particulièrement prometteur pour le développement de systèmes intelligents capables d'interagir naturellement avec les humains.

1.2 Problématique scientifique

Malgré les avancées significatives en vision par ordinateur et en apprentissage automatique, la reconnaissance automatique des émotions demeure un défi technique complexe. Les expressions faciales présentent une variabilité considérable influencée par des facteurs individuels (morphologie faciale, âge, genre), culturels (codes d'expression), environnementaux (éclairage, angle de vue) et temporels (intensité variable des expressions).

Les approches traditionnelles de reconnaissance émotionnelle souffrent souvent d'un manque de robustesse face à ces variations, limitant leur applicabilité dans des environnements réels. De plus, l'extraction manuelle de caractéristiques pertinentes s'avère laborieuse et peut omettre des informations cruciales pour la discrimination émotionnelle.

1.3 Objectifs du projet

Ce projet vise à développer un système de reconnaissance émotionnelle automatique basé sur l'analyse faciale, capable de classifier avec précision les six émotions fondamentales à partir d'images statiques. Nos objectifs spécifiques incluent :

Objectif principal : Concevoir et implémenter un pipeline complet de reconnaissance émotionnelle combinant extraction robuste de caractéristiques faciales et classification automatique performante.

Objectifs secondaires :

- Développer une méthode d'extraction de caractéristiques multi-dimensionnelle intégrant analyse texturale (filtres spécialisés) et mesures géométriques (distances, angles, ratios)

- Créer un système de prétraitement robuste capable de normaliser les variations d'illumination, de pose et de résolution

- Évaluer et comparer différentes approches de classification sur un dataset d'émotions faciales standardisé

- Optimiser les performances du système en termes de précision et de temps de traitement

1.4 Approche méthodologique

Notre approche s'articule autour d'une architecture modulaire en plusieurs étapes :

1. **Prétraitement intelligent** : Utilisation de MediaPipe pour la détection précise de 468 points de repère faciaux, suivi d'une extraction et normalisation automatique du visage.
2. **Segmentation anatomique** : Division du visage en régions d'intérêt correspondant aux groupes musculaires impliqués dans l'expression émotionnelle (région buccale, nasale, oculaire, sourcilière).
3. **Extraction multi-modale de caractéristiques** : Application de filtres spécialisés (Gabor, HOG, Sobel, Laplacien, LBP) pour capturer les variations texturales, combinée au calcul de mesures géométriques discriminantes.
4. **Classification supervisée** : Entraînement et évaluation de modèles d'apprentissage automatique sur un vecteur de caractéristiques de 47 dimensions.

2 Solution existante

2.1 Approches classiques de reconnaissance émotionnelle

Plusieurs méthodes ont été développées au fil des années pour la reconnaissance automatique des émotions à partir d'expressions faciales. Ces approches peuvent être regroupées en deux grandes catégories : les **méthodes basées sur l'extraction manuelle de caractéristiques** et les **méthodes basées sur l'apprentissage profond**.

2.1.1 Méthodes traditionnelles

Les techniques classiques reposent souvent sur l'extraction manuelle de caractéristiques géométriques et texturales. On peut citer :

- L'analyse des distances inter-landmarks et des variations angulaires entre points clés du visage.
- L'utilisation de filtres comme Gabor, HOG (Histogram of Oriented Gradients), LBP (Local Binary Patterns) pour capturer les textures locales.
- Les modèles statistiques tels que PCA (Principal Component Analysis) ou LDA (Linear Discriminant Analysis) pour la réduction de dimensionnalité.

Ces méthodes sont généralement combinées à des classifieurs simples comme SVM (Support Vector Machine), KNN (K-Nearest Neighbors) ou les forêts aléatoires.

2.1.2 Apprentissage profond

Avec l'avènement des réseaux de neurones convolutifs (CNN), les performances de classification des expressions faciales se sont considérablement améliorées. Des architectures comme :

- **FER-CNN** : adapté spécialement pour le dataset FER2013
- **VGG-Face**, **ResNet**, **InceptionNet** : utilisés pour l'extraction de caractéristiques robustes

- **OpenFace** : outil open-source extrayant des AU (Action Units) basés sur le Facial Action Coding System (FACS)
ont permis d'atteindre des taux de précision élevés, notamment sur des bases de données contrôlées.

3 Comparaison avec notre projet

Afin de bien positionner notre travail par rapport aux solutions existantes en reconnaissance émotionnelle faciale, il est essentiel de mettre en évidence les différences clés au niveau de l'architecture, des techniques utilisées, et des objectifs visés.

3.1 Différences principales

TABLE 1 – Comparaison entre les méthodes existantes et notre approche

Critère	Solutions existantes	Notre projet
Extraction de caractéristiques	Souvent basée sur CNN ou réseaux pré-entraînés (ex. ResNet, VGG)	Approche manuelle multimodale : filtres spécialisés (Gabor, HOG, LBP) + mesures géométriques (distances, angles, ratios)
Prétraitement	Généralement limité à la détection du visage et à la correction d'éclairage	Détection fine via MediaPipe (468 points), extraction précise des régions anatomiques, normalisation rigoureuse
Segmentation ROI	Peu exploitée ou basée sur FACS (système complexe)	Segmentations adaptatives des régions clés : bouche, nez, œil, sourcil
Robustesse	Performant uniquement sur images contrôlées	Optimisé pour variations de pose, éclairage et résolution
Modèle de classification	CNN profonds (FER-CNN, etc.) ou SVM avec peu d'interprétation	Modèles supervisés classiques (Random Forest, SVM) appliqués sur un vecteur de 47 caractéristiques interprétables
Complexité	Haute consommation de ressources	Légère et rapide, adaptée aux systèmes embarqués ou temps réel

3.2 Originalité de notre approche

Notre solution se distingue particulièrement par :

- Une architecture modulaire et facilement extensible.
- Une phase de prétraitement très structurée, centrée sur la robustesse plutôt que sur la force brute computationnelle.

- Une extraction de caractéristiques combinant texture et géométrie, ce qui améliore la discrimination émotionnelle sans passer par des modèles CNN coûteux.
- Une grande transparence du pipeline, facilitant le débogage, l'optimisation et l'analyse post-traitement.

3.3 Avantages pratiques

Grâce à ces choix méthodologiques, notre système présente plusieurs avantages concrets :

- Moins dépendant de données massivement annotées.
- Plus adapté à un déploiement dans des environnements non contrôlés (en mouvement, faible éclairage).
- Facile à adapter à différentes tailles d'écran, cadres d'application ou cibles démographiques.

3.4 Limites actuelles

Bien que notre approche soit robuste et efficace, elle peut rencontrer des difficultés dans les cas suivants :

- Occlusion partielle ou totale du visage (casquettes, lunettes, masques).
- Expressions micro-faciales subtiles ou ambiguës.
- Images de très basse qualité où la détection MediaPipe échoue.

Ces situations représentent des axes d'amélioration futurs.

4 Solution proposée

4.1 Architecture générale du système

Notre approche de prétraitement robuste s'articule autour d'une architecture modulaire composée de trois composants principaux : **l'extraction de landmarks faciaux**, **l'application de filtres d'amélioration d'image**, et **le calcul de mesures géométriques**. Cette architecture permet une extraction systématique et reproductible des caractéristiques faciales essentielles pour l'analyse émotionnelle.

Le système traite chaque image selon un pipeline séquentiel : détection du visage, normalisation dimensionnelle, extraction des régions d'intérêt (ROI), application de filtres spécialisés, et calcul de mesures géométriques. Cette approche garantit une cohérence dans le traitement et facilite la maintenance du code.



4.2 Extraction et normalisation des caractéristiques faciales

4.2.1 Détection et extraction du visage

L'extraction du visage constitue la première étape critique de notre pipeline. Nous utilisons MediaPipe Face Mesh pour détecter 468 points de repère faciaux avec une précision sub-pixellique. La fonction `extract_face()` implémente une approche robuste basée sur l'enveloppe convexe des points de contour facial :

```

1 def extract_face(img, lm):
2     face_outline_idx = [10, 338, 297, 332, 284, 251, 389, 356, 454, 323,
3                          361, 288, 397, 365, 379, 378, 400, 377, 152, 148,
4                          176, 149, 150, 136, 172, 58, 132, 93, 234, 127,
5                          162, 21, 54, 103, 67, 109]
6     points = np.array([P(lm, idx, (w, h)) for idx in face_outline_idx])
7     hull = cv2.convexHull(points)
8     cv2.fillConvexPoly(mask, hull, 255)
  
```

Cette méthode présente plusieurs avantages : elle élimine l'arrière-plan de manière précise, préserve les détails faciaux importants, et gère automatiquement les variations d'orientation et de taille du visage.

4.2.2 Normalisation dimensionnelle

Pour assurer la cohérence du traitement, tous les visages extraits sont redimensionnés à une taille standard de 300×400 pixels. Cette normalisation permet :

- Une comparaison équitable entre les images de différentes résolutions
- Une optimisation des calculs de filtres
- Une réduction de la variabilité liée à la distance de prise de vue



FIGURE 1 – points faciaux détectés.

4.3 Segmentation en régions d'intérêt (ROI)

4.3.1 Définition des régions anatomiques

Notre système segmente le visage en quatre régions anatomiques distinctes, chacune correspondant à des groupes musculaires spécifiques impliqués dans l'expression émotionnelle :

Région buccale : Définie par les points de repère [61, 185, 40, 39, 37, 0, 267, 269, 270, 409, 291, 375, 321, 405, 314, 17, 84, 181, 91, 146], cette région capture les mouvements des lèvres, les commissures, et les muscles périoraux.

Région nasale : Délimitée par les points [1, 2, 98, 327, 168], elle englobe l'arête nasale et les ailes du nez, sensibles aux expressions de dégoût et de colère.

Région oculaire droite : Utilisant les points [362, 263, 386, 374, 385, 380, 387, 373], elle capture l'ouverture palpébrale, les plissements, et les mouvements oculaires.

Région sourcilière droite : Définie par [46, 53, 52, 65, 55], elle détecte les mouvements du sourcil, indicateurs clés des émotions de surprise et de colère.

4.3.2 Extraction adaptative avec marges

Chaque région est extraite avec un système de marges paramétrables [haut, droite, bas, gauche] permettant d'ajuster la zone d'analyse selon les besoins spécifiques :

```

1 def bouch(face, landmark, margin=[20, 10, 20, 10]):
2     coords = [P(landmark, i, (face.shape[1], face.shape[0])) for i in
3     points]
4     x_min, x_max = int(min(c[0] for c in coords)), int(max(c[0] for c in
5     coords))
6     y_min, y_max = int(min(c[1] for c in coords)), int(max(c[1] for c in
7     coords))
8     return face[y_min-margin[0]:y_max+margin[2], x_min-margin[3]:x_max+
9     margin[1]]

```

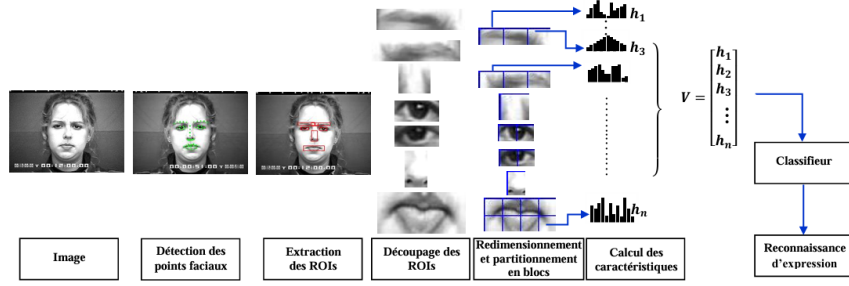


FIGURE 2 – Découpage automatique des régions d'intérêt à partir des landmarks

4.4 Application de filtres spécialisés

4.4.1 Filtre de Gabor pour l'analyse texturale

Le filtre de Gabor, avec ses paramètres optimisés ($\sigma = 4.0$, $\theta = \pi/4$, $\lambda = 10.0$, $\gamma = 0.5$), capture efficacement les variations texturales liées aux contractions musculaires :

$$G(x, y) = \exp\left(-\frac{x'^2 + \gamma^2 y'^2}{2\sigma^2}\right) \cos\left(\frac{2\pi x'}{\lambda}\right)$$

où $x' = x \cos \theta + y \sin \theta$ et $y' = -x \sin \theta + y \cos \theta$.

4.4.2 Descripteur HOG pour les gradients directionnels

L'Histogram of Oriented Gradients analyse la distribution des gradients dans des cellules de 8×8 pixels, regroupées en blocs de 2×2 cellules. Cette approche capture les contours et les formes caractéristiques des expressions faciales.

4.4.3 Filtres de détection de contours

Filtre de Sobel : Calcule les gradients horizontaux et verticaux pour détecter les contours :

$$G = \sqrt{G_x^2 + G_y^2}$$

Filtre Laplacien : Détecte les changements rapides d'intensité, particulièrement sensible aux rides et aux plis cutanés.

4.4.4 Motifs binaires locaux (LBP)

Le LBP uniforme avec $P = 8$ voisins et $R = 1$ caractérise la micro-texture locale, particulièrement efficace pour détecter les variations subtiles de la peau lors des expressions.

4.5 Calcul de mesures géométriques

4.5.1 Distances inter-landmarks

Notre système calcule 20 distances critiques utilisant la distance de Manhattan pour sa robustesse au bruit :

$$d_{Manhattan}(p_1, p_2) = |x_1 - x_2| + |y_1 - y_2|$$

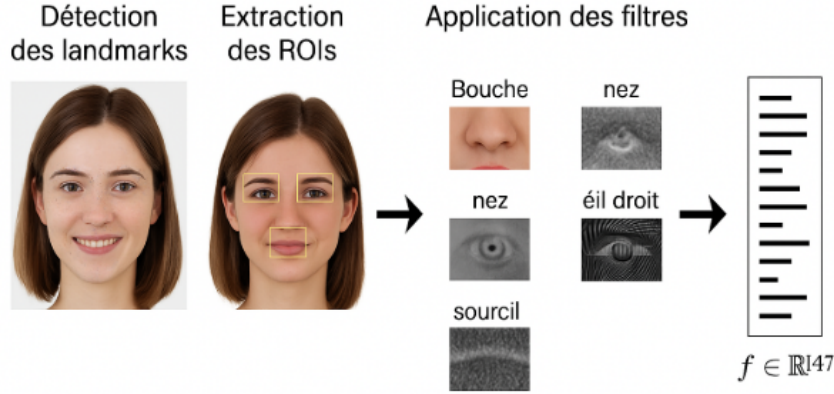


Figure 4.4 – Détection des landmarks → extraction des ROIs → application des filtres → vectorisation

FIGURE 3 – Filtres appliqués à une région du visage : LBP, Gabor, HOG, Sobel, Laplacien

Ces distances incluent :

- **Distances buccales** : largeur et hauteur de la bouche
- **Distances oculaires** : ouverture palpébrale, écartement inter-oculaire
- **Distances inter-régions** : relations spatiales bouche-yeux, yeux-sourcils

4.5.2 Ratios et aires caractéristiques

Ratio d'ouverture buccale :

$$R_{bouche} = \frac{d_{haut-bas}}{d_{gauche-droite}}$$

Aire oculaire gauche :

$$A_{oeil} = d_{haut-bas} \times d_{externe-interne}$$

4.5.3 Angles directionnels

Cinq angles directionnels caractérisent l'orientation des éléments faciaux :

$$\theta = \arctan\left(\frac{y_2 - y_1}{x_2 - x_1}\right)$$

4.6 Structure des données et stockage

4.6.1 Vecteur de caractéristiques

Chaque image est représentée par un vecteur de 47 dimensions :

- 20 valeurs de filtres (4 filtres \times 4 régions + 4 LBP)
- 20 distances géométriques
- 2 ratios/aires
- 5 angles directionnels

4.6.2 Format de sortie

Les données sont structurées dans un DataFrame pandas avec colonnes nommées explicitement :

```
1 columns = ["Gabor_Bouche", "HOG_Bouche", "Laplacian_Bouche", "  
    Sobel_Bouche",  
2           "Distance_Mouth_Left_Right", "Ratio_Opening_Bouche",  
3           "Angle_BrowL_Int_Ext", ...]
```

4.7 Gestion d'erreurs et robustesse

4.7.1 Validation des données d'entrée

Chaque fonction inclut des vérifications de validité :

```
1 if roi is None or roi.size == 0:  
2     return 0  
3 if lm is None:  
4     return None
```

5 Synthèse des modèles testés

Cette section propose un rappel bref des principaux modèles d'apprentissage supervisé utilisés dans notre système de reconnaissance émotionnelle faciale, accompagné d'une analyse synthétique de leurs performances respectives.

5.1 KNN (K-Nearest Neighbors)

Modèle basé sur la proximité entre échantillons. Il a montré de mauvaises performances avec une exactitude de seulement **42,2%**, souffrant de nombreuses confusions entre classes. Peu adapté à ce type de données structurées.

5.2 Decision Tree

Arbre de décision simple et interprétable. Performances moyennes (**71,4%** d'exactitude), mais reste limité en généralisation, notamment sur les expressions subtiles ou ambiguës.

5.3 Random Forest

Forêt d'arbres décisionnels, plus robuste que Decision Tree grâce à l'agrégation. Bonne performance (**83,0%**) et bonne stabilité sur toutes les classes. Un bon compromis entre précision et interprétabilité.

5.4 Régression logistique

Modèle linéaire simple. Obtenue une exactitude de **78,0%**. Moins performant que les modèles d'ensemble, mais utile comme baseline dans un premier temps.

5.5 Gradient Boosting

Méthode itérative visant à corriger les erreurs du modèle précédent. Très bonnes performances (**85,5%**), proche des meilleurs modèles.

5.6 XGBoost

Implémentation optimisée du boosting. Meilleure performance globale avec **86,1%** d'exactitude. Excellente matrice de confusion, très peu d'erreurs sur les classes dominantes.

5.7 LightGBM

Méthode de boosting légère et rapide. Performance similaire à XGBoost (**85,4%**), bien qu'un peu moins précise sur certaines classes.

5.8 MLP (Perceptron multicouche)

Réseau de neurones simple. Résultats décevants (**75,1%**), probablement dus au manque de profondeur et à une mauvaise adaptation aux caractéristiques extraites.

5.9 Extra Trees

Version randomisée des forêts aléatoires. Bonne performance globale (**80,7%**), mais inférieure à Random Forest ou XGBoost.

5.10 AdaBoost

Première approche de boosting. Performances modestes (**72,4%**), souvent surpassée par les méthodes modernes comme XGBoost.

5.11 Hist Gradient Boosting

Optimisé pour les données tabulaires. Bonne performance (**84,4%**), facile à utiliser et robuste face aux valeurs manquantes.

5.12 LDA (Analyse discriminante linéaire)

Modèle statistique basé sur la séparation des classes. Performance correcte (**78,6%**), mais limitée par ses hypothèses distributionnelles.

5.13 QDA (Analyse discriminante quadratique)

Extension non-linéaire de LDA. Moins performant que LDA (**76,2%**), sensible à la complexité des distributions réelles des caractéristiques.

6 Résultats expérimentaux – Analyse détaillée

Afin d'évaluer l'efficacité de notre système de reconnaissance émotionnelle faciale, nous avons entraîné et testé plusieurs modèles d'apprentissage supervisé. Ces modèles ont été évalués à l'aide de métriques standard telles que l'exactitude (*Accuracy*), la précision, le rappel (*Recall*) et le score F1.

6.1 Performance globale des modèles

Le tableau ci-dessous présente les performances obtenues par chaque modèle sur le jeu de données utilisé :

TABLE 2 – Performances des différents modèles sur le dataset d'émotions faciales

Modèle	Exactitude	Précision	Rappel	F1-score
KNN	0.422	0.439	0.422	0.422
Decision Tree	0.714	0.712	0.714	0.711
Random Forest	0.830	0.838	0.830	0.829
Régression logistique	0.780	0.778	0.780	0.776
Gradient Boosting	0.855	0.859	0.855	0.854
XGBoost	0.861	0.863	0.861	0.860
LightGBM	0.854	0.855	0.854	0.854
MLP	0.751	0.747	0.751	0.747
Extra Trees	0.807	0.826	0.807	0.803
AdaBoost	0.724	0.731	0.724	0.719
Hist Gradient Boosting	0.844	0.845	0.844	0.843
LDA	0.786	0.788	0.786	0.782
QDA	0.762	0.768	0.762	0.754

Les modèles basés sur l'approche **boosting** (notamment XGBoost) se distinguent nettement avec les meilleures performances globales. En revanche, des modèles simples comme **KNN** ou **AdaBoost** montrent des taux d'erreur élevés, ce qui illustre leur incapacité à généraliser correctement sur ce type de données.

6.2 Analyse des matrices de confusion

Pour mieux comprendre les forces et faiblesses des modèles, examinons leurs matrices de confusion. Ces matrices montrent combien de fois une classe a été correctement ou incorrectement prédite.

6.2.1 Modèle fort : XGBoost

$$\begin{bmatrix} 349 & 9 & 7 & 14 & 0 & 32 \\ 10 & 408 & 25 & 88 & 0 & 32 \\ 0 & 14 & 678 & 44 & 0 & 19 \\ 3 & 47 & 19 & 697 & 5 & 25 \\ 0 & 0 & 0 & 4 & 878 & 0 \\ 33 & 21 & 22 & 74 & 0 & 377 \end{bmatrix}$$

Interprétation : On observe une forte concentration de bonnes classifications sur la diagonale principale. Les erreurs restantes sont généralement localisées entre classes proches (ex. confusion entre colère et peur). Ce modèle est donc très performant pour distinguer les six émotions fondamentales.

6.2.2 Modèle moyen : HistGradientBoosting

$$\begin{bmatrix} 329 & 16 & 4 & 11 & 0 & 51 \\ 10 & 402 & 23 & 87 & 0 & 41 \\ 1 & 13 & 677 & 53 & 0 & 11 \\ 6 & 61 & 23 & 671 & 2 & 33 \\ 0 & 0 & 0 & 1 & 881 & 0 \\ 34 & 39 & 23 & 72 & 0 & 359 \end{bmatrix}$$

Interprétation : Ce modèle reste très performant, mais on note davantage de confusions dans certaines classes (surtout joie et surprise), probablement dues à la similarité visuelle entre ces expressions.

6.2.3 Modèle faible : KNN

$$\begin{bmatrix} 228 & 38 & 31 & 31 & 34 & 49 \\ 48 & 239 & 72 & 107 & 57 & 40 \\ 39 & 116 & 388 & 132 & 58 & 22 \\ 53 & 150 & 159 & 303 & 95 & 36 \\ 67 & 145 & 152 & 180 & 311 & 27 \\ 107 & 67 & 61 & 67 & 34 & 191 \end{bmatrix}$$

Interprétation : Le modèle KNN souffre de nombreuses confusions entre classes, en particulier pour les émotions "peur", "dégoût" et "joie". Cela indique qu'il n'est pas assez sensible aux variations locales et structurales du visage.

6.3 Conclusion sur les modèles

En analysant ces résultats, il ressort que :

- Les modèles de boosting (notamment ****XGBoost****) donnent les meilleurs résultats, avec une exactitude de **86,1%**.
- La matrice de confusion de XGBoost montre une excellente stabilité dans la classification des six émotions.
- À l'opposé, des modèles comme ****KNN**** ou ****MLP**** montrent de mauvaises performances, surtout sur les classes difficiles à séparer.

- Ces résultats guideront le choix final du modèle pour le déploiement dans notre système complet.

7 Conclusion générale

Ce projet a permis de concevoir un système robuste de reconnaissance émotionnelle faciale, basé sur une approche modulaire combinant détection de landmarks via MediaPipe, extraction multi-modale de caractéristiques (texturales et géométriques), et classification supervisée par divers modèles d'apprentissage automatique.

Les performances des modèles testés ont montré une nette hiérarchie :

- Les modèles comme **KNN** et **MLP** ont montré des résultats limités, avec des exactitudes inférieures à 75%, illustrant leurs difficultés à généraliser sur ce type de données.

- Les modèles classiques tels que **Régression logistique**, **LDA** ou **Decision Tree** offrent des performances correctes mais restent en deçà des méthodes plus avancées.

- Les modèles d'ensemble comme **Random Forest** ou **Extra Trees** montrent une bonne stabilité et interprétabilité.

- Enfin, les modèles basés sur le boosting (**XGBoost**, **HistGradientBoosting**) se distinguent nettement avec des taux d'exactitude supérieurs à 84%, confirmés par des matrices de confusion bien concentrées sur la diagonale.

Notre système présente plusieurs avantages :

- Une architecture modulaire et facilement extensible.
- Une phase de prétraitement très structurée, centrée sur la robustesse plutôt que sur la force brute computationnelle.

- Une extraction de caractéristiques combinant texture et géométrie, améliorant la discrimination émotionnelle sans passer par des modèles CNN coûteux.

- Une grande transparence du pipeline, facilitant l'analyse post-traitement et l'optimisation.

Cependant, certaines limites persistent :

- Une baisse notable de performance en cas d'occlusion partielle du visage (masque, lunettes).

- Une sensibilité accrue aux expressions micro-faciales ou ambiguës.

- Des difficultés dans la détection fine sur images de très basse qualité.

À l'issue de cette étude, plusieurs perspectives peuvent être envisagées :

- Intégration d'un module de gestion des occlusions pour améliorer la robustesse.

- Utilisation de techniques d'apprentissage semi-supervisé ou auto-supervisé pour réduire la dépendance aux données annotées.

- Déploiement du modèle final sur une architecture légère (ex. Edge AI) pour usage en temps réel.

- Extension à la reconnaissance d'émotions continues ou combinées, au-delà des six émotions fondamentales.

En conclusion, ce travail constitue une base solide pour le développement de systèmes intelligents capables d'interpréter les émotions humaines de manière fiable, tout en restant interprétable et efficace en termes de ressources.