

Master 2 : MALIA

Projet Recent advances in ML

Présenté par :

SOME André

Chargé du cours : Carlos Fernando Crispim Junior

Le 17 mars 2025

TABLE DES MATIÈRES

1	Introduction	2
2	Dataset	2
3	Méthodologie	3
3.1	Prétraitement des données	3
3.2	Architecture du modèle	3
3.3	Entraînement et hyperparamètres	4
3.4	Évaluation du Modèle	4
4	Expériences et résultats	5
4.1	Résultats quantitatifs	5
4.2	Analyse de la Matrice de Confusion	6
4.3	Résultats qualitatifs (Visualisation Grad-CAM)	7
4.4	Visualisation des résultats	7
5	Conclusion	8

1 Introduction

Le cancer de la peau, et en particulier le mélanome, représente une menace sérieuse pour la santé mondiale en raison de sa capacité à évoluer rapidement et à devenir mortel s'il n'est pas détecté à temps. La détection précoce est donc cruciale pour améliorer les chances de survie des patients. Cependant, le diagnostic manuel par les dermatologues peut être sujet à des erreurs dues à la variabilité des lésions cutanées et à la charge de travail importante. Dans ce contexte, l'intelligence artificielle (IA), notamment les réseaux de neurones convolutionnels (CNN) et les techniques de vision par ordinateur offrent une solution prometteuse pour assister les professionnels de santé en automatisant et en améliorant la précision du diagnostic.

L'objectif de ce projet est de :

- Développer un modèle de deep learning capable de classifier des images médicales selon plusieurs catégories (par exemple : lésions bénignes, malignes et indéterminées).
- Exploiter une architecture CNN pré-entraînée via le transfer learning afin de bénéficier de connaissances issues d'un vaste jeu de données.
- Optimiser le processus d'entraînement par des stratégies de régularisation, d'augmentation de données et de fine-tuning pour améliorer la robustesse et la précision du modèle.
- Fournir une analyse quantitative et qualitative permettant d'identifier les points forts et les limites de l'approche proposée.

2 Dataset

Pour ce projet, nous avons utilisé le dataset ISIC (International Skin Imaging Collaboration), une ressource largement reconnue dans le domaine de la dermatologie. Ce dataset provient d'un challenge international et contient des images annotées de lésions cutanées, accompagnées de métadonnées détaillées. Il est composé d'un grand nombre d'images (par exemple, plusieurs milliers, selon la version exacte utilisée), réparties en trois classes principales : bénin, malin et indéterminé. La répartition exacte des classes reflète généralement la prévalence naturelle de ces types de lésions.

Le dataset est accessible publiquement via le site officiel de l'ISIC à l'adresse suivante : <https://challenge.isic-archive.com/data/>. Pour ce projet, nous nous sommes appuyés sur la version 2017 du dataset, qui fournit des images dermoscopiques de haute qualité ainsi que des annotations validées par des experts. Les données d'entraînement, de validation et de test ont

déjà été constituées et sont directement disponibles sur le site, ce qui garantit une séparation prédéfinie et cohérente des ensembles. Ces sous-ensembles sont organisés comme suit :

Pour préparer les données à l'entraînement du modèle, le dataset a été divisé en trois sous-ensembles :

- **Entraînement** : utilisé pour ajuster les poids du modèle.
- **Validation** : permet de surveiller les performances pendant l'entraînement et d'ajuster les hyperparamètres.
- **Test** : évalue la performance finale du modèle sur des données jamais vues.

Les annotations des images sont fournies dans des fichiers CSV, qui associent chaque image à son label correspondant (bénin, malin ou indéterminé). Ces informations ont été exploitées pour garantir une correspondance précise entre les images et leurs catégories lors de l'entraînement et de l'évaluation.

3 Méthodologie

La méthodologie de ce projet comprend plusieurs étapes techniques, allant du prétraitement des données à l'entraînement et à l'évaluation du modèle.

3.1 Prétraitement des données

Avant d'entraîner le modèle, les images ont été prétraitées pour garantir leur compatibilité avec le réseau de neurones. Les étapes incluent :

- **Normalisation** : les valeurs des pixels ont été ramenées à une échelle entre 0 et 1 pour faciliter la convergence du modèle.
- **Augmentation des données** : des transformations comme la rotation, le décalage et le retournement horizontal ont été appliquées pour augmenter la taille de l'ensemble d'entraînement et réduire le risque de surapprentissage.

3.2 Architecture du modèle

Nous avons choisi d'utiliser **ResNet-50**, un réseau de neurones convolutionnels pré-entraîné sur le dataset ImageNet, en raison de sa capacité à extraire des caractéristiques complexes des images. Pour adapter ce modèle à notre tâche :

- Les couches de base de ResNet-50 ont été gelées (non entraînées) pour conserver les connaissances génériques apprises sur ImageNet.
- Des couches personnalisées ont été ajoutées à la fin du réseau pour effectuer la classification en trois classes (bénin, malin, indéterminé). Cette approche, appelée transfer learning, est particulièrement adaptée aux projets avec des datasets médicaux limités, car elle exploite les connaissances préexistantes tout en les ajustant à notre problème spécifique.

3.3 Entraînement et hyperparamètres

Le modèle a été entraîné avec les configurations suivantes :

- **Optimiseur** : Adam, qui adapte le taux d'apprentissage pour accélérer la convergence.
- **Fonction de perte** : categorical crossentropy, adaptée à une classification multi-classes.
- **Batch size** : 32 images par itération
- **Nombre d'époques** : 50, avec une stratégie d'early stopping pour arrêter l'entraînement si les performances sur l'ensemble de validation cessent de s'améliorer. Les performances ont été évaluées à l'aide de métriques telles que la précision globale, la précision par classe et le rappel, afin de mesurer la capacité du modèle à identifier correctement chaque type de lésion.

3.4 Évaluation du Modèle

L'évaluation inclut des résultats quantitatifs (précision, rappel, AUC-ROC) et qualitatifs (visualisations Grad-CAM) sur les données de test.

Métriques d'évaluations

Les métriques suivantes ont été calculées pour évaluer les performances du modèle :

- **Précision (Precision)** : Proportion des prédictions positives correctes parmi toutes les prédictions positives pour une classe donnée.
- **Rappel (Recall)** : Proportion des vrais positifs correctement identifiés parmi tous les échantillons réels d'une classe.
- **F1-score** : Moyenne harmonique de la précision et du rappel, reflétant l'équilibre entre ces deux métriques.

- **Support** : Nombre d'échantillons réels par classe dans l'ensemble de validation.
- **Accuracy** : Proportion globale des prédictions correctes.
- **AUC-ROC** : Mesure de la capacité du modèle à discriminer entre les classes.

Grad-CAM

La technique de visualisation Grad-CAM (Gradient-weighted Class Activation Mapping) sera utilisée pour identifier les zones d'une image qui ont le plus contribué à la prédiction d'un modèle d'apprentissage profond. Elle produit une carte de chaleur (heatmap) superposée à l'image originale, utilisant une échelle de couleurs allant du bleu (faible influence) au rouge (forte influence).

4 Expériences et résultats

4.1 Résultats quantitatifs

Le tableau 1 récapitule les principales métriques de performance obtenues sur l'ensemble de validation pour chaque classe.

Pour la classe **Benign**, une précision de 0,88 indique que 88% des images prédites comme bénignes le sont réellement. Le rappel de 0,83 signifie que 83% des images réellement bénignes ont été correctement identifiées. Le F1-score de 0,85 atteste d'une performance robuste et équilibrée pour cette classe, qui constitue la majorité des échantillons (481 sur 600).

En revanche, la classe **Indeterminate** présente des résultats nuls (précision, rappel et F1-score à 0,00), ce qui s'explique par sa très faible représentation (seulement 2 individus), limitant ainsi la capacité du modèle à apprendre et à identifier cette catégorie.

Pour la classe **Malignant**, une précision de 0,43 signifie que seulement 43% des prédictions pour cette catégorie sont correctes, tandis qu'un rappel de 0,55 indique que 55% des images réellement malignes sont détectées. Le F1-score de 0,48 reflète une performance modérée, marquée par un déséquilibre entre précision et rappel.

Analyse globale des métriques :

- **Accuracy (0,80)** : Une exactitude globale de 80% est obtenue, mais cette mesure est biaisée par la prédominance de la classe *Benign*, qui représente 79% des données.
- **Macro moyenne (0,44, 0,46, 0,45)** : Les moyennes non pondérées indiquent une faible performance globale, soulignant les difficultés du modèle à bien classer les catégories minoritaires.

- **Moyenne pondérée (0,79, 0,77, 0,78)** : Ces valeurs, proches de l’accuracy, confirment l’influence dominante de la classe majoritaire.
- **AUC-ROC (0,63)** : La valeur de 0,63 démontre capacité modérée globale du modèle à discriminer entre les classes, suggérant une amélioration dans la séparation des catégories par rapport aux analyses précédentes.

TABLE 1 – Résultats de classification

Classe	Précision	Rappel	F1-score	Support
Benign	0.88	0.83	0.85	481
Indeterminate	0.00	0.00	0.00	2
Malignant	0.43	0.55	0.48	117
Accuracy	-	-	0.77	600
Macro avg	0.44	0.46	0.45	600
Weighted avg	0.79	0.77	0.78	600
AUC-ROC :	0.6275			

4.2 Analyse de la Matrice de Confusion

La matrice de confusion (Figure 1) illustre la répartition des prédictions du modèle par rapport aux classes réelles. Sur 481 échantillons de la classe *Benign*, 398 sont correctement prédits et 83 sont confondus avec la classe *Malignant*, soit un rappel de 0,81. Pour la classe *Indeterminate*, aucun des 2 échantillons n’est correctement identifié (1 est prédit comme *Benign* et 1 comme *Malignant*), ce qui donne un rappel de 0,00. Enfin, sur 117 échantillons de la classe *Malignant*, 64 sont correctement classés et 53 sont confondus avec la classe *Benign*, aboutissant à un rappel de 0,56. Ces résultats montrent que le modèle reconnaît efficacement la classe *Benign*, mais rencontre des difficultés pour la classe *Indeterminate* et affiche une performance modérée pour la classe *Malignant*.

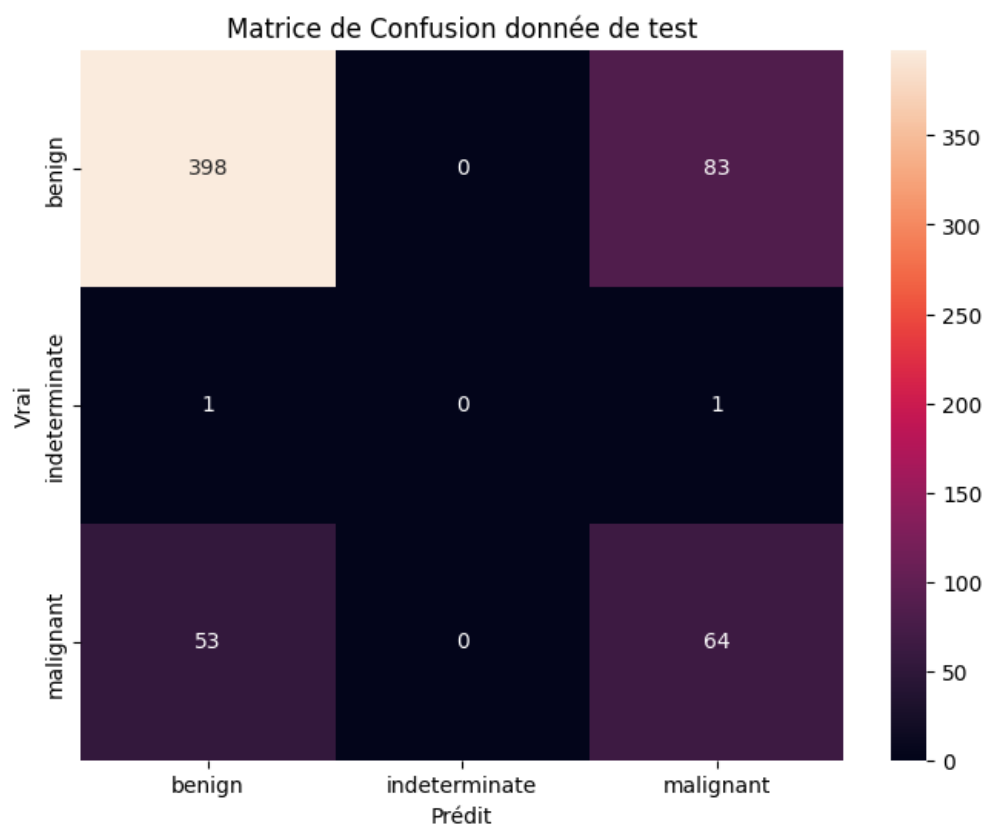


FIGURE 1 – Matrice de confusion pour les données de test

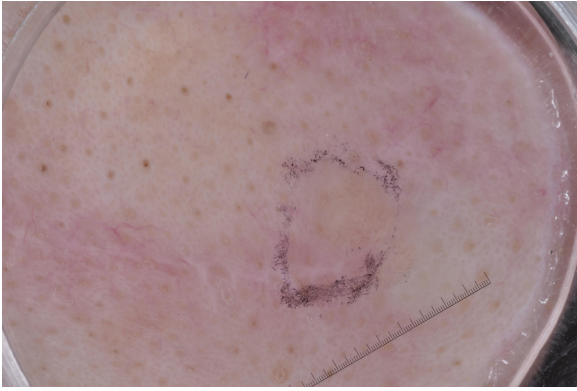
4.3 Résultats qualitatifs (Visualisation Grad-CAM)

La visualisation Grad-CAM montre que la prédiction "maligne" est principalement influencée par une région centrale-gauche de l'image, marquée en rouge, qui contient probablement une anomalie significative comme une lésion ou des cellules irrégulières. Les zones orange et jaune environnantes apportent un soutien contextuel, tandis que les régions bleues périphériques sont ignorées par le modèle. La ligne diagonale de points blancs suggère que des caractéristiques linéaires sont également prises en compte. Cette interprétation permet de confirmer que le modèle se focalise sur des zones potentiellement pertinentes.

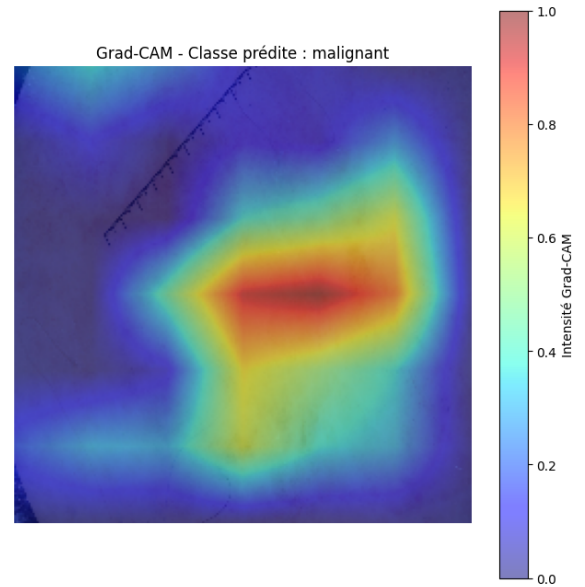
4.4 Visualisation des résultats

On observe une progression de la précision aussi bien sur l'ensemble d'entraînement que sur l'ensemble de validation, signe que le modèle apprend effectivement à classer les images. Les fluctuations sont relativement normales en cours d'apprentissage, et l'écart entre les deux courbes (entraînement vs. validation) reste modéré, ce qui indique un surapprentissage limité.

La perte diminue de manière régulière pour les deux ensembles, montrant une convergence



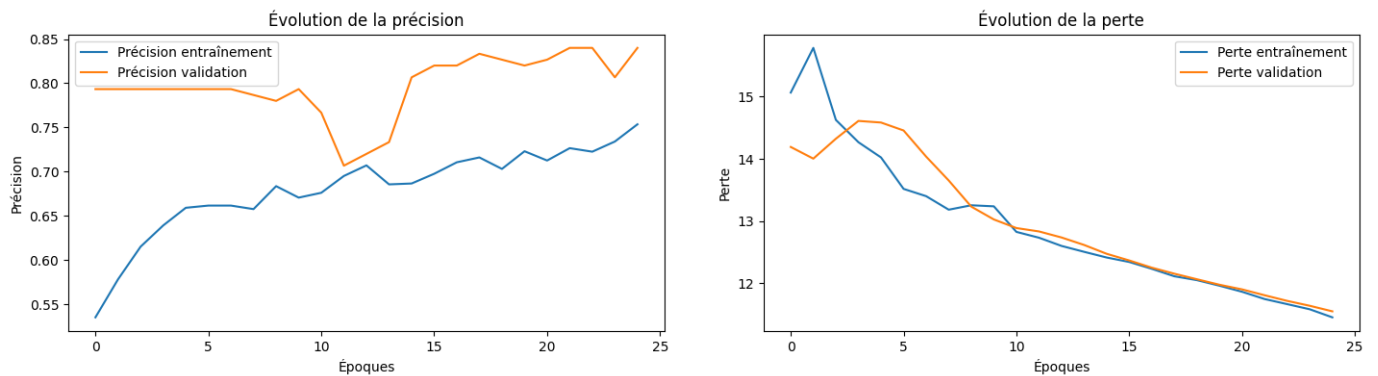
(a) Image d'une lésion cancéreuse



(b) Prédiction du modèle sur l'image

FIGURE 2 – Comparaison entre l'image originale et la prédiction du modèle

progressive. Le fait que la courbe de validation reste proche de celle de l'entraînement suggère que le modèle généralise correctement, sans dérive importante vers l'overfitting.



5 Conclusion

En conclusion, ce projet a permis d'explorer l'application du deep learning pour la classification des lésions cutanées dans le but d'améliorer la détection précoce du mélanome. L'utilisation d'un modèle **ResNet-50** pré-entraîné, adapté par transfert d'apprentissage, a offert des résultats prometteurs pour la catégorie majoritaire *benign*. Cependant, les performances restent limitées pour les classes *malignant* et *indeterminate*, principalement en raison du déséquilibre des données. L'analyse des métriques et la visualisation Grad-CAM ont permis de mettre en lumière les zones clés influençant les prédictions du modèle, renforçant ainsi la compréhension

de son fonctionnement. Ces observations soulignent à la fois le potentiel de l'intelligence artificielle dans le domaine médical et les défis persistants liés à l'optimisation de la classification des cas moins représentés. Des pistes d'amélioration, telles que l'enrichissement du dataset, l'ajustement des hyperparamètres et l'utilisation de techniques avancées d'augmentation de données, semblent prometteuses pour renforcer la robustesse et la fiabilité du modèle.