

Fusion d’approches classiques et d’apprentissage profond pour la classification du mélanome

Jules Collenne

Aix-Marseille Université

jules.collenne@etu-univ.amu.fr

Encadrant Rabah Iguernaissi

Aix-Marseille Université

rabah.iguernaissi@univ.amu.fr

**Mémoire présenté pour l’obtention du Master
Intelligence Artificielle et Apprentissage Automatique
de l’Université d’Aix-Marseille
Juin 2021**

Abstract

Le mélanome est un cancer cutané potentiellement mortel qu’il est possible de diagnostiquer à un stade précoce à l’aide d’images dermoscopiques. Pour ce faire, trois principales méthodes de diagnostic par ordinateur existent. Les approches classiques, qui utilisent l’extraction de features expertes, les approches par réseaux de neurones convolutifs (CNN), et la fusion de ces deux approches. Cet article présente une méthode par fusion utilisant deux modèles. Le premier est développé à partir d’extraction de features expertes innovantes modélisant l’asymétrie des lésions, tandis que le second est une approche ensembliste de CNN issus de l’état de l’art. Le premier modèle atteint un score AUC de 0.876, le second de 0.938 tandis que la fusion les surpasse avec un AUC de 0.942 sur un sous-ensemble de la base ISIC 2019.

1 Introduction

Le mélanome est une tumeur maligne se présentant sous la forme d’une lésion pigmentée cutanée. Parmi les cancers de la peau, il est l’un des plus meurtrier et constitue un problème de santé publique majeur : En France, plus de 1 700 décès sont rapportés en 2017, et son taux d’incidence augmente de 2% tous les ans dans les pays développés (nbm, 2021). En France, cette augmentation est estimée à plus de 15 000 nouveaux cas par an, avec un taux de mortalité de 1.7 pour les hommes et 0.9 pour les femmes pour 100 000 individus. Pourtant une simple chirurgie locale permet de soigner le patient lors des premiers stades de la maladie. Mais son diagnostic n’est pas aisé. À un stade précoce, il est difficile de les différencier des autres lésions mélanocytaires bénignes appelées naevus ou plus communément grains de beauté, qui peuvent être présents en grande quantité et qu’il est inutile de retirer (cicatrices inutiles, potentielles complications de chirurgie). Ainsi, de nombreux travaux cherchent à diagnostiquer au mieux ce cancer, notamment à l’aide d’algorithmes de classification visant à différencier les mélanomes des naevus. Ces techniques peuvent permettre d’aider les dermatologues à réaliser le diagnostic plus rapidement, et ainsi augmenter le nombre de patients analysés.

Actuellement, les CNN permettent de diagnostiquer la maladie avec des performances équivalentes à celles des dermatologues (Brinker et al., 2019). Mais bien que ces modèles soient très performants, plusieurs problèmes persistent. Particulièrement, le manque d’interprétabilité de leurs résultats a pour conséquence de ralentir leur usage par manque de confiance de la part des dermatologues. De plus, le temps d’entraînement et d’inférence des modèles peut être extrêmement long et coûteux en ressources et ils nécessitent d’importantes bases de données. Il est alors intéressant d’étudier différentes pistes afin de pallier ces problèmes. Par exemple, l’usage de features expertes peut permettre une meilleure compréhension des prédictions des classifieurs par les dermatologues, et un moindre coût lié à l’entraînement et l’usage des modèles. L’élaboration de descripteurs, en lien avec des dermatologues, peut permettre d’extraire des caractéristiques que l’on sait utiles au diagnostic, et ainsi guider la classification.

Dans cet article, nous avons implémenté différentes approches afin d'en faire une étude comparative. Il s'agit d'abord d'une approche classique basée sur des features expertes caractérisant l'asymétrie de couleurs des lésions, puis d'une approche par CNN, et enfin d'une fusion des deux approches. Nous comparons tous ces modèles entre eux ainsi qu'avec une méthode ensembliste de CNN représentant l'actuel état de l'art (Ha et al., 2020). Les performances, mais aussi la quantité de calculs et l'interprétabilité des modèles sont comparés. Notre modèle a donc pour but de rendre le diagnostic du mélanome plus rapide afin de permettre aux dermatologues d'analyser un nombre plus important de lésions sans perdre en performance. Notre modèle d'approche classique pourra aussi être utilisé afin d'analyser automatiquement de grandes quantités d'images sans avoir besoin d'aide humaine. Une fois les naevus symétriques écartés, un dermatologue pourra ensuite analyser les instances de mélanomes afin de réaliser le diagnostic.

Nous étudions d'abord l'état de l'art pour chaque approche, puis expliquons nos méthodologies, avant de présenter les résultats de chaque modèle développé.

2 État de l'art

Comme évoqué précédemment, plusieurs méthodes coexistent actuellement dans la littérature. Les approches classiques, celles par CNN et la fusion des deux. Les premières approches de classification assistées par ordinateur sont dites classiques. Ces dernières reposent sur l'extraction de features expertes. Pour ce type de méthodes, les connaissances dermatologiques sont utilisées afin de créer des caractéristiques discriminant au mieux les naevus des mélanomes. Ces features sont ensuite transmises en entrée d'un classifieur qui réalise le diagnostic automatiquement.

La plupart des travaux précédents s'inspirent de règles dermatologiques connues pour caractériser les mélanomes dont la plus populaire est la règle ABCDE "Asymétrie, Bords irréguliers, Couleurs différentes et irrégulières, Diamètre, Évolution". En pratique, les descripteurs vont être la couleur moyenne de l'image, l'écart-type, une analyse de teintes sombres et l'irrégularité des contours... Toutefois, peu d'articles étudient réellement l'asymétrie. L'asymétrie de contours a été étudiée par (Garnavi et al., 2012), celle de teintes sombres par (Chang et al., 2005) mais l'asymétrie de couleur dans la lésion n'est pas vraiment exploitée, alors que ce concept est très important pour différencier les mélanomes des naevus. Dans ce contexte, le critère important lors de la classification est la disparité de couleurs, que ce soit un changement de teinte ou encore la présence de couleurs différentes. Par ailleurs, dans ces méthodes, les classifieurs utilisés sont généralement des SVM comme dans l'article (Rastgoo et al., 2015), un Random Forest (Janney.J and Roslin, 2020), ou des méthodes de Gradient Boosting dans l'article (Celebi et al., 2007) par exemple.

L'apport de cet article pour les approches classiques est de proposer de nouvelles features innovantes permettant de modéliser l'asymétrie de couleur au sein de la lésion. Les modèles développés sont entraînés afin de discriminer les mélanomes des naevus en utilisant seulement ce critère, pour que les prédictions réalisées soient compréhensibles pour les dermatologues tout en ayant de bonnes performances. Le modèle utilisé est un ANN dont l'architecture est présentée à la figure 6.

Quant à la classification d'images par CNN, cette approche est actualisée chaque année par de nouveaux algorithmes toujours plus performants. AlexNet (Krizhevsky et al., 2012), Inception (Szegedy et al., 2014) puis ResNet (He et al., 2015) et enfin EfficientNet (Tan and Le, 2020) et son amélioration EfficientNetV2 (Tan and Le, 2021) améliorent d'année en année les performances sur les tâches de vision par ordinateur dont la référence est ImageNet (Deng et al., 2009). Rappelant qu'en 2011, le meilleur classifieur AlexNet avait une accuracy top 1 à 0.63 tandis qu'il est aujourd'hui à 0.90 grâce à EfficientNet.

Afin d'étudier le cas du mélanome en particulier, il existe bien évidemment la littérature scientifique dans le domaine, mais aussi le challenge SIIM-ISIC ayant lieu chaque année. Cette compétition, en ligne sur le site Kaggle, permet à toute personne intéressée par le sujet d'évaluer son modèle de classification sur une base de donnée disponible au public (appelée base de données SIIM-ISIC ou simplement ISIC).

Les vainqueurs de cette compétition s'engagent à rendre public leur modèle, leur code source ainsi qu'un article expliquant leur programme. Il est donc aisé d'analyser l'état de l'art dans la classification du mélanome à partir des meilleurs modèles de cette compétition.

Par exemple, les vainqueur d'ISIC 2020 Ha et al. (2020) ont utilisé une méthode ensembliste de réseaux de neurones regroupant plusieurs modèles EfficientNet fine-tunés. Une méthode similaire a donc été développée dans nos expériences afin de comparer nos différents modèles.

Enfin, certains articles se sont aussi intéressés à la fusion des deux approches, mais ces derniers sont peu nombreux. Généralement, la fusion utilise les méta données des patients tel que leur âge, sexe, et position de la lésion sur le corps. C'est le cas dans l'article de (Yildirim Yayilgan et al., 2020) ainsi que celui de (Hagerty et al., 2019). Leur descripteurs évaluent différentes caractéristiques comme la couleur de la lésion et sa forme, mais ne prennent pas en compte l'asymétrie. Comparés aux modèles seuls, leurs résultats sont toutefois bien meilleurs lors de la fusion des deux approches. Cette conclusion laisse donc penser qu'il est possible d'améliorer les résultats avec de meilleures descripteurs et d'une meilleure fusion.

L'apport de cet article est donc la création de features innovantes, la classification de ces dernières à l'aide d'un ANN, la fusion de ce dernier avec les CNN actuels, puis la comparaison de chacun de ces modèles.

3 Méthodes

Cette partie présente les données utilisées et la méthodologie de chacun des modèles développé en expliquant d'abord le modèle classique, puis les CNN et enfin la fusion des deux.

3.1 Données

La plus grande base de données publique de dermoscopies, ISIC, élaborée par l'académie du même nom (International Skin Imaging Collaboration) a été utilisée. Cette dernière est actualisée chaque année, nous avons donc récupéré celle de 2019. Il s'agit d'un ensemble d'images dermoscopiques classées en deux catégories : Mélanomes et Naevus. Les images proviennent de sources différentes, certaines ont des artefacts tel que des poils et des bulles, et la luminosité, saturation, et teinte peuvent grandement varier d'une image à l'autre. L'approche classique nécessitant une segmentation des lésions, une dermatologue s'est chargée d'enlever les images contenant trop d'artefacts qui rendaient la segmentation impossible. La base finale ne contient donc pas toutes les images d'ISIC 2019, mais seulement un sous-ensemble assaini.

La base résultante compte alors 7 600 images, dont 6 300 naevus et 1 300 mélanomes. Le problème principal réside dans la quantité d'image par classe : Il y a presque 5 fois plus de naevus que de mélanomes (un ratio de 4.59 mélanomes par naevus). Cette disproportion, si elle n'est pas prise en compte, peut engendrer une baisse considérable de performances pour nos modèles de classification. Ces derniers ont donc à chaque fois un poids favorisant les mélanomes.

La base de données contient également les images segmentées (la peau a été enlevée, et seule la lésion est présente), ainsi que les masques de ces segmentations (images binaires représentant la segmentation). Cette segmentation a été réalisée par les chercheurs de l'équipe Image et Modèle du Laboratoire Informatique et Système avec un réseau de neurones U-Net. Plusieurs exemples d'images de la base de données sont affichées figures 1 et 2.

3.2 Approche classique

Pour caractériser le concept d'asymétrie de couleur, nous avons choisi de découper la lésion en quatre parties égales appelées quadrants en suivant les axes principaux de symétrie. L'idée est de comparer les différents quadrants à la fois par rapport au centre mais aussi par rapport aux axes. Ces différents calculs sont réalisés uniquement sur la lésion en utilisant la segmentation faite par U-Net, afin de ne pas prendre en compte la peau environnante. Plusieurs descripteurs ont été implémentés afin de caractériser cette notion d'asymétrie au niveau des couleurs.

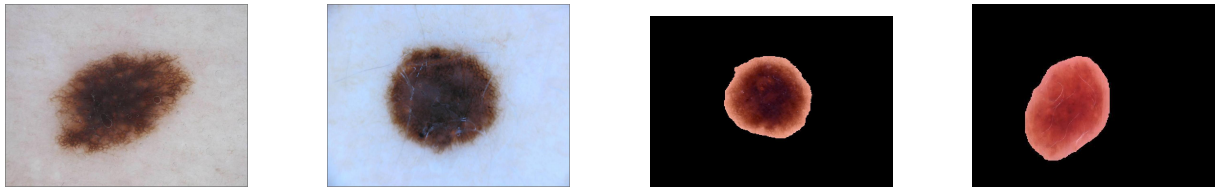


FIGURE 1 – Exemple d’images dermoscopiques dans le cas des naevus. À gauche, deux images originales et à droite deux naevus segmentés. On remarque que la teinte d’une image varie grandement d’une image à l’autre



FIGURE 2 – Exemple de dermoscopies dans le cas des mélanomes. À gauche : Deux images originales, à droite : deux images segmentées

3.2.1 Axes de symétrie et des quadrants

La première étape consiste à calculer les axes de symétrie des lésions afin de créer les quadrants. Pour ce faire, deux méthodes ont été essayées :

- PCA
- Axe le plus symétrique

La méthode PCA consiste à voir chaque pixel de la lésion segmentée comme un point dans l’espace, puis d’appliquer la PCA sur ce nuage de point afin d’avoir les axes réduisant un maximum la variance.

La méthode de l’axe le plus symétrique est calculé en essayant de nombreux axes en gardant celui réduisant au maximum la différence d’aire des deux moitiés de la lésion. On peut imaginer une image pliée en deux selon cet axe dont les pixels différents seraient comptabilisés. Tous les axes entre 0 et 180 degrés sont essayés avec un pas de 10 degrés.

La deuxième méthode a été retenue car étant la plus rapide et offrant des axes plus pertinents.

Un simple calcul du déterminant permet ensuite de connaître la position relative de chaque pixel par rapport aux axes afin de créer les quadrants. Une fois les lésions découpées, nous pouvons calculer les features sur chacun des quadrant, puis les comparer entre eux. La figure 3 montre les différentes étapes simplifiées. De plus, les quadrants ont tous un indice allant de 0 à 3. Pour toutes les lésions, le quadrant 0 est toujours symétriquement opposé au quadrant 3 par rapport au centre, de même pour les quadrants 1 et 2.

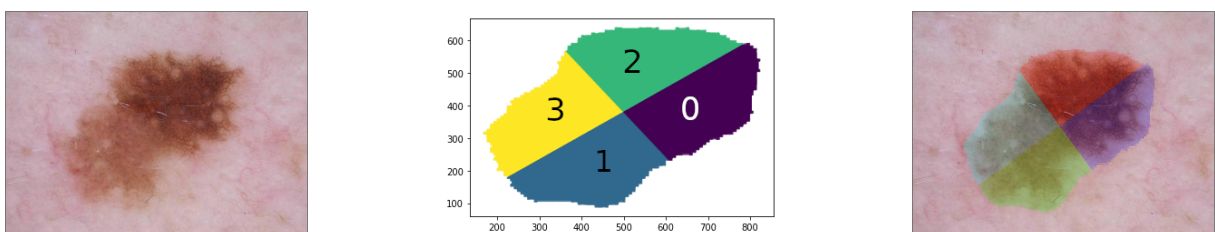


FIGURE 3 – De gauche à droite : Dermoscopie d’un naevus, nuage de points de la lésion colorié par quadrant avec leur indice, superposition des quadrants sur la lésion

3.2.2 Features

Les features calculées par quadrants sont les suivantes :

- Couleur moyenne
- Ecart-type
- Asymétrie statistique (skewness)
- Kurtosis
- Entropie
- Histogrammes de couleurs avec 10 bins

La figure 4 montre deux exemples de la couleur moyenne par quadrant.



FIGURE 4 – Couleur moyenne par quadrant : A gauche d'un naevus, à droite d'un mélanome

Ensuite, nous allons définir deux modélisation de l'asymétrie : Une par rapport aux axes, et l'autre par rapport au centre. Le vecteur final donné en entrée aux classifieurs sera la concaténation de ces deux modélisations de l'asymétrie.

3.2.3 Asymétrie par rapport aux axes

Pour chaque feature F , l'asymétrie par rapport aux axes est définie par :

$$A_{axes} = \frac{abs(F_0 - F_1) + abs(F_2 - F_3) + abs(F_0 - F_2) + abs(F_1 - F_3)}{4}$$

Avec F_i la feature calculée sur le quadrant i d'une lésion.

Nous rappelons que le quadrant 0 est toujours opposé au quadrant 3 par rapport au centre, de même pour le quadrant 1 et 2.

3.2.4 Asymétrie par rapport au centre

Pour chaque feature F , l'asymétrie par rapport au centre est définie par :

$$A_{centre} = \frac{abs(F_0 - F_3) + abs(F_1 - F_2))}{2}$$

Avec F_i la feature calculée sur le quadrant i d'une lésion.

Feature innovante : Cercles sur image superpixel

La dernière feature réalisée n'utilise pas les quadrants et consiste en une comparaison de pixels symétriques par rapport au centre.

Des paires de pixels symétriques entre eux par rapport au centre sont récupérés, puis comparés.

D'abord, la lésion est transformée en image superpixel avec l'algorithme SLIC (Achanta et al., 2012) afin de réduire le bruit que pourraient contenir les pixels choisis. Chaque image contient environ 200 superpixels.

Des cercles de rayons différents centrés sur la lésion ainsi que des lignes sont placés sur l'image, comme présenté sur la figure 5. La différence en valeur absolue est ensuite effectuée entre les pixels symétriques entre eux par rapport au centre. Enfin, les valeurs sont moyennées par cercle. La taille du vecteur en sortie est donc de $n_{circles} * 3$ (3 correspondant aux trois canaux RGB).

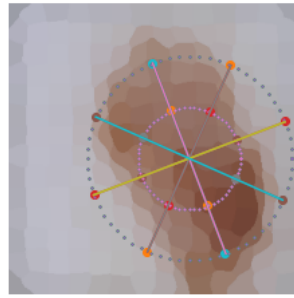


FIGURE 5 – Représentation des cercles sur superpixels sur un naevus

3.2.5 Pré-traitement des features et espaces de couleurs

L'usage de différents espaces de couleurs peut dans certains cas permettre d'atteindre de meilleurs résultats lors de la classification. Pour chacune des features, nous avons utilisé les espaces de couleur RGB, HSV, YCbCr et CieLab. Bien qu'il s'agisse de la même image, sa représentation est différente, mettant en avant certaines caractéristiques pouvant faciliter la classification. Par exemple, HSV (Hue, Saturation, Value soit Teinte, Saturation et Valeur) est très utilisé en traitement de l'image car cette modélisation permet de regrouper facilement les couleurs similaires malgré les changements de luminosité. Plusieurs expériences ont été menées et l'usage de différents espaces de couleurs augmente effectivement les résultats comme le montre la table 1.

Espaces de couleurs	Précision équilibrée	AUC
RGB	0.72	0.85
HSV	0.78	0.86
Cielab	0.74	0.83
YCbCr	0.75	0.85
RGB+HSV+CieLab+YCbCr	0.79	0.88

TABLE 1 – Scores sur la base de test en fonction des espaces de couleurs utilisés. Résultats obtenus avec l'ANN présenté figure 6

Avant la concaténation, les features sont normalisées à l'aide de l'algorithme MinMax. Une fois normalisées entre 0 et 1, elles sont concaténées puis centrées en zéro. Les valeurs du vecteur final sont donc comprises entre -0.5 et 0.5.

3.2.6 Modèle choisi

Le modèle final est un réseau de neurone (ANN) dont l'architecture est présentée figure 6. Les instances données en entrée au modèle sont de taille 714, ce qui correspond à la concaténation des features sur les espaces de couleurs différents. L'entraînement est effectué sur 50 epochs, avec une taille de batch de 64, et un poids par classe de 4.59 pour les mélanomes et 1 pour les naevus. La fonction de loss utilisée est la categorical cross entropy et l'optimizer est Adam avec un learning rate de 0.001. Différents modèles ont été essayés et leurs résultats sont présentés en annexe sur la table 5. Enfin, le modèle compte 11 474 paramètres.

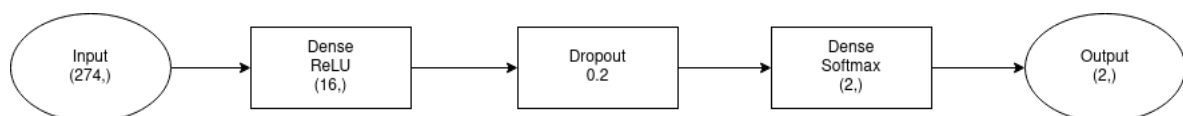


FIGURE 6 – Diagramme du réseau de neurones utilisé pour l'approche classique

3.3 Approche par réseaux de neurones convolutifs

Deux manières de classifications ont été utilisées dans la partie réseaux de neurones. D'abord l'utilisation d'un CNN seul, EfficientNetB2, puis une méthode ensembliste regroupant plusieurs EfficientNet (de B0 et B7). Ces modèles ont été choisis car ce sont actuellement les meilleurs dans de nombreux problèmes de vision par ordinateur, tout en étant relativement léger en terme de nombre de paramètres. Leur fonctionnement est rappelé en annexe figure 16 et leur architecture est présentée en annexes sur la figure 17.

3.3.1 Entraînement des CNN

L'entraînement se base sur deux parties. D'abord, les différents EfficientNet sont récupérés avec les poids calculés sur ImageNet et leur poids sont gelés. La dernière couche est retirée et remplacée par une couche de softmax pour notre classification binaire, qui sera alors la seule couche entraînée lors d'un premier entraînement.

Une fois cette couche entraînée, un deuxième entraînement a lieu avec tous les poids dégelés et un petit learning rate afin de réaliser le fine tuning. Enfin, les poids sont sauvegardés afin de réutiliser les modèles facilement pour l'approche ensembliste.

Pour causes de contraintes matérielles, le deuxième entraînement n'a pas lieu sur les modèles B5, B6 et B7 car leurs entrées sont trop lourdes.

Le premier entraînement est réalisé avec 15 epochs, l'optimiseur Adam avec un learning rate de 0.001 et la fonction de perte binary crossentropy. Pour pallier la disproportion des données, un poids de 4.59 a été donné à la classe mélanome, tandis que la classe naevus a un poids de 1. Lors du deuxième apprentissage, le learning rate est à 10^{-4} et dure 5 epochs. Les autres paramètres sont les même que le premier entraînement.

3.3.2 Image augmentation

Pour avoir un bon apprentissage, l'augmentation d'images est nécessaire. Il s'agit de créer de nouvelles images à partir de celles existantes afin d'améliorer la généralisation du modèle. La méthode utilisée effectue des opérations aléatoires sur les images avant de les faire passer par les couches de convolutions. Plusieurs opérations sont effectuées : rotation, translation, image miroir et changement de contraste. Pour avoir une augmentation de donnée bien diversifiée, ces opérations ont toutes une part d'aléatoire, que ce soit dans leur application (certaines images vont avoir une rotation, d'autres non) que dans leur degré de modification (la rotation est aléatoire entre 0 et 360 degrés).

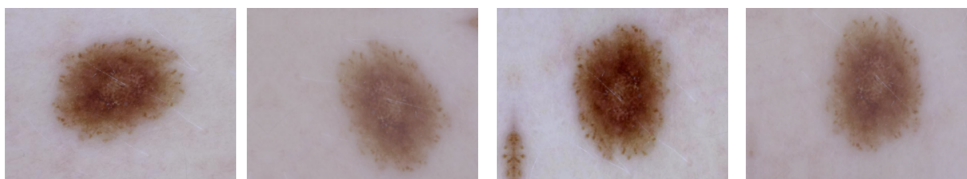


FIGURE 7 – Image originale d'un naevus (à gauche) suivis de ses augmentations

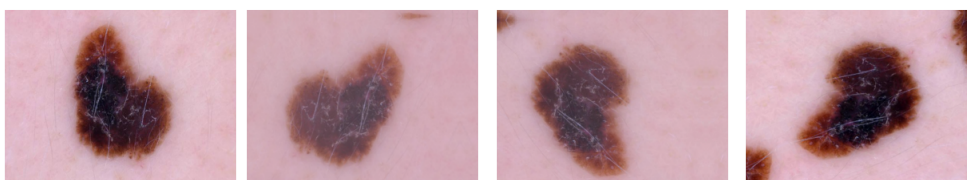


FIGURE 8 – Image originale d'un mélanome (à gauche) suivis de ses augmentations

3.3.3 Méthode ensembliste

L'état de l'art consiste à entraîner indépendamment plusieurs classifieurs, puis à les regrouper pour réaliser les prédictions. Afin de réaliser ceci, deux méthodes de late fusion ont été réalisées :

- Faire la moyenne des prédictions de tous les modèles
- Réaliser un vote, la classe prédite pour une instance donnée étant celle récoltant le plus de votes

La figure 14 en annexe montre le fonctionnement de la late fusion dans le cas de la moyenne des prédictions.

Les modèles utilisés pour cette méthode sont les EfficientNet de B0 à B7. L'entraînement a été le même pour tous à la différence de la taille d'entrée des images. Les instances sont redimensionnées pour chaque modèle à la taille leur correspondant. La liste de la taille d'entrée pour chacun se trouve sur le tableau 6 en annexe. Quant au nombre de paramètres, ils varient entre 5 millions pour B0 et 66 millions pour B7, contrairement au précédent ANN disposant d'un peu plus de 11 000 paramètres.

3.4 Fusion des deux approches

Enfin, il est possible de fusionner les CNN avec l'ANN entraîné sur les features expertes. Le but est d'utiliser l'information des features experte pour aider le CNN à corriger ses erreurs sur certaines instances que nous savons asymétriques. La fusion est réalisable de plusieurs manières :

- Late fusion avec moyennes des prédictions
- La fusion hybride : Entraîner les modèles indépendamment, retirer leur dernière couche, puis concaténer leurs sorties. Celles-ci sont données en entrée à un ANN final (voir figure 15)

L'ANN final de la fusion hybride est composé de deux couches denses ReLU de taille 32 et 16 puis d'une couche Dense d'activation Softmax finale sortant deux scalaires pour la classification. L'entraînement est effectué avec les deux modèles déjà entraînés et leurs poids gelés. Seul l'ANN final apprend lors de 20 epochs. Le batch size est à 32, l'optimiseur est Adam et la loss est la binary cross entropy.

4 Expériences et résultats

Tous les modèles ont été entraînés sur la même base d'entraînement, puis évalués avec les mêmes métriques sur la même base de test. Les entraînements ont été effectués sur une NVIDIA Tesla K80, à l'aide des bibliothèques Tensorflow 2.5.0 et Keras 2.5.0.

4.1 Métriques utilisées

L'aire sous la courbe ROC (AUC) est une métrique très utilisée dans ce genre de problème de classification. Elle est utilisée notamment dans le challenge SIIM-ISIC 2020. C'est une bonne manière d'évaluer les performances du modèle, mais qui peut être biaisée par la disproportion de classes. L'AUC n'est donc pas suffisant pour évaluer nos modèles. Deux autres métriques sont alors utilisées : La sensibilité et la spécificité. Très utilisées en médecine, elles permettent de connaître la proportion d'instances bien classées d'une part pour les mélanomes (sensibilité), d'autre part pour les naevus (spécificité).

$$Sensibilité = \frac{VP}{VP + FN}$$

$$Spécificité = \frac{VN}{VN + FP}$$

Avec VP les vrais positifs, VN les vrais négatifs, FP les faux positifs et FN les faux négatifs.

Nous utilisons également la précision équilibrée qui consiste à faire la moyenne de la sensibilité et de la spécificité, et nous permet d'avoir une bonne vision de la proportion de classes bien prédites :

$$PrécisionÉquilibrée = \frac{Sensibilité + Spécificité}{2}$$

4.2 Approches classiques

Le modèle final de l'approche classique est l'ANN entraîné sur le vecteur de caractéristiques. Les résultats obtenus sont très satisfaisants. Le modèle final issu de la concaténation des features a de bien meilleurs résultats que chacune des features indépendantes. Il atteint alors une précision équilibrée de 0.80 ainsi qu'un AUC de 0.88 comme le montrent la table 2 et la figure 9.

Modèle	Sensibilité	Spécificité	Précision équilibrée	AUC
Features par quadrant sans histogrammes	0.61	0.80	0.71	0.80
Histogrammes par quadrants	0.87	0.66	0.77	0.82
Cercles sur superpixels	0.73	0.74	0.74	0.81
Modèle final	0.82	0.77	0.80	0.88

TABLE 2 – Scores par feature sur la base de test avec un ANN

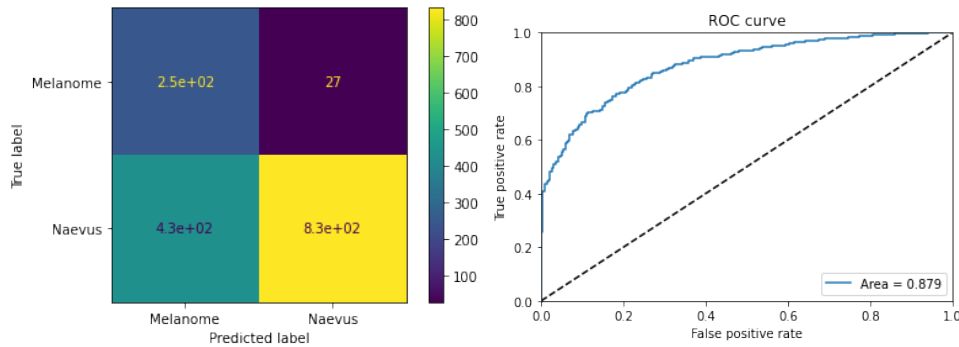


FIGURE 9 – Matrice de confusion et courbe ROC sur la base de test avec l'ANN final

4.3 Réseaux de neurones

Les scores des CNN sont présentés sur le tableau 3. On remarque que le meilleur modèle seul est EfficientNetB2 avec une précision équilibrée à 0.82 ainsi qu'un AUC à 0.91 (voir figure 10). Il est toutefois largement dépassé par la méthode ensembliste qui dépasse chacun des modèles seuls (voir figure 11). Quant au vote majoritaire, il n'a pas de score AUC car il ses prédictions sont binaires (0 ou 1), mais il atteint la même précision équilibrée que la moyenne des prédictions. Le regroupement de plusieurs modèles a donc effectivement permis d'augmenter considérablement les performances, bien qu'avec un nombre de calculs plus élevés.

Modèles	Précision équilibrée	AUC
EfficientNetB0	0.79	0.888
EfficientNetB1	0.80	0.895
EfficientNetB2	0.82	0.910
EfficientNetB3	0.80	0.900
EfficientNetB4	0.81	0.909
EfficientNetB5	0.79	0.877
EfficientNetB6	0.78	0.872
EfficientNetB7	0.78	0.870
Ensemble (moyennes prédictions)	0.85	0.938
Ensemble (vote majoritaire)	0.85	/

TABLE 3 – Liste des modèles utilisés et leur score respectif sur la base de test. Un score en gras est la valeur maximale de la colonne

Pour une meilleure estimation des performances de ces modèles, le score en cross validation sur 5 folds a été réalisé et sont affichés en annexe table 7.

Il est par ailleurs intéressant de noter que l'ANN défini dans l'approche classique obtient des résultats relativement similaires à ces modèles, alors que sa taille est largement plus petite.

4.4 Fusion des deux approches

Les méthodes de fusions différentes ont donné des résultats plutôt différents, décrit dans ce chapitre.

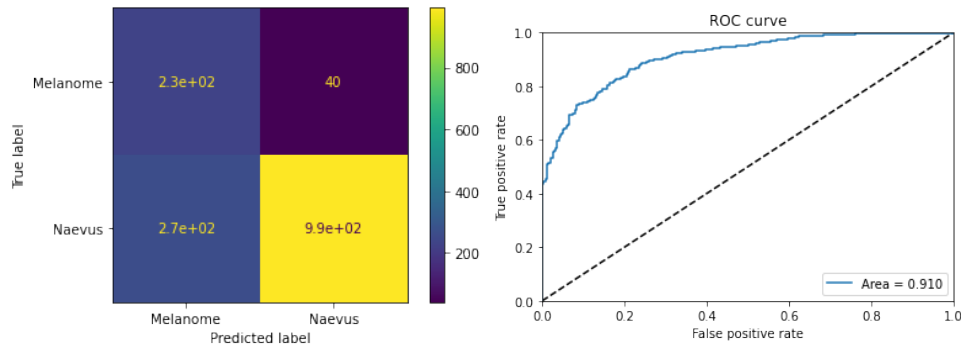


FIGURE 10 – Matrice de confusion et courbe ROC sur la base de test avec B2

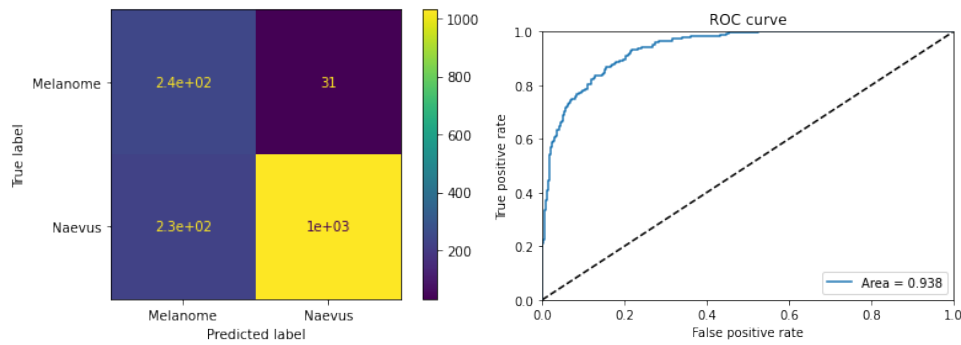


FIGURE 11 – Matrice de confusion et courbe ROC sur la base de test avec l'ensemble de CNNs

4.4.1 Late fusion ANN+B2

La late fusion a obtenu de très bons résultats sur la base de test. La méthode utilisée est la moyenne des prédictions de l'ANN entraîné sur les handcrafted feature et celles d'EfficientNetB2.

Alors qu'EfficientNetB2 seul atteint un score AUC de 0.91 et l'ANN un AUC de 0.88, la fusion avec l'ANN permet d'atteindre un score AUC de 0.92. Cette dernière a donc amélioré un petit peu les résultats des deux modèles indépendants. Le modèle de l'approche classique a ainsi réussi à corriger certaines instances sur lesquelles le CNN se trompait et vice-versa, sans pour autant fusionner leurs erreurs.

Les résultats n'atteignent pas ceux de la méthode ensembliste de CNN, mais sont toutefois convenable au vu du nombre extrêmement plus faible de calculs. Là où la méthode ensembliste requiert 8 modèles, la fusion entre ANN et CNN ne compte que deux modèles dont un ANN très petit.

4.4.2 Fusion hybride ANN+B2

La fusion hybride entre ANN et B2 a eu des résultats équivalent à la late fusion, avec une précision équilibrée à 0.83 et un score AUC à 0.92. Le choix de l'ANN final, nombre de couches, taille des couches, nombre d'epochs sont de nombreux paramètres qu'il est possible de modifier et qui pourraient potentiellement améliorer les résultats dans des travaux ultérieurs. Les performances actuelles sont toutefois meilleures que les deux modèles indépendants.

4.4.3 Late fusion : Tous les CNN + ANN

Enfin, les meilleurs résultats sont détenus par la fusion de tous les modèles. Il s'agit de tous les EfficientNet ainsi qu'à l'ANN choisi. Bien que l'ANN se retrouve parmi 7 CNN, son ajout augmente grandement les résultats de la méthode ensembliste de CNN. Cela est probablement dû au fait que les CNN font à peu près les mêmes erreurs alors que l'ANN apporte des prédictions totalement différentes qui peuvent s'avérer très utiles, car on sait que les méthodes ensemblistes ont besoin de diversité pour fonctionner.

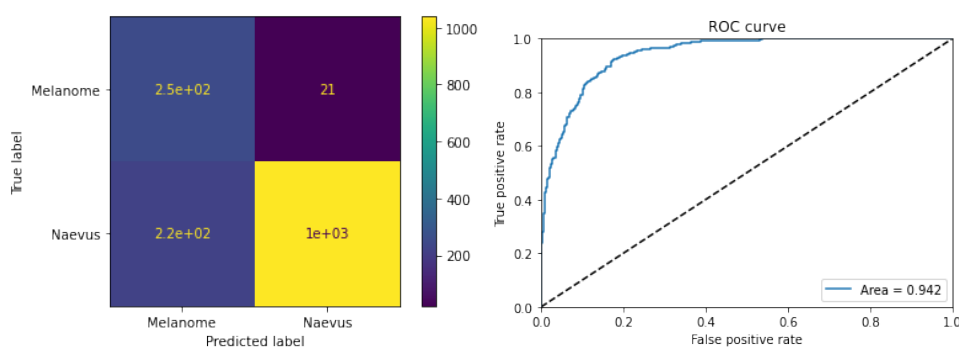


FIGURE 12 – Matrice de confusion et courbe ROC sur la base de test avec l’ensemble des modèles en late fusion (EfficientNet de B0 à B7 + ANN)

4.5 Résultats finaux

Enfin, les résultats finaux sont présentés 4.

Modèle	Sensibilité	Spécificité	Précision équilibrée	AUC
ANN handcrafted features	0.82	0.77	0.80	0.876
EfficientNetB2	0.81	0.81	0.81	0.905
Fusion hybride ANN+B2	0.91	0.74	0.83	0.921
Late fusion ANN+B2	0.85	0.80	0.83	0.915
Ensemble de CNN	0.88	0.80	0.84	0.938
Ensemble de CNN + ANN (late fusion)	0.92	0.82	0.87	0.942

TABLE 4 – Scores par modèles sur la base de test

L’ajout de l’ANN à l’ensemble des CNN a permis d’améliorer les résultats. Seul, l’ANN a des performances proches des CNN mais avec un nombre de paramètres extrêmement plus faible (seulement une dizaine de millier contre plusieurs dizaines de millions pour les CNN), et ses choix sont entièrement basé sur l’asymétrie de couleur, ce qui rend les prédictions compréhensibles pour les dermatologues.

5 Conclusions et perspectives

Dans cet article, nous avons vu qu’il était possible d’améliorer les résultats de CNN à l’aide d’ajout de features expertes pour la classification du mélanome. Les descripteurs choisis sont donc pertinents dans ce contexte. En théorie, les modèles profonds sont capables de faire une approximation de tous types de fonctions si la quantité de donnée est suffisante, mais quand ce n’est pas le cas, il est certain que les features expertes peuvent faciliter l’apprentissage de ces modèles en les incitant à se concentrer sur des aspects de l’image que nous savons discriminant. Notre modèle d’approche classique pourra permettre de détecter automatiquement si une lésion est anormalement asymétrique, et ainsi prévenir le dermatologue qui effectuera le diagnostic en conséquence.

A l’avenir, des améliorations seront possible tels de nouvelles features pertinentes caractérisant la forme par exemple, de meilleurs réseaux de neurones, un apprentissage plus long et une data augmentation plus diverse. La fusion de deux modèles est aussi un domaine complexe, et bien d’autres techniques existantes pourront probablement améliorer les classifications.

References

- [Achanta et al.2012] Radhakrishna Achanta, Appu Shaji, Kevin Smith, Aurelien Lucchi, Pascal Fua, and Sabine Süsstrunk. 2012. Slic superpixels compared to state-of-the-art superpixel methods. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(11) :2274–2282.
- [Brinker et al.2019] Titus J. Brinker, Achim Hekler, Alexander H. Enk, Carola Berking, Sebastian Haferkamp, Axel Hauschild, Michael Weichenthal, Joachim Klode, Dirk Schadendorf, Tim Holland-Letz, Christof von Kalle, Stefan Fröhling, Bastian Schilling, and Jochen S. Utikal. 2019. Deep neural networks are superior to dermatologists in melanoma image classification. *European Journal of Cancer*, 119 :11–17.
- [Celebi et al.2007] M. E. Celebi, H. A. Kingravi, B. Uddin, H. Iyatomi, Y. A. Aslandogan, W. V. Stoecker, and R. H. Moss. 2007. A methodological approach to the classification of dermoscopy images. *Comput Med Imaging Graph*, 31(6) :362–373, Sep.
- [Chang et al.2005] Y. Chang, R. J. Stanley, R. H. Moss, and W. Van Stoecker. 2005. A systematic heuristic approach for feature selection for melanoma discrimination using clinical images. *Skin Res Technol*, 11(3) :165–178, Aug.
- [Deng et al.2009] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. 2009. Imagenet : A large-scale hierarchical image database. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pages 248–255.
- [Garnavi et al.2012] R. Garnavi, M. Aldeen, and J. Bailey. 2012. Computer-aided diagnosis of melanoma using border and wavelet-based texture analysis. *IEEE Trans Inf Technol Biomed*, 16(6) :1239–1252, Nov.
- [Ha et al.2020] Qishen Ha, Bo Liu, and Fuxu Liu. 2020. Identifying melanoma images using efficientnet ensemble : Winning solution to the siim-isic melanoma classification challenge.
- [Hagerty et al.2019] Jason R. Hagerty, R. Joe Stanley, Haidar A. Almubarak, Norsang Lama, Reda Kasmi, Peng Guo, Rhett J. Drugge, Harold S. Rabinovitz, Margaret Oliviero, and William V. Stoecker. 2019. Deep learning and handcrafted method fusion : Higher diagnostic accuracy for melanoma dermoscopy images. *IEEE Journal of Biomedical and Health Informatics*, 23(4) :1385–1391.
- [He et al.2015] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2015. Deep residual learning for image recognition.
- [Janney.J and Roslin2020] Bethanney Janney.J and S.Emalda Roslin. 2020. Classification of melanoma from dermoscopic data using machine learning techniques. *Multimedia Tools and Applications*, 79(5) :3713–3728, Feb.
- [Krizhevsky et al.2012] Alex Krizhevsky, Ilya Sutskever, and Geoffrey Hinton. 2012. Imagenet classification with deep convolutional neural networks. *Neural Information Processing Systems*, 25, 01.
- [nbm2021] 2021. Epidémiologie des cancers cutanés. <https://www.e-cancer.fr/Professionnels-de-sante/Depistage-et-detection-precoc/Detection-precoc-des-cancers-de-la-peau/Epidemiologie>. Accessed : 02-06-2021.
- [Rastgoo et al.2015] M. Rastgoo, R. Garcia, O. Morel, and F. Marzani. 2015. Automatic differentiation of melanoma from dysplastic nevi. *Comput Med Imaging Graph*, 43 :44–52, Jul.
- [Szegedy et al.2014] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. 2014. Going deeper with convolutions.
- [Tan and Le2020] Mingxing Tan and Quoc V. Le. 2020. Efficientnet : Rethinking model scaling for convolutional neural networks.
- [Tan and Le2021] Mingxing Tan and Quoc V. Le. 2021. Efficientnetv2 : Smaller models and faster training.
- [Yildirim Yayilgan et al.2020] Sule Yildirim Yayilgan, Blend Arifaj, Masoomah Rahimpour, Jon Hardeberg, and Lule Ahmedi. 2020. Pre-trained cnn based deep features with hand-crafted features and patient data for skin lesion classification. *Lecture Notes in Computer Science*, 5805, 09.

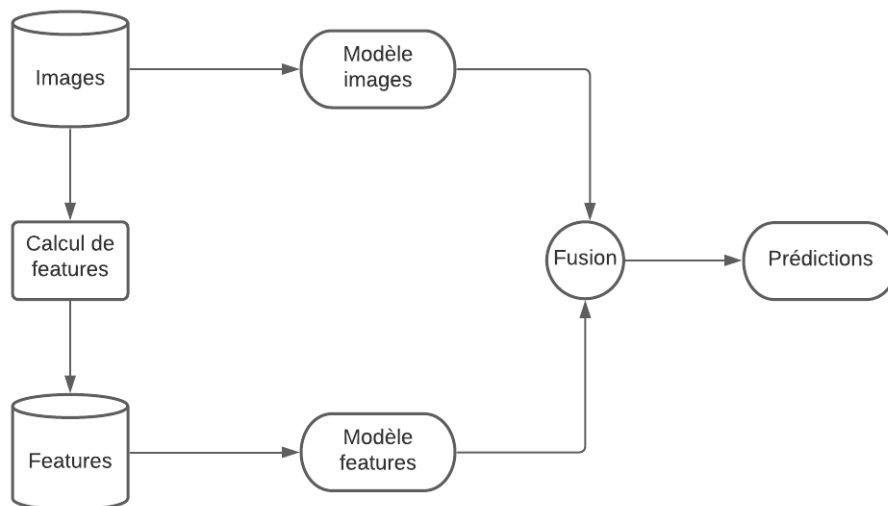


FIGURE 13 – Principe général de la fusion entre deux modèles

Modèle	Précision équilibrée	AUC
LinearSVM	0.75	0.845
SVM	0.76	0.855
GradientBoosting	0.80	0.875
ANN final	0.80	0.876

TABLE 5 – Performances par modèle sur les features

Annexes

Modèle	Resolution d'image en entrée
EfficientNetB0	224
EfficientNetB1	240
EfficientNetB2	260
EfficientNetB3	300
EfficientNetB4	380
EfficientNetB5	456
EfficientNetB6	528
EfficientNetB7	600

TABLE 6 – Résolution d'image par modèle

Modèles	Sensibilité	Spécificité	Précision équilibrée	AUC
EfficientNetB0	0.858	0.736	0.797	0.889
EfficientNetB1	0.824	0.728	0.776	0.873
EfficientNetB2	0.792	0.733	0.763	0.852
EfficientNetB3	0.896	0.591	0.744	0.866
EfficientNetB4	0.847	0.705	0.776	0.892
EfficientNetB5	0.903	0.723	0.813	0.907
EfficientNetB6	0.893	0.717	0.805	0.895
EfficientNetB7	0.905	0.662	0.784	0.894

TABLE 7 – Moyenne des scores en cross validation (5 fold) des différents CNN

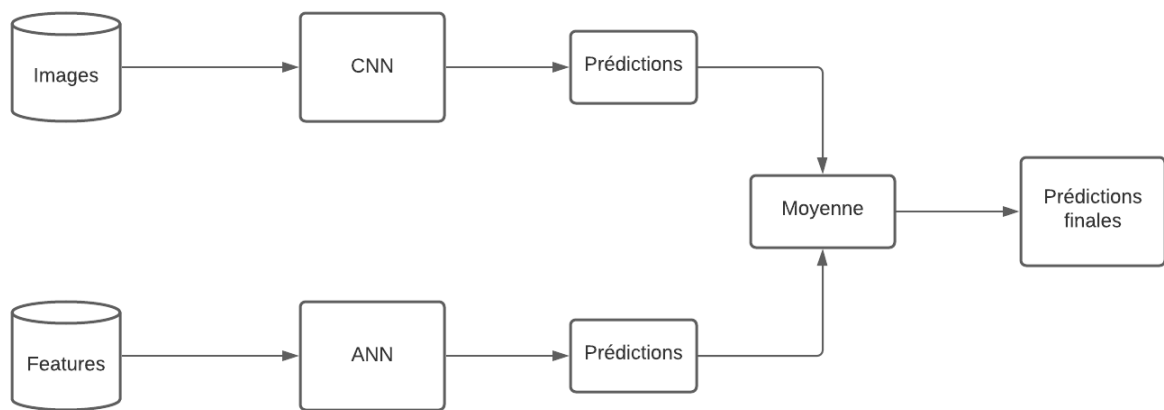


FIGURE 14 – Late fusion

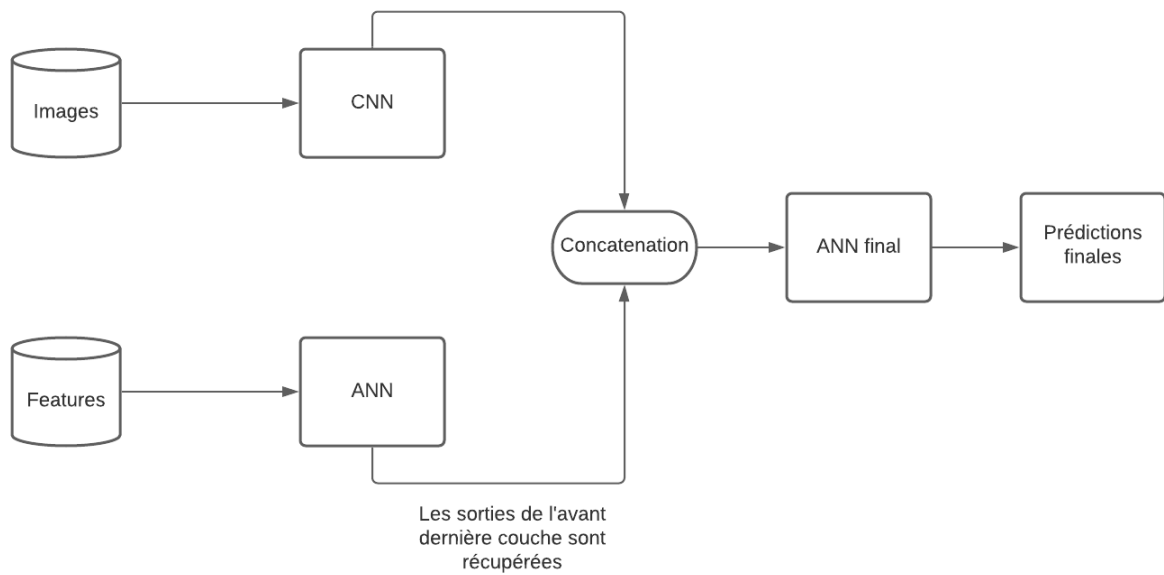


FIGURE 15 – Fusion hybride

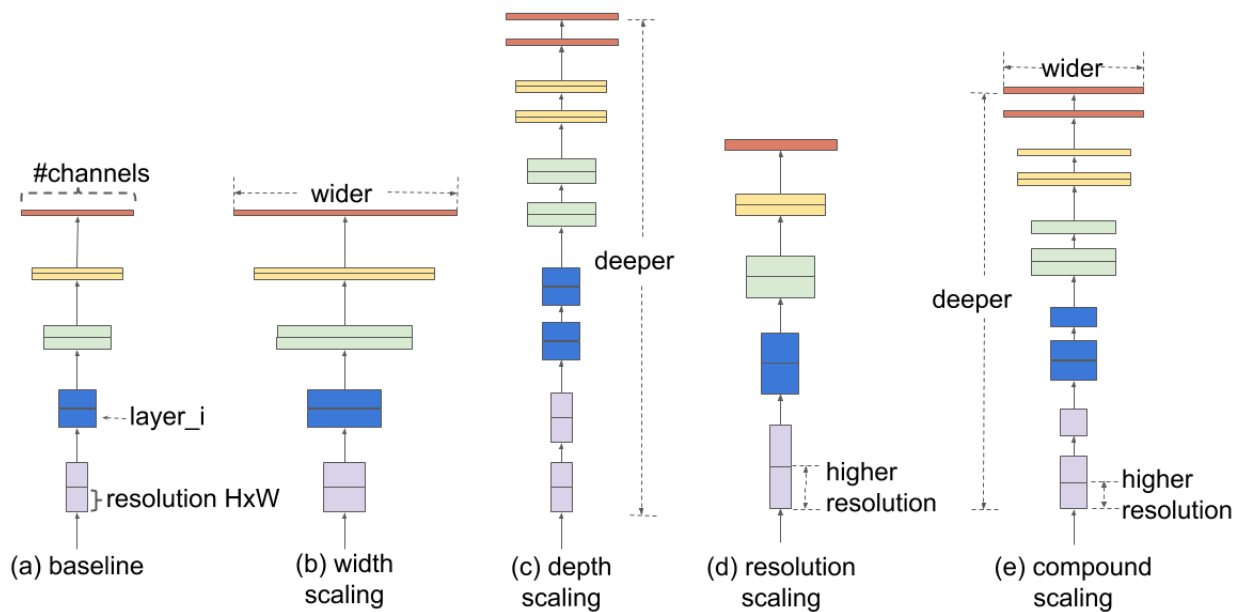


FIGURE 16 – Le modèle EfficientNet utilise des dimensions optimales trouvés en GridSearch permettant d’avoir une architecture à la fois performante et légère

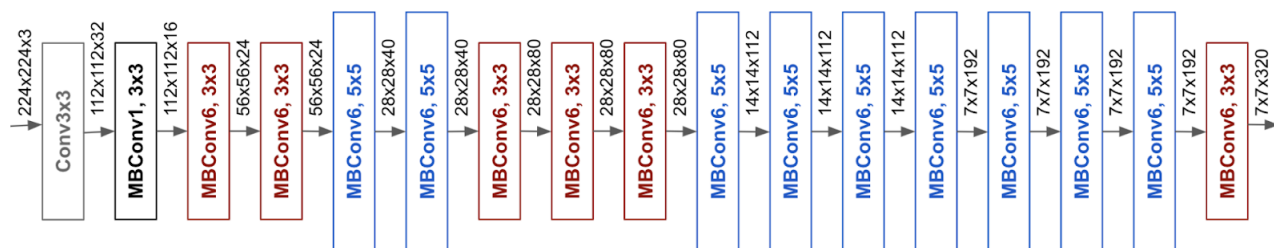


FIGURE 17 – Architecture d’EfficientNetB0