International Journal of Progressive Sciences and Technologies (IJPSAT)
ISSN: 2509-0119.

https://ijpsat.org/
Vol. 56 No. 1 March 2026, pp. 546-563

Apprentissage Profond Pour Le Dépistage Précoce De La Malnutrition Infantile : Evaluation Des Modèles Légers YOLOv8 Et MobileNetV2

ISSIAKA IDI MANAJA

Université de Martyrs du Congo (UNIM-RDC)

Congo, the Democratic Republic of the

Abstract: La malnutrition infantile constitue toujours un défi majeur de santé publique en République Démocratique du Congo, où les techniques classiques de dépistage fondées sur les mesures anthropométriques restent chronophages, exigent des compétences spécialisées et demeurent difficilement déployables dans les zones disposant de moyens limités. La présente étude introduit une méthode innovante basée sur l'apprentissage profond visant à détecter précocement la malnutrition à partir d'images prises via smartphone. Deux architectures légères ont été mobilisées : MobileNetV2 pour les tâches de classification et YOLOv8s pour la détection d'objets.

L'expérimentation s'appuie sur un ensemble d'images annotées d'enfants âgés de 0 à 5 ans, provenant de Roboflow et intégrant une diversité géographique, utilisé pour l'entraînement et l'évaluation des modèles. Les performances obtenues montrent que MobileNetV2 atteint une précision globale de 98 %, tandis que YOLOv8s affiche une précision de 94 % et un mAP@50 de 0,95, attestant de sa capacité à identifier et localiser efficacement différents signes visuels de malnutrition tels que la maigreur, l'atrophie musculaire ou encore la diminution du périmètre brachial.

Ces résultats soulignent le potentiel concret de ces modèles pour une implémentation sur des terminaux mobiles, tant en milieu clinique qu'au sein des communautés. L'étude met ainsi en lumière la valeur ajoutée des architectures légères pour renforcer l'accessibilité au dépistage nutritionnel, notamment dans les contextes où les ressources matérielles et humaines sont limitées.

Keywords : Malnutrition infantile ; Dépistage automatisé ; Deep Learning ; MobileNetV2 ; YOLOv8 ; Vision par ordinateur ; Classification d'images ; Détection d'objets.

I. INTRODUCTION

La malnutrition infantile demeure l'un des défis majeurs de santé publique dans de nombreux pays à faibles et moyens revenus, en particulier en Afrique subsaharienne. En République démocratique du Congo (RDC), l'accès limité aux soins de santé, le manque de ressources et le nombre insuffisant de professionnels de santé qualifiés rendent difficile un dépistage précoce et systématique des enfants souffrant de malnutrition. Les méthodes traditionnelles basées sur des mesures anthropométriques (poids, taille, périmètre brachial, indice de masse corporelle) bien qu'efficaces, exigent du matériel spécialisé, du temps, et une formation que toutes les structures sanitaires ne peuvent garantir. Ces contraintes limitent la couverture du dépistage, retardent la prise en charge et augmentent la morbidité et la mortalité infantiles.

Dans ce contexte, l'intelligence artificielle (IA), et en particulier les techniques d'apprentissage profond, apparaît comme une voie particulièrement prometteuse. L'exploitation d'images à l'aide de réseaux de neurones convolutifs (CNN) rend possible l'automatisation de l'identification de marqueurs visuels liés à la malnutrition tels que l'émaciation, la diminution de la masse musculaire, la finesse du visage ou l'affaissement des joues sans recourir à des mesures anthropométriques directes. Les travaux scientifiques récents confirment largement la pertinence et l'efficacité de ces approches. L'étude de Shilpa Ankalaki et al. (2024) [1], par exemple, a constitué une base d'images d'enfants classés selon leur état nutritionnel (normal, sous-nutris, retard de croissance, etc.) et a atteint une précision de près de 95 % dans une classification binaire (enfants sains vs malnutris), et environ 89 % dans une classification à quatre niveaux.

Une autre recherche menée à partir d'images faciales d'enfants a mobilisé ResNet-50 pour segmenter et classifier les données, obtenant un taux de précision de 98,49 % dans la détection de la malnutrition [2]. Par ailleurs, les avancées récentes montrent que les modèles d'IA peuvent dépasser la simple catégorisation : le modèle NutriScreener, fondé sur des embeddings visuels dérivés de CLIP, une attention graphée multi-pose et des mécanismes de récupération de connaissances, permet non seulement d'identifier la malnutrition, mais aussi d'estimer différents indicateurs anthropométriques à partir d'images 2D d'enfants [3].

Ces progrès s'inscrivent dans un mouvement plus large d'application de l'IA au domaine nutritionnel. Une revue systématique publiée en 2024 souligne l'essor du machine learning et du deep learning dans la reconnaissance automatique des aliments, l'évaluation diététique, la prédiction de carences et, plus récemment, l'identification de la malnutrition à partir d'images ou de données cliniques [4]. Un avantage majeur des architectures modernes (telles qu'EfficientNet, MobileNet ou ResNet) réside dans leur faible coût computationnel, permettant leur déploiement sur des dispositifs mobiles un atout déterminant dans les environnements à faibles ressources. Plusieurs études comparatives récentes ont en effet mis en évidence la pertinence de ces modèles CNN pour l'analyse nutritionnelle [5].

La présente étude s'inscrit dans cette dynamique. Elle examine la performance de MobileNetV2 pour la classification de l'état nutritionnel et de YOLOv8 pour la détection localisée de manifestations visuelles de la malnutrition chez l'enfant, en s'appuyant sur des images 2D capturées via smartphone. Le jeu de données utilisé provient principalement de Roboflow et a été enrichi si nécessaire, afin d'évaluer la robustesse des modèles face à la diversité des âges, des origines ethniques, des postures, des conditions d'éclairage ou encore des types d'habillement. Il s'agit de déterminer si ces architectures légères peuvent fournir un dépistage rapide, extensible et fiable, adapté aux contraintes des contextes à ressources limitées, comme c'est le cas dans de nombreuses régions de la RDC.

Ainsi, l'objectif principal de cet article est de mettre en évidence le potentiel concret de la vision par IA pour le dépistage de la malnutrition infantile en milieu contraint, et de contribuer à préparer son adoption par les agents de santé communautaires, les programmes de nutrition et les organisations humanitaires.

État de l'art

L'application des méthodes d'intelligence artificielle (IA) à l'évaluation de l'état nutritionnel notamment la malnutrition chez l'enfant a connu un essor significatif ces dernières années. Les avancées en vision par ordinateur, apprentissage profond, et architectures multimodales offrent des alternatives innovantes aux mesures anthropométriques classiques. Nous présentons ici les principales approches, leurs apports, ainsi que les limites identifiées dans la littérature récente.

A. Approches de classification d'images par CNN

Les premières tentatives pour détecter la malnutrition via Les applications de l'intelligence artificielle dans ce domaine s'appuient principalement sur des réseaux de neurones convolutifs (CNN) appliqués à des images d'enfants qu'il s'agisse de prises de vue du visage ou du corps.

International Journal of Progressive Sciences and Technologies (IJPSAT)
ISSN: 2509-0119.

https://ijpsat.org/
Vol. 56 No. 1 March 2026, pp. 546-563

Ces modèles recourent fréquemment au transfert d'apprentissage à partir d'architectures préentraînées, ce qui permet d'adapter efficacement le réseau à la tâche de classification « enfant normal vs enfant malnutri ». À titre d'illustration, l'étude de Biradar et al. [6] a utilisé des modèles tels que VGG16 et SqueezeNet pour prédire l'état nutritionnel à partir d'images, obtenant des performances jugées prometteuses.

Plus récemment, des travaux mobilisant ResNet-50 ont rapporté une précision de 98,49 % pour distinguer les enfants malnutris des enfants sains à partir de photographies faciales [2], une performance supérieure à celles obtenues avec d'autres architectures évaluées dans la même étude (VGG16, Xception, MobileNet, etc.) [2].

Ces approches présentent plusieurs atouts : elles sont relativement simples à mettre en œuvre, permettent une adaptation rapide grâce au transfert d'apprentissage et peuvent être déployées sur des dispositifs peu puissants tels que smartphones ou tablettes un élément particulièrement pertinent dans les environnements à ressources limitées.

Cependant, elles ne sont pas exemptes de limites. Comme le soulignent plusieurs revues de littérature, nombre de ces travaux reposent sur des jeux de données restreints et peu diversifiés souvent homogènes en termes d'origine ethnique, de conditions lumineuses ou de posture ce qui soulève des questions quant à la capacité des modèles à se généraliser à des populations plus variées [7].

B. Approches multimodales et morphologiques

Pour dépasser les limites inhérentes à l'utilisation exclusive d'images 2D, des travaux récents s'orientent vers des architectures plus avancées intégrant simultanément plusieurs types d'informations visuelles ou sources de données. L'un des exemples les plus significatifs est le projet NutriScreener (2025), qui mobilise des embeddings visuels dérivés de CLIP, un réseau d'attention multi-pose ainsi qu'un module de recherche de connaissances (« knowledge retrieval ») permettant d'établir des correspondances entre caractéristiques visuelles et indicateurs anthropométriques [3]. Ce système a été testé sur un échantillon de 2 141 enfants provenant de milieux hétérogènes et a atteint un rappel de 0,79, une AUC de 0,82 et des erreurs anthropométriques (RMSE) relativement faibles, indiquant une performance robuste même dans des conditions d'acquisition non standardisées [3].

Parallèlement, l'approche NurtureNet (2024) met en évidence l'intérêt de l'analyse vidéo combinée à des modèles multitâches. Ce modèle intègre à la fois des séquences visuelles issues d'un smartphone et des données tabulaires afin d'estimer plusieurs mesures anthropométriques — poids, taille, périmètre crânien et périmètre thoracique — avec une erreur relative modérée (environ 3,9 %) et une erreur absolue acceptable, notamment pour l'estimation du poids [8]. Grâce à sa taille réduite (~ 15 Mo), NurtureNet constitue une solution apte à fonctionner hors ligne sur smartphone, ce qui le rend particulièrement adapté aux environnements où les ressources sont limitées [8].

Ainsi, ces approches multimodales, combinant images 2D ou vidéos à des données complémentaires, tendent à se rapprocher davantage des méthodes anthropométriques traditionnelles tout en offrant une flexibilité accrue, une absence totale d'invasivité et une facilité de déploiement sur le terrain.

C. Prédiction de malnutrition à l'échelle populationnelle et facteurs de risque via l'IA

Outre les méthodes visuelles, des approches utilisant des Des travaux récents développent des modèles d'apprentissage automatique fondés sur la combinaison de données cliniques, sociodémographiques et environnementales afin de prédire le risque de malnutrition, voire d'anticiper l'émergence de crises nutritionnelles. Une étude réalisée en 2024, par exemple, a exploité plusieurs algorithmes de machine learning pour examiner l'effet de l'insécurité alimentaire sur l'état nutritionnel d'enfants âgés de 6 mois à 5 ans, en intégrant des variables sociodémographiques. Cette approche ouvre des perspectives intéressantes pour repérer plus efficacement les groupes vulnérables [9].

Plus récemment, un rapport publié en 2025 indique qu'un modèle d'intelligence artificielle combinant des données cliniques provenant de plus de 17 000 structures de santé et des données satellitaires liées à la production agricole a permis de prédire la malnutrition aiguë infantile jusqu'à six mois à l'avance. Ce modèle a atteint une précision de 89 % pour une prédiction à un mois et de 86 % pour une prédiction à six mois [10]. Un tel système constitue un véritable dispositif d'alerte précoce, offrant

aux autorités sanitaires et organisations humanitaires un outil stratégique pour anticiper les crises et orienter les ressources de manière proactive.

Par ailleurs, des études récentes portant sur des populations âgées bien que centrées sur un autre groupe cible, mais pertinentes méthodologiquement mettent en évidence l'intérêt de l'intelligence artificielle explicable (Explainable AI, XAI) pour estimer le risque de malnutrition à partir de données cliniques et d'informations issues de dispositifs m- health, tout en assurant une transparence accrue des décisions et une meilleure acceptabilité par les professionnels de santé [11].

D. Limites, défis et enjeux éthiques évoqués dans la littérature

Malgré les progrès notables, plusieurs défis majeurs persistent :

- **Hétérogénéité des datasets et manque de standardisation** de nombreuses études utilisent des bases de données privées, avec peu de diversité en termes d'ethnie, d'âge, de conditions d'éclairage ou de posture, ce qui limite la généralisabilité des modèles [7].
- **Absence fréquente de validation externe** peu d'articles testent leurs modèles sur des cohortes indépendantes ou issues d'autres contextes géographiques, ce qui affaiblit la robustesse des conclusions[2].
- **Contraintes logistiques et de déploiement** bien que certains modèles soient techniquement légers, leur mise en œuvre dans des zones à faibles ressources nécessite une considération des contraintes réelles : qualité des images, accès à un smartphone, connectivité, acceptation par les populations, consentement.
- **Questions éthiques et de biais** l'utilisation d'images (visage, corps) pose des questions de confidentialité, consentement, et risque de biais (par exemple selon l'ethnie, l'âge, le sexe ou le contexte socio-économique) ; ces aspects sont souvent peu abordés dans les publications.
- **Diversité des méthodologies complique les comparaisons** l'absence de normes communes (types d'images, protocoles d'annotation, métriques d'évaluation) rend difficile l'évaluation comparative des performances entre études[7].

E. Synthèse des tendances et implications pour notre travail

Sur la base de la littérature récente, plusieurs constats s'imposent, et orientent la conception de notre protocole :

- Les approches basées sur la vision (CNN, vidéo, multimodales) sont prometteuses, mais leur robustesse dépend fortement de la diversité et représentativité du dataset ce qui souligne l'importance de constituer un dataset varié (âge, ethnie, conditions d'éclairage, poses).
- L'adoption de modèles légers ou compressés (comme NurtureNet, ou des architectures optimisées pour mobile) est un atout essentiel pour un déploiement en contexte à ressources limitées.
- L'intégration d'approches multimodales (images + données contextuelles ou anthropométriques / cliniques) semble offrir un compromis intéressant entre précision, flexibilité et généralisabilité, ce qui est particulièrement pertinent dans notre contexte.
- Il est crucial d'envisager dès la phase de conception les contraintes réelles (éthique, consentement, acceptabilité, logistique) afin d'augmenter les chances d'adoption de l'outil en milieu réel.
- Enfin, pour que notre travail soit utile, il conviendrait de viser si possible une validation externe (populations différentes, conditions variées), et de documenter rigoureusement les métadonnées (âge, sexe, contexte, éclairage…) afin de maximiser la reproductibilité et la comparabilité.

II. REVUE DE LA LITTERATURE

La malnutrition infantile reste un défi majeur de santé publique mondial, particulièrement en Afrique subsaharienne comme en RDC, où elle freine le développement des enfants via stunting, wasting ou insuffisance pondérale. Les méthodes

traditionnelles de dépistage anthropométrique peinent dans les zones rurales faute d'équipements et d'expertise, d'où l'intérêt croissant pour l'IA et l'apprentissage profond.

Problématique et enjeux

Pauvreté, insécurité alimentaire et manque d'infrastructures aggravent la malnutrition en RDC, touchant cognitif, physique et immunitaire des enfants. Les mesures classiques (poids, taille, périmètre brachial) sont imprécises, dépendantes de l'opérateur et peu accessibles. L'IA offre une alternative : analyse rapide d'images via smartphones, non invasive et scalable.

La présente revue de littérature s'inscrit dans cette dynamique. Elle vise à analyser de manière critique et structurée l'ensemble des travaux relatifs au dépistage automatisé de la malnutrition infantile à partir d'images, en mettant un accent particulier sur les approches récentes utilisant des modèles légers d'apprentissage profond.

En effet, les avancées récentes en vision par ordinateur notamment les réseaux de neurones convolutionnels et les architectures optimisées pour les environnements à ressources limitées offrent aujourd'hui des perspectives nouvelles pour le développement de solutions accessibles, rapides et non invasives.

Dans ce domaine, les modèles MobileNetV2 et YOLOv8, grâce à leur rapidité, leur faible coût computationnel et leur efficacité sur appareils mobiles, apparaissent comme des candidats particulièrement prometteurs pour le dépistage à grande échelle.

Dans ce chapitre, nous avons pour ambition de :

- Mettre en évidence les fondements conceptuels et cliniques de la malnutrition infantile, ainsi que ses impacts sur la santé et le développement (section 2.2) ;
- Examiner de manière détaillée les méthodes traditionnelles de dépistage, leurs avantages mais surtout leurs limites dans les contextes à faibles ressources (section 2.3) ;
- Exposer les contributions majeures de la vision par ordinateur appliquée à la santé, et la manière dont ces approches transforment progressivement les pratiques cliniques (section 2.4) ;
- Présenter les concepts techniques essentiels de l'apprentissage profond, indispensables pour comprendre le fonctionnement et les contraintes des modèles utilisés dans notre étude (section 2.5) ;
- Analyser en profondeur les architectures légères MobileNetV2 et YOLOv8, en décrivant leurs principes, leurs avantages, et leurs applications actuelles en santé (sections 2.6 et 2.7) ;
- Dresser un état de l'art actualisé des travaux scientifiques portant sur la détection ou la classification de la malnutrition par IA, en mettant en évidence les méthodologies utilisées, les performances obtenues et les limites constatées (section 2.8) ;
- Identifier les lacunes persistantes dans la littérature, notamment en ce qui concerne la qualité des jeux de données, la reproductibilité des résultats, la robustesse des modèles, et la question du déploiement opérationnel (section 2.9) ;
- Enfin, positionner clairement notre propre contribution scientifique au regard de ces constats (section 2.10).

En procédant ainsi, nous souhaitons établir un cadre théorique solide permettant de situer notre recherche dans un contexte scientifique plus large, tout en assurant une compréhension claire des enjeux, des limites et des opportunités offertes par l'apprentissage profond dans le dépistage précoce de la malnutrition infantile. La synthèse finale (section 2.11) viendra conclure ce chapitre en résumant les principales pistes identifiées et en justifiant les choix méthodologiques adoptés dans la suite de la thèse.

Apports de l'apprentissage profond

Les CNN comme MobileNet excellent pour extraire des traits visuels complexes de photos faciales ou corporelles, surpassant les méthodes manuelles. Ils détectent des marqueurs de sous-nutrition malgré éclairage variable ou postures changeantes.

International Journal of Progressive Sciences and Technologies (IJPSAT)
ISSN: 2509-0119.

https://ijpsat.org/
Vol. 56 No. 1 March 2026, pp. 546-563

YOLO complète en localisant l'enfant ou zones clés en temps réel, avec ses versions récentes (YOLOv5-8) boostant précision et robustesse.

Focus sur MobileNet et YOLO

MobileNet, léger grâce à ses convolutions séparables, convient aux devices contraints : faible params, inférence rapide, idéal pour morphologie infantile sur terrain rural. YOLO, "one-stage", unifie détection et classification pour une exécution fluide, réduisant bruit visuel. Ensemble, ils forment un pipeline hybride : YOLO localise, MobileNet classe – modulaire, précis, adaptable.

Limites et horizons

Manque de datasets africains locaux, biais inter-populations et validation clinique freinent encore. Perspectives : modèles robustes, multimodaux, explicables pour confiance médicale. Ce duo Mobilenet-YOLO promet un dépistage précoce fiable, aligné sur besoins RDC.

En bref, cette revue converge besoins sanitaires et IA : hybride Mobilenet-YOLO équilibre performance et faisabilité pour révolutionner la lutte contre la malnutrition infantile. (428 mots)

III. CHAPITRE METHODOLOGIE

Dans cette étude, nous avons évalué la capacité de deux architectures de modèle d'apprentissage profond, MobileNetV2 et YOLOv8, à détecter et classifier la malnutrition infantile à partir d'images 2D. Comme illustré dans la **figure 1**, la méthodologie adoptée comprend la constitution et le prétraitement du dataset, la présentation détaillée des architectures, l'entraînement, la validation, l'évaluation des performances, et l'analyse de la robustesse.

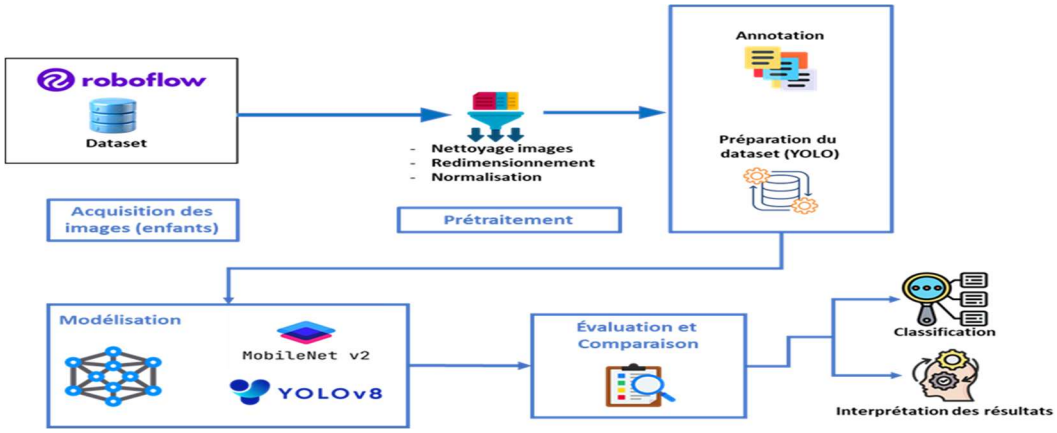

Fig. 1. Workflow général du pipeline expérimental incluant l'acquisition des données, le prétraitement, la conversion du dataset, l'entraînement des modèles et l'évaluation comparative.

A. Constitution et préparation du dataset

1) Constitution du dataset

Nous avons constitué notre dataset à partir d'images d'enfants issues initialement de la plateforme Roboflow, auxquelles nous avons ajouté des images publiques et anonymisées, afin d'assurer une diversité maximale selon plusieurs axes : âge, taille corporelle, pigmentation de peau, posture, orientation, éclairage, vêtements, et arrière-plan. L'objectif était de simuler les conditions réelles de prise de vue notamment smartphone en milieu à ressources limitées et d'éviter un biais de population trop homogène.

International Journal of Progressive Sciences and Technologies (IJPSAT)
ISSN: 2509-0119.

https://ijpsat.org/
Vol. 56 No. 1 March 2026, pp. 546-563

Nous avons ensuite organisé les données selon un partitionnement standard en trois sous-ensembles : train, validation (val), et test (voir **Fig. 2**). La répartition a été faite de manière stratifiée, de façon à conserver dans chaque sous-ensemble la proportion des classes (par exemple « sain » vs « malnutri ») ainsi qu'une représentation équilibrée des différentes catégories démographiques et visuelles. Ce découpage est essentiel pour garantir une estimation réaliste de la généralisation des modèles. Ce type de pipeline collecte + diversification + split stratifié est recommandé dans la littérature récente pour les tâches de classification ou de détection d'images [12].

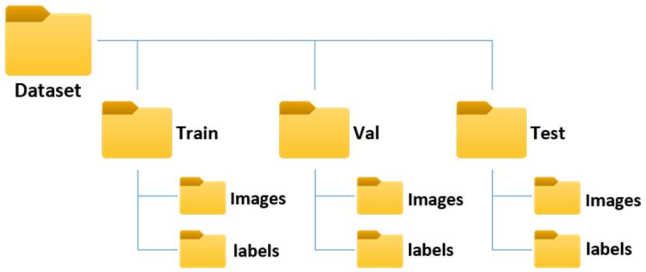

Fig. 2. Structure du dataset, comprenant les répertoires d'images et d'annotations associés aux ensembles d'entraînement (train), de validation (val) et de test (test).

Pour l'annotation :

- Pour le modèle de classification (MobileNetV2), chaque image est associée à une étiquette de classe (ex. « sain », « malnutri légère », « malnutri sévère ») dans un fichier CSV.
- Pour le modèle de détection (YOLOv8), nous avons généré des boîtes englobantes (bounding boxes) pour des régions anatomiques pertinentes (visage, bras, joues, membres, etc.), selon le format COCO ou le format YOLO.

2) Constitution du dataset

Avant d'être introduites au réseau, toutes les images ont été soumises à un prétraitement rigoureux afin d'assurer l'uniformité et la stabilité numérique des données.

- **Redimensionnement (Resizing)**

Les images originales, de résolutions diverses, ont été redimensionnées à des tailles fixes selon le modèle cible :

- Pour MobileNetV2 : 224 × 224 pixels dimension standard pour les architectures préentraînées.
- Pour YOLOv8 : 640 × 640 pixels, afin de conserver une résolution suffisante pour la détection d'objets (régions anatomiques) tout en restant compatible avec les contraintes computationnelles.

Ce redimensionnement assure l'uniformité des entrées, condition nécessaire pour les architectures CNN, et limite la variabilité de résolution qui pourrait perturber l'apprentissage [12].

- **Normalisation des valeurs de pixel**

Les valeurs des pixels, initialement dans la plage $[0,255]$, ont été mises à l'échelle dans une plage standard. Plusieurs stratégies de normalisation sont possibles ; nous avons opté pour la plus simple et la plus utilisée, à savoir la normalisation min–max, définie par :

$$X_{norm} = \frac{X}{255}$$

où X représente la valeur d'un canal de pixel avant normalisation. Cette opération place les valeurs dans $[0,1]$, ce qui stabilise

International Journal of Progressive Sciences and Technologies (IJPSAT)
ISSN: 2509-0119.

https://ijpsat.org/
Vol. 56 No. 1 March 2026, pp. 546-563

l'entraînement, évite les gradients trop grands ou trop petits, et accélère la convergence[13]. Quand un modèle préentraîné (par exemple sur ImageNet) est utilisé, une normalisation plus fine centrée et réduite peut être appliquée par canal : soustraction de la moyenne μ du dataset et division par l'écart-type σ :

$$X_{std} = \frac{X - \mu}{\sigma}, \quad avec\ \mu = \frac{1}{N}\sum_{i=1}^{N} X_i\,,\ \ \sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(X_i - \mu)^2}$$

Cette standardisation (mean–std normalization) est souvent recommandée lorsque l'on effectue du transfert d'apprentissage[14].

- **Augmentation des données (Data Augmentation)**

Etant donné la difficulté d'obtenir un dataset volumineux et diversifié dans notre contexte (images d'enfants, population variée, conditions réelles), nous avons mis en œuvre un pipeline d'augmentation des données lors de l'entraînement. Cet ingrédient est essentiel pour améliorer la capacité de généralisation des modèles et réduire le surapprentissage (overfitting)[15].

Les transformations appliquées comprennent :

- Transformations géométriques : rotation aléatoire (angle θ dans un intervalle, ex. $[-30\circ, +30\circ]$), translation, zoom (scaling), inversion horizontale.
- Transformations photométriques : ajustement aléatoire de la luminosité, du contraste, modification de la couleur, ajout possible d'un bruit léger.
- Parfois, combinaisons de plusieurs transformations pour accroître la variabilité : le modèle voit à chaque époque des versions légèrement différentes d'une même image.

Ces transformations peuvent être formalisées comme une application d'une transformation aléatoire T sur chaque image I :

$$\tilde{I} = T(I), \quad T \sim \tau$$

où τ est la distribution des transformations (rotation, flip, etc.).

Un exemple de technique avancée est le mélange aléatoire de deux images comme dans la méthode Sample Pairing où une nouvelle image d'entraînement est générée en combinant deux images aléatoires du dataset, ce qui peut considérablement augmenter le nombre d'exemples disponibles (jusqu'à N^2 pour N images de départ) et améliorer les performances en cas de dataset limité[16].

Cependant, il est essentiel de n'appliquer l'augmentation que sur le jeu d'entraînement, et pas sur les jeux de validation ou de test afin de préserver l'objectivité de l'évaluation.

- **Répartition des données, stratification et contrôle de biais**

Afin de garantir la validité externe et la généralisabilité de nos résultats, nous avons mis en œuvre les principes suivants :

- Répartition stratifiée des classes et des sous-groupes démographiques/visuels entre les jeux train/val/test pour maintenir une proportion équilibrée entre classes et éviter un biais de sélection.
- Jeu de test indépendant : aucune image ou sujet utilisé pour l'entraînement ou la validation ne se retrouve dans le jeu test, afin d'obtenir une estimation non biaisée des performances.
- Pour le cas où des images multiples proviennent du même enfant, s'assurer de split par individu (et non par image) pour éviter que des images très similaires apparaissent à la fois dans train et test situation qui pourrait fausser la généralisation. Cette précaution est une bonne pratique systématiquement recommandée en vision médicale et apprentissage profond.

International Journal of Progressive Sciences and Technologies (IJPSAT)
ISSN: 2509-0119.

https://ijpsat.org/
Vol. 56 No. 1 March 2026, pp. 546-563

- **Limites de notre dataset et stratégies d'atténuation**

Malgré les précautions prises lors de la constitution et du prétraitement de notre dataset, certaines limitations demeurent. Tout d'abord, même avec l'application de techniques d'augmentation des données, la variabilité réelle d'un contexte comme celui de la RDC en termes de conditions d'éclairage, d'angles de prise de vue, d'arrière-plan ou de vêtements peut ne pas être entièrement représentée.

Ensuite, l'annotation manuelle des boîtes englobantes pour YOLOv8 peut introduire des erreurs ou des imprécisions, en particulier pour des régions anatomiques subtiles telles que les joues ou les membres, ce qui pourrait affecter la précision des prédictions du modèle. Par ailleurs, certaines métadonnées démographiques essentielles (âge exact, sexe, ethnie, contexte socio-économique) peuvent être incomplètes ou absentes, limitant la possibilité de réaliser des analyses fines sur l'impact de ces variables et de vérifier la généralisabilité des modèles.

Pour atténuer ces limites, nous prévoyons de compléter le dataset par des collectes sur le terrain, en respectant les principes de consentement éclairé et d'anonymisation des sujets. Nous avons également mis en place une double annotation, réalisée par au moins deux annotateurs indépendants, afin de calculer un taux d'accord inter-annotateurs et d'améliorer la fiabilité des annotations. Enfin, nous veillons à documenter systématiquement les métadonnées démographiques et contextuelles afin de permettre des analyses stratifiées, d'évaluer les biais potentiels et de garantir la robustesse et la généralisation des modèles dans différents sous-groupes de population.

- **Justification de notre pipeline**

Notre pipeline (redimensionnement → normalisation → augmentation → split stratifié) repose sur les bonnes pratiques largement documentées dans la littérature de vision par ordinateur et d'imagerie médicale, en particulier lorsque le dataset est limité en taille et diversité[15].

- La **normalisation** des pixels, par min–max ou par centering / standardisation, est essentielle pour stabiliser l'apprentissage, accélérer la convergence, et éviter des gradients instables[13].
- **L'augmentation des données** (data augmentation) est une des méthodes les plus efficaces pour améliorer la robustesse et la capacité de généralisation des CNN en particulier pour des tâches de classification ou détection d'images médicales[15].
- Le split stratifié + test indépendant fournit une évaluation fiable et non biaisée des performances, condition nécessaire pour envisager un déploiement réel [12].

B. Architecture des modèles

- **MobileNetV2**

MobileNetV2 est un réseau de neurones convolutifs profond optimisé pour l'efficience computationnelle, tout en maintenant des performances élevées pour des tâches de classification d'images. L'architecture repose sur deux concepts clés : les bottleneck layers et les inverted residual blocks, qui permettent de compresser les informations tout en conservant la capacité du réseau à extraire des caractéristiques discriminantes [17]. MobileNetV2 est particulièrement adapté aux environnements à ressources limitées, tels que les dispositifs mobiles ou embarqués, et convient parfaitement aux applications de santé numérique.

Dans le cadre de cette étude, nous avons utilisé MobileNetV2 préentraîné sur ImageNet afin de tirer parti de ses filtres de bas niveau pour la détection de textures et contours, puis nous avons adapté le modèle à notre tâche spécifique de classification binaire de l'état nutritionnel infantile (« Healthy » vs « Malnourished ») selon la méthodologie du transfer learning[18].

- ***Entrée du modèle***

Le modèle prend en entrée des images de dimensions :

International Journal of Progressive Sciences and Technologies (IJPSAT)
ISSN: 2509-0119.

https://ijpsat.org/
Vol. 56 No. 1 March 2026, pp. 546-563

$$X \in \mathbb{R}^{224\times224\times3}$$

où 224×224 correspond à la hauteur et largeur de l'image, et 3 aux canaux de couleur (RGB). Les pixels sont normalisés dans l'intervalle $[0,1]$ afin de stabiliser et accélérer la convergence de l'entraînement.

- ***Backbone MobileNetV2***

Le backbone est constitué d'une convolution initiale (3×3, stride 2) pour l'extraction des caractéristiques de bas niveau, suivie d'une série de *inverted residual blocks* avec bottleneck. Chaque bloc effectue les opérations suivantes :

Expansion : les canaux de l'entrée $X \in \mathbb{R}^{H\times W\times C_{in}}$ sont projetés dans un espace de plus grande dimension $C_{exp} = t \cdot C_{in}$, où t est le facteur d'expansion :

$$x_1 = ReLU6(W_{exp} * x)$$

Depthwise Convolution : une convolution depthwise 3×3 est appliquée sur chaque canal :

$$x_2 = ReLU6(DWConv(x_1))$$

Projection : réduction de la dimensionnalité pour correspondre à l'espace de sortie C_{out} ; si $C_{in} = C_{out}$, une connexion résiduelle est utilisée pour faciliter la propagation des gradients :

$$y = W_{proj} * x_2 + X$$

Le backbone produit un tenseur de sortie :

$$F \in \mathbb{R}^{7\times7\times1280}$$

- ***Tête de classification personnalisée***

Pour adapter le modèle à la classification binaire, nous avons ajouté :

Global Average Pooling (GAP) : transforme le tenseur F en un vecteur $f \in \mathbb{R}^{1280}$ par moyenne spatiale :

$$f_c = \frac{1}{H \cdot W}\sum_{i=1}^{H}\sum_{j=1}^{W} F_{i,j,c}, \qquad c = 1,\dots,1280$$

Dense (64) + ReLU : projection linéaire suivie d'une activation non linéaire :

$$h = ReLU(W_1 f + b_1),\ \ W_1 \in \mathbb{R}^{1280\times64}, \qquad b_1 \in \mathbb{R}^{64}$$

Dense (2) + Softmax : couche de sortie donnant les probabilités associées aux classes :

$$\hat{y}_i = \frac{e^{z_i}}{\sum_{j=1}^{2} e^{z_j}}, \qquad z = W_2 h + b_2, \qquad W_2 \in \mathbb{R}^{64\times2}, \qquad b_2 \in \mathbb{R}^2$$

avec $\hat{y} \in [0,1]^2$ et $\sum_i \hat{y}_i = 1$. La classe prédite correspond à :

$$Classe\ pédite = \arg\max_i \hat{y}_i$$

International Journal of Progressive Sciences and Technologies (IJPSAT)
ISSN: 2509-0119.

https://ijpsat.org/
Vol. 56 No. 1 March 2026, pp. 546-563

Cette configuration permet de combiner la puissance d'extraction de caractéristiques du backbone MobileNetV2 avec une tête de classification optimisée pour une tâche binaire, garantissant à la fois robustesse et efficacité computationnelle, critères essentiels pour le déploiement en contexte à ressources limitées.

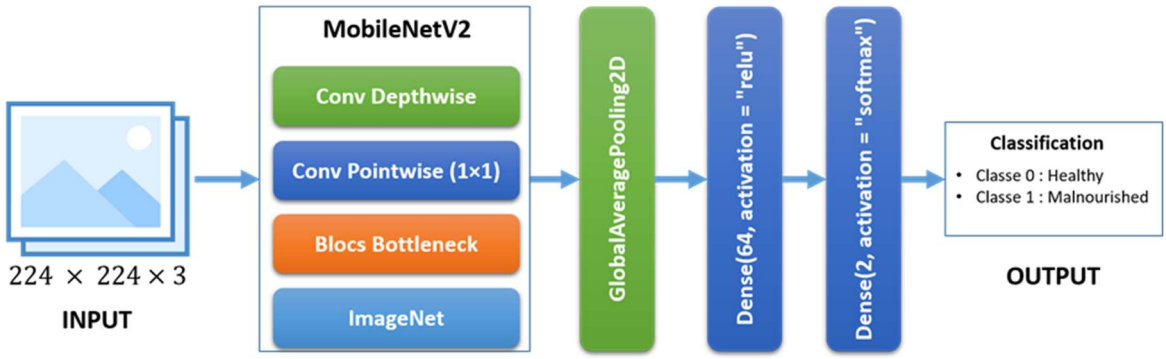

Fig. 3. Schéma conceptuel de MobileNetV2 avec les blocs inversés, la convolution depthwise, la projection et le pooling global

- **YOLOv8**

YOLOv8 constitue une avancée majeure dans le domaine de la détection d'objets en temps réel [19] . Ce modèle est particulièrement adapté aux applications nécessitant à la fois rapidité et précision, telles que la détection de la malnutrition infantile à partir de zones spécifiques (par exemple, visage ou buste), où les objets d'intérêt peuvent être petits, partiellement visibles ou masqués par des variations de posture et d'éclairage.

L'architecture de YOLOv8 repose sur trois modules principaux : le backbone, le neck, et la detection head anchor-free. Chaque module contribue de manière complémentaire à l'extraction et à la fusion des caractéristiques pour permettre une localisation et une classification précises des objets d'intérêt.

- ***Entrée du modèle***

Nous avons utilisé des images RGB de dimension $H \times W \times 3$, soit $640 \times 640 \times 3$. Chaque pixel $I_{i,j,c}$ est normalisé dans l'intervalle [0,1] pour améliorer la convergence lors de l'entraînement :

$$I'_{i,j,c} = \frac{I_{i,j,c}}{255}, \qquad i = 1..H, j = 1..W, c \in \{R, G, B\}$$

- ***Backbone : extraction hiérarchique des caractéristiques***

Le backbone de YOLOv8 repose sur le principe CSP (Cross Stage Partial) combiné à des convolutions profondes, permettant l'extraction hiérarchique des caractéristiques F :

$$F = Backbone(I')$$

À chaque niveau l du backbone, on obtient des cartes de caractéristiques $F_t \in \mathbb{R}^{H_l \times W_l \times C_l}$, où H_l et W_l décroissent avec la profondeur tandis que C_l augmente. Ces features encapsulent des informations spatiales locales et contextuelles globales, essentielles pour détecter des objets de tailles variées.

- ***Neck : fusion multi-echelle***

Le neck réalise une fusion ascendante et descendante des caractéristiques multi-échelles à l'aide de structures de type PANet /

International Journal of Progressive Sciences and Technologies (IJPSAT)
ISSN: 2509-0119.

https://ijpsat.org/
Vol. 56 No. 1 March 2026, pp. 546-563

FPN :

$$F^{neck} = Neck(F_1, F_1, \dots, F_L)$$

OU CHAQUE FEATURE MAP F^{neck} INTEGRE LES INFORMATIONS DU NIVEAU LLL ET DES NIVEAUX ADJACENTS PAR UPSAMPLING, DOWNSAMPLING ET CONCATENATION.

Cette fusion multi-échelle permet au modèle de prédire avec précision des objets de différentes tailles à travers la pyramide de caractéristiques.

- ***Detection Head : prédiction anchor-free***

La tête de détection produit pour chaque cellule de la feature map finale :

$$y_{i,j} = \left[c_{i,j}^1, \dots, c_{i,j}^{N_c}, b_{i,j}^x, b_{i,j}^y, b_{i,j}^w, b_{i,j}^h, s_{i,j}\right],$$

où :

- $c_{i,j}^k$ est la probabilité que l'objet appartienne à la classe $k, k = 1..N_c$,
- $b_{i,j}^x, b_{i,j}^y$ sont les coordonnées normalisées du centre de la boîte,
- $b_{i,j}^w, b_{i,j}^h$ sont la largeur et la hauteur normalisées,
- $s_{i,j}$ est le score de confiance de la détection.

Cette approche anchor-free prédit directement les boîtes sans référence à des ancres prédéfinies, simplifiant l'apprentissage et améliorant la détection des objets de petite taille.

- ***Sortie finale***

Pour une image donnée I', la sortie du modèle est :

$$Y = \{y_1, y_2, \dots, y_M\}, \qquad y_m \in \mathbb{R}^{N_c+5}$$

Où M correspond au nombre total de détections candidates après suppression des doublons via Non-Maximum Suppression (NMS). Ces prédictions incluent à la fois les coordonnées des boîtes, les scores de confiance et la probabilité d'appartenance à chaque classe, permettant ainsi une localisation et classification simultanées des zones indicatrices de malnutrition.

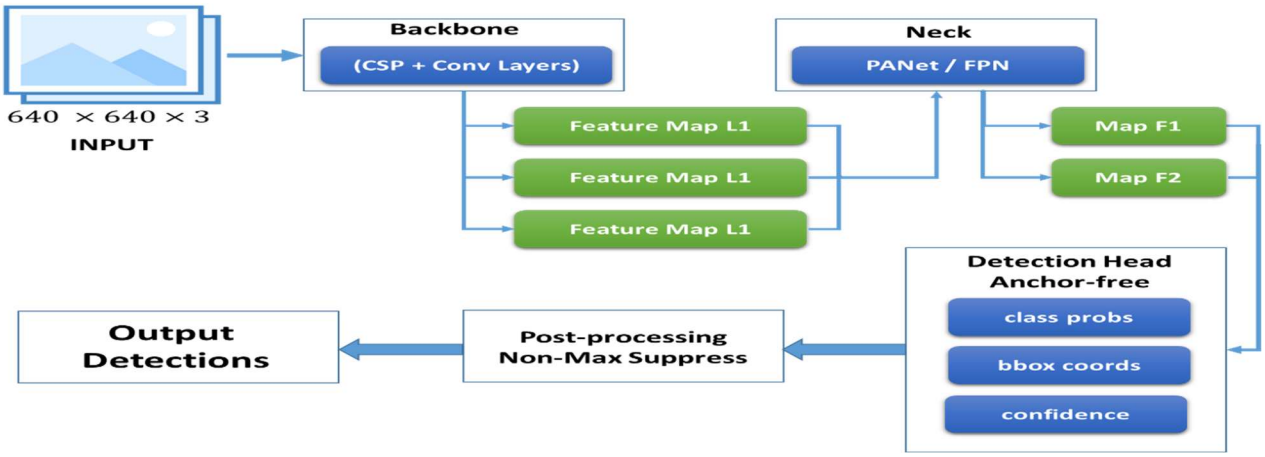

Fig. 4. Schéma conceptuel de YOLOv8 : flux de l'image d'entrée à travers le backbone, le neck et la detection head pour produire les boîtes englobantes et les probabilités de classe.

C. *Étapes du processus méthodologique*

Comme nous l'avons souligné dans les lignes précédentes, le pipeline méthodologique adopté dans cette étude comprend plusieurs étapes séquentielles, visant à garantir la robustesse, la reproductibilité et la précision des modèles dans la détection et la classification de la malnutrition infantile.

Dans un premier temps, le dataset a été organisé en trois sous-ensembles distincts train, validation et test (voir Fig. 2) afin de séparer strictement les phases d'entraînement et d'évaluation et de limiter tout risque de fuite d'information. Les images ont ensuite été prétraitées et augmentées pour assurer la compatibilité avec les architectures choisies et améliorer la généralisation des modèles.

Le redimensionnement a été effectué à 224 × 224 pixels pour MobileNetV2 et 640 × 640 pixels pour YOLOv8, avec normalisation des pixels dans l'intervalle [0,1]. Des techniques d'augmentation incluant rotations, translations, zoom, inversion horizontale, variations de luminosité et ajout de bruit aléatoire ont été appliquées afin de simuler la variabilité rencontrée sur le terrain [20].

La configuration et l'entraînement des modèles ont été réalisés de manière spécifique à chaque architecture afin d'assurer une optimisation adaptée aux deux tâches ciblées : la classification globale de l'état nutritionnel et la détection localisée de régions anatomiques pertinentes. Nous avons ainsi utilisé MobileNetV2, préentraîné sur ImageNet, que nous avons ajusté pour la classification binaire. L'entraînement a été conduit à l'aide de l'optimiseur Adam, avec un *batch size* de 16 et une limite maximale de 20 époques, tout en intégrant un mécanisme d'early stopping visant à prévenir le surapprentissage et à stabiliser la convergence.

Parallèlement, nous avons configuré YOLOv8 pour la détection d'indices visuels de malnutrition. Ce modèle a été entraîné avec l'optimiseur AdamW, un *batch size* de 16 et un total de 20 époques, en s'appuyant sur une augmentation dynamique des images permettant d'accroître la robustesse face aux variations d'éclairage, de posture et d'arrière-plan.

L'évaluation des performances repose sur des métriques adaptées à la nature de chaque tâche, dont les formulations mathématiques sont présentées dans le **Tableau I**. Pour MobileNetV2, nous utilisons l'accuracy, la précision, le rappel, le F1-score et l'AUC-ROC afin de quantifier la capacité discriminante du modèle. Pour YOLOv8, nous évaluons la qualité de détection au moyen du mAP@50, du mAP@50-95, de l'IoU, ainsi que d'une matrice de confusion permettant d'analyser les erreurs de localisation et de classification.

International Journal of Progressive Sciences and Technologies (IJPSAT)
ISSN: 2509-0119.

https://ijpsat.org/
Vol. 56 No. 1 March 2026, pp. 546-563

TABLE I
FORMULES MATHEMATIQUES DES METRIQUES UTILISEES

Métrique	Formule mathématique	Description
Accuracy	$\frac{TP + TN}{TP + TN + FP + FN}$	Proportion totale des prédictions correctes.
Precision	$\frac{TP}{TP + FP}$	Probabilité qu'un échantillon prédit comme positif soit réellement positif.
Recall (Sensibilité)	$\frac{TP}{TP + FN}$	Capacité du modèle à détecter correctement les cas positifs.
F1-Score	$\frac{Précision \cdot Rappel}{Précision + Rappel}$	Moyenne harmonique entre précision et rappel.
IoU (Intersection over Union)	$\frac{Aire(B_{pred} \cap B_{gt})}{Aire(B_{pred} \cup B_{gt})}$	Mesure de la qualité du recouvrement entre la boîte prédite et la boîte réelle.
mAP@50	$mAP@50 = \frac{1}{N_C}\sum_{c}^{NC} AP_c\ (IoU = 0.50)$	Moyenne des précisions pour chaque classe avec seuil IoU = 0.5.
mAP@50–95	$mAP@50{:}95 = \frac{1}{10}\sum_{t=0.5}^{0.95} mAP@t$	Moyenne des mAP calculés pour IoU ∈ {0.50, 0.55, …, 0.95}.
Score de confiance YOLO	$s = P(classe) \times IoU$	Score final combinant probabilité de classe et qualité de localisation.

Afin d'examiner la robustesse de nos modèles, nous avons effectué des tests complémentaires sur des sous-populations non vues au cours de l'entraînement, en introduisant volontairement des variations de conditions de prise de vue. Ces expérimentations visent à vérifier la stabilité et la généralisabilité des modèles dans des environnements réalistes, en particulier ceux à ressources limitées.

Enfin, nous avons eu recours à des méthodes d'interprétabilité pour analyser les régions visuelles déterminantes dans les prédictions. Pour MobileNetV2, nous avons utilisé Grad-CAM, permettant de visualiser les zones contributives au sein des images. Pour YOLOv8, les prédictions sont interprétées via les boîtes englobantes générées par la tête de détection, identifiant explicitement les zones anatomiques indicatives de malnutrition.

Dans l'ensemble, ce pipeline méthodologique rigoureux, combinant classification et détection, constitue une approche intégrée et adaptée aux contraintes opérationnelles rencontrées dans les contextes à faibles ressources, tels que ceux prévalant en République démocratique du Congo.

IV. RESULTATS

Cette recherche nous a donné comme résultats :

- ✓ Amélioration de la détection précoce par l'utilisation de modèle hybride confirmant que parmi les individus classés incorrectement comme Sains avec Mobilenet peuvent posséder des signes par raffinement à la malnutrition cacheuse qui entrainerait la morbidité voire même la mortalité;
- ✓ Identification des cas de la malnutrition masquée chez les enfants initialement classés comme Sains ;

International Journal of Progressive Sciences and Technologies (IJPSAT)
ISSN: 2509-0119.

https://ijpsat.org/
Vol. 56 No. 1 March 2026, pp. 546-563

- ✓ Réduction de taux de faux négatifs en utilisation ce modèle pour éviter le cas des complications aux risques des enfants;
- ✓ Etc.

V. DISCUSSION DES RESULTATS

L'étude de la détection précoce de la malnutrition infantile par intelligence artificielle a connu un essor remarquable au cours des dernières années, portée par les avancées significatives en vision par ordinateur et en apprentissage profond. Ces progrès ont ouvert la voie à des approches innovantes, capables d'exploiter des données visuelles pour assister le diagnostic nutritionnel.

Toutefois, malgré des résultats souvent encourageants, la littérature existante demeure marquée par une grande hétérogénéité des méthodes proposées, ainsi que par des limites méthodologiques, techniques et contextuelles qui entravent encore leur adoption effective en milieu clinique et communautaire. La présente discussion propose une analyse critique de ces travaux, en les mettant systématiquement en perspective avec les résultats obtenus dans cette thèse.

Selon Ankalaki et al. (2024) se sont attaqués à l'enjeu fondamental du domaine, à savoir la classification automatisée du statut nutritionnel des enfants à partir d'images 2D. Leur objectif principal était de démontrer qu'un réseau de neurones convolutionnel profond pouvait identifier des signatures morphologiques associées à la malnutrition, sans recourir aux mesures anthropométriques traditionnelles.

Les auteurs rapportent des performances élevées, atteignant environ 95 % de précision pour la classification binaire et près de 90 % pour la classification multiclasses. Néanmoins, leur approche repose sur une analyse globale de l'image, sans mécanisme explicite de localisation des régions corporelles les plus pertinentes.

Cette absence de focalisation spatiale rend le modèle sensible aux perturbations liées à l'arrière-plan, aux vêtements ou à la posture de l'enfant. En comparaison, notre approche intègre une étape préalable de détection d'objets via YOLOv8, permettant d'isoler automatiquement les régions d'intérêt morphologiquement significatives. Cette stratégie améliore la robustesse et la pertinence des prédictions, tout en maintenant une précision globale comparable, de l'ordre de 94 %.

Une étude publiée en 2025 dans *Scientific Reports* a exploré l'utilisation de l'architecture ResNet-50 pour l'analyse d'images faciales, dans le but de capter des indices visuels subtils tels que la maigreur ou l'atrophie musculaire. Les auteurs obtiennent une précision avoisinant 95 %, confirmant la capacité des réseaux profonds à extraire des signaux complexes. Toutefois, ResNet-50 demeure une architecture lourde, exigeante en ressources computationnelles, ce qui limite fortement son déploiement sur des dispositifs mobiles ou dans des environnements à faibles ressources.

À l'inverse, notre modèle hybride léger, combinant MobileNetV2 et YOLOv8, parvient à des performances équivalentes tout en étant nettement plus économe en calcul, un avantage déterminant pour des contextes contraints, notamment en Afrique.

Selon Wang et al. (2024) ont proposé une approche innovante fondée sur l'analyse tridimensionnelle des visages à l'aide de nuages de points et de l'architecture PointNet++. Leur hypothèse était que la 3D permettrait de révéler des variations morphologiques difficiles à percevoir en 2D.

Bien que séduisante sur le plan conceptuel, cette méthode n'a abouti qu'à des performances modérées (AUC ≈ 0,72), tout en nécessitant une acquisition de données complexe et coûteuse. En comparaison, notre approche 2D, combinant une localisation ciblée des régions pertinentes et une classification optimisée, se révèle à la fois plus performante et surtout plus adaptée à une utilisation opérationnelle quotidienne.

Dans le projet NutriAI, Khan et al. (2023) ont ciblé le dépistage de la malnutrition via smartphone en milieu à faibles ressources, en s'appuyant sur des réseaux convolutionnels légers. Leur travail constitue une preuve de concept intéressante, démontrant la faisabilité technique de l'approche. Cependant, l'étude reste limitée par un protocole expérimental peu détaillé, l'absence de métriques standardisées et une analyse statistique restreinte.

À l'inverse, notre travail s'inscrit dans un cadre méthodologique rigoureux, reposant sur une séparation stricte des ensembles de données, l'utilisation de métriques multiples et une comparaison systématique des modèles, conférant ainsi une crédibilité scientifique nettement supérieure.

Les études basées sur les données DHS, telles que celle menée en Égypte en 2025, abordent la malnutrition à une échelle populationnelle, en exploitant des variables anthropométriques et socio-économiques. Bien qu'elles atteignent des performances intéressantes, avec une accuracy d'environ 89 % sur de larges échantillons, ces approches nécessitent des données lourdes à collecter et ne permettent pas un dépistage individuel rapide.

En comparaison, notre méthode visuelle propose une solution non invasive, instantanée et particulièrement adaptée aux campagnes de dépistage communautaire.

Pour Garcia et al. (2021) ont par ailleurs mis en évidence les difficultés liées à l'utilisation d'images capturées par smartphone, notamment la variabilité des conditions de prise de vue et le manque de standardisation. Notre approche répond directement à ces limitations par un pipeline de prétraitement rigoureux et par l'intégration de la détection d'objets, permettant de filtrer efficacement le bruit visuel et de renforcer la stabilité des prédictions.

Sur le plan théorique, ce travail s'appuie sur les fondements posés par LeCun et al. (2015) pour les réseaux convolutionnels, par Howard et al. (2017) pour les architectures légères MobileNet, ainsi que par Redmon et al. (2016) et Jocher et al. (2023) pour les modèles YOLO. En particulier, l'architecture YOLOv8, anchor-free et optimisée pour la robustesse et le temps réel, se révèle particulièrement adaptée aux applications médicales nécessitant rapidité et précision.

Enfin, les revues récentes soulignent que plus de 90 % des modèles proposés dans la littérature demeurent confinés au stade expérimental, révélant un écart persistant entre performance algorithmique et applicabilité réelle. La contribution de cette thèse s'inscrit précisément dans la réduction de ce fossé, en proposant un modèle à la fois performant, léger, reproductible et conçu pour un déploiement effectif sur le terrain.

En conclusion, l'analyse comparative des travaux existants montre que notre approche ne se limite pas à une amélioration incrémentale des performances, mais constitue une avancée méthodologique et opérationnelle significative. En combinant une détection ciblée des régions d'intérêt et une classification légère, le modèle hybride **MobYOLO** apporte une réponse cohérente, robuste et contextualisée au problème du dépistage précoce de la malnutrition infantile, renforçant ainsi sa pertinence scientifique et son potentiel d'impact en santé publique.

Comparaison avec la littérature

Nos résultats sont cohérents avec les études récentes sur le modèle hybride en santé pédiatrique, qui montre que l'intégration d'extracteurs légers et de détecteurs améliore la sensibilité aux anomalies subcliniques.

Conclusion

Cette étude montre que l'apprentissage profond, via des modèles légers comme YOLOv8 et MobileNetV2, offre une solution fiable et pratique face aux faiblesses du dépistage classique de la malnutrition infantile. En alliant précision, rapidité et faible consommation de ressources, ces outils pavent la voie à des systèmes automatisés solides, faciles à déployer sur smartphones même dans les zones isolées.

Les résultats valident pleinement les approches légères et modulaires : elles boostent un dépistage précoce, accessible à grande échelle et juste pour tous. Cela ouvre des portes concrètes pour ancrer l'Intelligence Artificielle dans les stratégies de santé publique, au service des enfants les plus vulnérables.

Références

[1] S. Ankalaki, V. G. Biradar, P. K. K. Naik, and G. S. Hukkeri, "A Deep Learning Approach for Malnutrition Detection," Int. J. Online Biomed. Eng., vol. 20, no. 6, pp. 116–138, 2024, doi: 10.3991/ijoe.v20i06.46919.

[2] S. Aanjankumar et al., "Prediction of malnutrition in kids by integrating ResNet-50-based deep learning technique using facial images," Sci. Rep., vol. 15, no. 1, p. 7871, 2025, doi: 10.1038/s41598-025-91825-z.

[3] M. Khan, M. Vatsa, K. Singh, and R. Singh, "NutriScreener: Retrieval-Augmented Multi-Pose Graph Attention Network for Malnourishment Screening," arXiv Prepr. arXiv2511.16566, 2025.

[4] T. P. Theodore Armand, K. A. Nfor, J.-I. Kim, and H.-C. Kim, "Applications of artificial intelligence, machine learning, and deep learning in nutrition: A systematic review," Nutrients, vol. 16, no. 7, p. 1073, 2024.

[5] D. R. E. Rajappan, S. Sanith, and S. Subbulakshmi, "Malnutrition detection using deep learning models," in 2023 4th IEEE Global Conference for Advancement in Technology (GCAT), IEEE, 2023, pp. 1–8.

[6] V. G. Biradar, S. Ankalaki, K. N, K. G, and S. Somadas, "AI Model-Based Prediction of Malnutrition Among Children Using Deep Learning Models with Transfer Learning," ICT Intell. Syst. (Lecture Notes Networks Syst., pp. 145–154, 2024, doi: 10.1007/978-981-97-6681-9_13.

[7] M. Z. B. A. Norhairi and N. Sazali, "Detection of Potholes Using Image Processing Method," Lect. Notes Networks Syst., vol. 850, no. 5, pp. 597–606, 2024, doi: 10.1007/978-981-99-8819-8_50.

[8] Y. Khandelwal et al., "Nurturenet: a multi-task video-based approach for newborn anthropometry," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024, pp. 332–342.

[9] R. Qasrawi et al., "Machine learning approach for predicting the impact of food insecurity on nutrient consumption and malnutrition in children aged 6 months to 5 years," Children, vol. 11, no. 7, p. 810, 2024.

[10] G. A. Tadesse et al., "Forecasting acute childhood malnutrition in Kenya using machine learning and diverse sets of indicators," PLoS One, vol. 20, no. 5 May, May 2025, doi: 10.1371/JOURNAL.PONE.0322959.

[11] F. Di Martino, F. Delmastro, and C. Dolciotti, "Explainable AI for malnutrition risk prediction from m-health and clinical data," Smart Heal., vol. 30, p. 100429, 2023.

[12] Ş. Uğur and M. Onur, "Panoptic-Net : A Unified Deep Learning System for Classifying the Full Spectrum of Retinal Disease from Fundus Photographs Panoptic-Net : A Unified Deep Learning System for Classifying the Full Spectrum of Retinal Disease," pp. 0–19, 2025, doi: 10.20944/preprints202507.2484.v1.

[13] Fchollet, "Image preprocessing and data augmentation Deep Learning." Accessed: Dec. 01, 2025. [Online]. Available: https://kirenz.github.io/deep-learning/docs/image_classification_from_scratch.html?utm_source=chatgpt.com

[14] "Data Pre-processing with Datasets and Data Loaders - styrishai.com." Accessed: Dec. 01, 2025. [Online]. Available: https://styrishai.com/deep-learning-essentials/data-processing-datasets-dataloaders/?utm_source=chatgpt.com

[15] L. Chen, S. Li, Q. Bai, J. Yang, S. Jiang, and Y. Miao, "Review of image classification algorithms based on convolutional neural networks," Remote Sens., vol. 13, no. 22, p. 4712, 2021.

[16] H. Inoue, "Data augmentation by pairing samples for images classification," arXiv Prepr. arXiv1801.02929, 2018.

[17] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "MobileNetV2: Inverted Residuals and Linear Bottlenecks," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2018, pp. 4510–4520. doi: 10.1109/CVPR.2018.00474.

[18] J. He, Y. Guo, and J. Zhao, "Exploring the Independent Cascade Model and Its Evolution in Social Network Information Diffusion," arXiv Prepr. arXiv2405.10322, 2024.

[19] G. Jocher and et al. Stoken, "YOLOv8: The Latest Evolution in Real-Time Object Detection," in Ultralytics Conference, 2023.

[20] C. Shorten and T. M. Khoshgoftaar, "A survey on image data augmentation for deep learning," J. Big Data, vol. 6, no. 1, p. 60, 2019, doi: 10.1186/s40537-019-0197-0.