

Eagle AgriP : système hybride d'aide au diagnostic précoce des maladies foliaires par fusion intermédiaire images et données agroclimatiques : Cas du Manioc en République démocratique du Congo

KAMGANG BERTIN Simplicie¹, BAZA NTOYE Tostaos⁴, KALANGA KABASELE Franchise², SUMAILI KIBASHA Grace⁴ et ASSANI KAHUYA Bob-Martin⁴, MUKWA FAMA TONGO Lyna et ILUNGA MBUYAMBA Elisée³

¹ Ecole de Télécommunications et Télédétection Spatiale (ETS/UPN) Kinshasa, République Démocratique du Congo

² Institut National d'Étude et Recherche Agronomiques (INERA), Kinshasa, République Démocratique du Congo

³ Institut Supérieur Techniques Appliquées (ISTA-KIN)

⁴ Institut Supérieur Pédagogique Technique (ISPT-KIN)

Résumé

Les maladies du manioc, notamment la mosaïque africaine du manioc (CMD) et la maladie de la striure brune (CBSD), constituent une contrainte phytosanitaire importante pour la production vivrière en République démocratique du Congo (RDC). Le diagnostic conventionnel dépend de l'observation humaine et, dans les cas complexes, d'analyses spécialisées dont l'accès et le délai peuvent limiter la couverture territoriale. Cet article propose l'architecture scientifique d'Eagle AgriP, un système d'information phytopathologique associant l'analyse d'images foliaires par apprentissage profond, des variables agroclimatiques et une composante de géolocalisation/SIG. La branche visuelle s'appuie sur MobileNetV2 et l'apprentissage par transfert ; la branche tabulaire représente les conditions environnementales et rejoint la représentation visuelle dans une couche de fusion. Le système vise à produire une classification indicative — plante saine, CMD, CBSD ou cas incertain —, un score de confiance et une information géographique destinée à prioriser la vérification sur le terrain. Un protocole de validation est proposé, avec séparation des données par parcelle ou site, comparaison entre modèle image seul et modèle multimodal, mesures de précision, rappel, F1-score macro, calibration et évaluation externe. Aucune performance expérimentale nouvelle n'est revendiquée dans cet article : les résultats devront être établis à partir d'un jeu de données documenté et d'une validation indépendante. La contribution principale réside dans une architecture adaptée aux contraintes de connectivité, de ressources et de surveillance phytosanitaire en RDC, tout en maintenant la validation experte comme garde-fou.

Mots-clés : phytopathologie numérique ; manioc ; CMD ; CBSD ; MobileNetV2 ; apprentissage multimodal ; SIG ; télédétection ; RDC

Abstract

Cassava diseases, particularly cassava mosaic disease (CMD) and cassava brown streak disease (CBSD), threaten food production in the Democratic Republic of the Congo (DRC). This paper presents Eagle AgriP, a proposed plant-health information system combining deep-learning analysis of leaf images, agroclimatic variables, geolocation and geographic information systems. The visual branch uses MobileNetV2 with transfer learning, while a multilayer perceptron encodes tabular environmental data; the representations are fused to support preliminary classification and risk-oriented field follow-up. The paper defines a reproducible validation protocol, including site-level data splitting, image-only versus multimodal comparisons, macro-F1, recall, calibration and external validation. No new empirical performance is claimed. The contribution is a context-aware architecture for phytopathology surveillance in the DRC, with expert confirmation for uncertain or high-impact cases.

Keywords : phytopathologie numérique ; manioc ; CMD ; CBSD ; MobileNetV2 ; apprentissage multimodal ; SIG ; télédétection ; RDC

Date of Submission: 11-10-2026

Date of Acceptance: 24-10-2026

La station de Luki est envisagée comme source de variables agroclimatiques de référence dans le projet, sous réserve de vérifier la disponibilité, la résolution temporelle, la proximité géographique et la qualité des séries réellement accessibles. Les données de station ne doivent pas être considérées comme représentatives de toute la RDC sans étude de transférabilité.

2.2. Classes et sources de données

Le jeu de données doit associer à chaque observation un identifiant, une classe diagnostique, la date, la variété si connue, le stade de développement, le site/parcelle, les coordonnées si consenties, les conditions de prise de vue et la source de confirmation. Les classes minimales proposées sont : saine, CMD, CBSD et autre/indéterminée. La classe « autre/indéterminée » est importante pour éviter de forcer le modèle à attribuer une maladie connue à un symptôme inhabituel.

Source	Variables / contenu	Contrôles nécessaires
Images terrain	Feuilles prises au smartphone ou à la caméra	Neteté, éclairage, distance, doublons, arrière-plan
Annotations expertes	Classe, symptômes, degré de certitude, méthode de confirmation	Guide d'annotation, double lecture d'un sous-échantillon
Données agroclimatiques	Température, humidité relative, pluie, rayonnement, vent	Unités, horodatage, valeurs manquantes, proximité de la parcelle
Métadonnées spatiales	Site, parcelle, coordonnées GPS, date	Consentement, précision GPS, accès restreint si nécessaire
Télédétection (optionnelle)	Bandes multispectrales, NDVI, séries temporelles	Résolution, nuages, correction atmosphérique, validation terrain

Les images issues d'une base publique, telle que PlantVillage, peuvent servir au pré-entraînement ou à une première expérimentation, mais elles ne remplacent pas un échantillon congolais représentatif. Les différences de variété, de fond, de lumière, de matériel photographique et de prévalence peuvent provoquer un décalage de domaine.

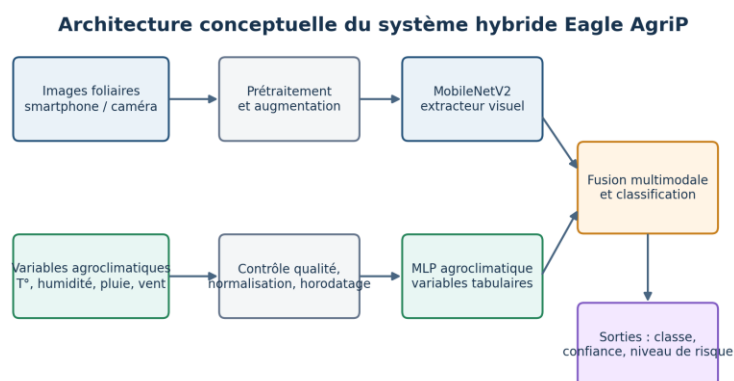


Figure 1. Architecture conceptuelle du modèle hybride Eagle AgriP. Schéma de conception, non résultat expérimental.

2.3. Prétraitement et organisation des données

Les images sont contrôlées, redimensionnées à 224×224 pixels pour l'entrée de MobileNetV2, puis normalisées selon les exigences du modèle pré-entraîné. L'augmentation de données peut inclure des rotations limitées, des retournements et de légères variations photométriques, à condition de ne pas transformer les symptômes au point de modifier leur sens agronomique. Les images provenant d'une même plante, d'une même parcelle ou d'une même séquence doivent être regroupées avant la partition des données afin d'éviter qu'une quasi-duplicata soit présente simultanément dans l'entraînement et le test.

Une partition indicative est 70 % pour l'entraînement, 15 % pour la validation et 15 % pour le test, à ajuster selon la taille des données et les regroupements disponibles. Lorsque plusieurs sites sont disponibles, une validation externe par site ou par période est préférable. Les variables continues sont normalisées à partir des seules statistiques calculées sur l'ensemble d'entraînement ; les valeurs manquantes doivent être imputées selon une règle documentée, avec indicateur de manque si cela est pertinent.

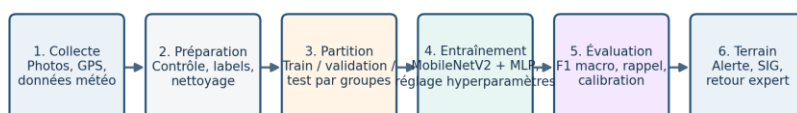
2.4. Architecture du modèle

La branche visuelle utilise MobileNetV2, architecture légère fondée sur des blocs résiduels inversés et des convolutions séparables en profondeur, adaptée aux contraintes de calcul des dispositifs mobiles (Sandler et al., 2018). L'apprentissage par transfert consiste à initialiser l'extracteur avec des poids pré-entraînés, à geler

d'abord une partie des couches, puis à dégeler sélectivement certaines couches lors d'un ajustement fin si les données le permettent.

La branche agroclimatique est un MLP recevant un vecteur de variables normalisées. Les deux vecteurs de représentation sont concaténés, puis transmis à une tête de classification comprenant une ou plusieurs couches denses, une régularisation appropriée et une sortie softmax. Le nombre de paramètres, les couches gelées, les hyperparamètres et la version exacte du modèle doivent être consignés dans le rapport expérimental.

Cycle de développement, validation et déploiement



Boucle d'amélioration : les erreurs confirmées par les experts enrichissent la base, avec traçabilité des versions et contrôle de la qualité des annotations.

Figure 2. Cycle reproductible de collecte, d'entraînement, de validation et d'amélioration du système.

2.5. Formalisation mathématique

Soit x_i une image foliaire, z_i le vecteur agroclimatique associé et y_i la classe de référence. L'extracteur visuel produit $v_i = f\theta(x_i)$ et la branche tabulaire produit $a_i = g\phi(z_i)$. La représentation fusionnée est $h_i = [v_i; a_i]$, puis la distribution de classes est $p_i = \text{softmax}(W h_i + b)$. La prédiction est $\hat{y}_i = \text{argmax}_k p_{ik}$. Les paramètres θ , ϕ , W et b sont estimés par minimisation d'une fonction de perte, par exemple l'entropie croisée pondérée :

$$L = -(1/N) \sum_i \sum_k w_k y_{ik} \log(p_{ik})$$

où N désigne le nombre d'exemples, y_{ik} l'indicateur de la classe réelle, p_{ik} la probabilité prédite et w_k un poids de classe éventuel. Les poids ne doivent être employés qu'en réponse à un déséquilibre documenté et leur effet doit être évalué sur la validation.

Pour une surveillance territoriale, le score de risque ne doit pas être assimilé directement à la probabilité d'épidémie. Une formulation exploratoire peut combiner la probabilité de classe, la confiance calibrée, l'incidence observée, la sévérité et des facteurs environnementaux. La fonction de risque doit être définie avant l'évaluation et validée avec les phytopathologues.

2.6. Évaluation et comparaisons

Le protocole doit comparer au minimum : (M1) MobileNetV2 image seule ; (M2) MLP agroclimatique seul ; (M3) fusion image + agroclimat. Un modèle de référence simple (par exemple régression logistique ou forêt aléatoire sur caractéristiques disponibles) peut aider à vérifier que la complexité supplémentaire apporte un gain réel. Les hyperparamètres sont sélectionnés sur la validation, jamais sur le jeu de test.

Indicateur	Définition / usage
Précision (precision)	$TP / (TP + FP)$: proportion de prédictions positives correctes.
Rappel (recall)	$TP / (TP + FN)$: capacité à retrouver les cas réellement positifs.
F1-score	$2 \times \text{précision} \times \text{rappel} / (\text{précision} + \text{rappel})$.
F1 macro	Moyenne non pondérée des F1 par classe, utile si les classes sont déséquilibrées.
Matrice de confusion	Détaille les confusions entre saine, CMD, CBSD et autre/indéterminée.
Calibration	Vérifie si les confiances annoncées correspondent aux fréquences observées.
Latence et taille	Mesurent la faisabilité sur serveur et/ou smartphone en connectivité limitée.

Les intervalles de confiance doivent être calculés lorsque la taille d'échantillon le permet, idéalement par bootstrap au niveau de la parcelle ou du site plutôt qu'au niveau de l'image isolée. Les résultats doivent inclure les effectifs par classe, les exclusions, les données manquantes, la version du code et les conditions matérielles.

2.7. Déploiement, SIG et supervision

Le système proposé comprend une interface de collecte mobile, une API ou un service d'inférence, un stockage contrôlé, un tableau de bord et une couche SIG. La collecte peut fonctionner en mode différé : les observations

sont mises en file d'attente hors ligne puis synchronisées lorsque le réseau redevient disponible. Les coordonnées exactes doivent être protégées ; les cartes publiques peuvent agréger les observations à une maille plus large.

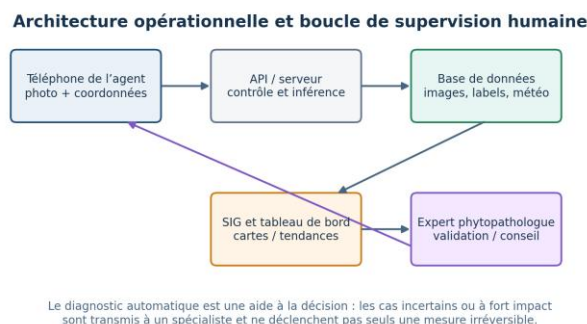


Figure 3. Architecture opérationnelle proposée avec supervision experte et boucle de retour.

2.8. Indicateurs de terrain

L'évaluation algorithmique ne suffit pas. Le pilote terrain doit mesurer le délai entre observation et avis expert, la proportion de cas nécessitant une confirmation, le taux d'alertes jugées pertinentes, la couverture des parcelles, le coût par diagnostic, l'usage effectif par les agents et les effets éventuels sur les décisions de lutte intégrée. L'incidence peut être calculée par $I = (\text{nombre de plants malades} / \text{nombre de plants observés}) \times 100$. La sévérité doit suivre une échelle définie et stable ; si une échelle ordinale 0–5 est utilisée, la sévérité normalisée peut être rapportée en pourcentage du score maximal, en précisant exactement la formule et les unités.

Matériel	Fonction	Données / rôle dans le projet
Smartphone	Acquisition d'images	Photographies de feuilles de manioc en conditions réelles
Capteurs IoT	Mesure environnementale	Température, humidité relative, pluviométrie, rayonnement, vent (caractéristiques météorologiques : Tableau Error! Reference source not found.)
Station de Luki	Relevé climatique	Pluviométrie, température, humidité de référence
Passerelle / API mobile	Transmission	Acheminement sécurisé des données vers le serveur
Serveur GPU	Calcul	Entraînement et inférence du modèle hybride (CNN + MLP)
Base Souveraine Locale	Stockage	Hébergement national des images, relevés et diagnostics

Tableau 3.1 : Matériels du dispositif de collecte et de traitement d'Eagle AgriP.

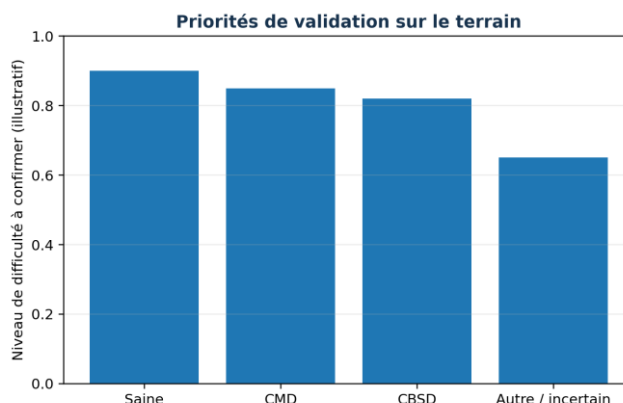


Schéma qualitatif de planification — valeurs illustratives, non issues d'une expérience.

Figure 4. Exemple qualitatif de priorités de validation ; les valeurs illustrées ne constituent pas des mesures empiriques.

III. Résultats attendus et stratégie de validation

Étant donné que le document de base ne fournit pas ici un jeu de données expérimental complet, les résultats de performance ne sont pas rapportés comme acquis. Les livrables attendus sont : un corpus documenté, une version reproductible du modèle, une comparaison des trois configurations M1–M3, une analyse des erreurs par classe et par site, une évaluation de calibration, un prototype de tableau de bord et un rapport d'utilisabilité.

Question évaluée	Résultat à rapporter	Critère de prudence
La fusion améliore-t-elle le diagnostic ?	Différence de F1 macro/rappel entre M1 et M3 avec incertitude	Un gain moyen peut masquer une baisse sur une classe rare
Le modèle généralise-t-il ?	Performances sur site ou période non vus	Ne pas confondre test aléatoire et validation externe
Les alertes sont-elles utiles ?	Délai, pertinence et proportion confirmée par expert	Une alerte n'est pas une confirmation de laboratoire
Le système est-il utilisable ?	Taux de complétion, temps de saisie, retours utilisateurs	Tester avec les agents et producteurs concernés

IV. Discussion

La valeur scientifique d'Eagle AgriP repose moins sur l'usage isolé d'un réseau neuronal que sur l'intégration de sources hétérogènes, la traçabilité des données et l'organisation de la réponse phytosanitaire. L'information agroclimatique peut aider à contextualiser une image, mais son apport doit être démontré par comparaison contrôlée : une corrélation saisonnière peut améliorer artificiellement les scores si la partition des données n'est pas rigoureuse. De même, un modèle performant sur des images propres peut échouer sur le terrain lorsque les feuilles sont partiellement occultées, vieillissantes ou photographiées dans des conditions lumineuses variables.

L'architecture légère de MobileNetV2 constitue un compromis plausible entre capacité de représentation et coût de calcul. Cependant, le choix définitif dépendra des ressources disponibles, de la taille du corpus local, de la qualité des annotations et de la nécessité d'un fonctionnement hors ligne. Les résultats publiés sur des jeux de données externes ne peuvent pas être transférés directement à la RDC sans validation locale.

Le SIG apporte une valeur opérationnelle lorsqu'il permet de repérer des regroupements d'observations, de prioriser les inspections et de documenter l'évolution temporelle. Il faut néanmoins distinguer la carte des prédictions du modèle, la carte des cas confirmés et la carte de risque épidémiologique. Ces trois objets n'ont ni la même signification ni les mêmes exigences de validation.

4.1. Limites et menaces à la validité

- Représentativité : les bases publiques peuvent sous-représenter les variétés, pratiques et environnements congolais.
- Qualité diagnostique : les étiquettes fondées uniquement sur l'apparence peuvent contenir des erreurs ; une confirmation experte ou de laboratoire est nécessaire pour les cas ambigus.
- Fuite de données : plusieurs images d'une même plante ou parcelle peuvent donner des scores artificiellement élevés si elles sont réparties entre train et test.
- Variables environnementales : une station éloignée ou des données temporellement décalées peuvent représenter imparfaitement les conditions du plant.

- Équité d'accès : le coût du smartphone, la couverture réseau, la langue et la littératie numérique peuvent limiter l'adoption.
- Interprétabilité : la confiance du réseau n'est pas une preuve causale et doit être accompagnée de contrôles et de supervision humaine.

V. Gouvernance des données et considérations éthiques

Le protocole doit documenter le consentement des producteurs ou détenteurs des parcelles, la finalité de la collecte, la durée de conservation, les droits d'accès et la procédure de suppression. Les coordonnées et informations identifiantes doivent être minimisées, protégées et partagées selon des règles explicites. Les images ne doivent pas être réutilisées pour une finalité différente sans base légitime et information adéquate. Les recommandations de lutte doivent rester cohérentes avec les pratiques de lutte intégrée, les règles phytosanitaires applicables et les conseils des spécialistes.

VI. Conclusion

Cet article propose une architecture scientifique et un protocole de validation pour Eagle AgriP, système d'information phytopathologique destiné à soutenir le diagnostic précoce de la CMD et de la CBSD en RDC. L'approche associe un extracteur visuel MobileNetV2, une branche agroclimatique MLP, une fusion multimodale, la géolocalisation et un SIG de surveillance. Sa crédibilité dépendra d'une validation locale rigoureuse, de données traçables, d'une séparation des jeux par site ou parcelle et de la confirmation experte des cas incertains. La priorité immédiate est donc la constitution d'un corpus congolais de qualité et la conduite d'un pilote prospectif ; aucune performance ne devrait être annoncée avant cette étape. En pratique, Eagle AgriP est à considérer comme un outil d'aide à la décision et de priorisation de la surveillance, non comme un substitut autonome au diagnostic phytopathologique.

Références bibliographiques

- [1]. Dean, R., Van Kan, J. A. L., Pretorius, Z. A., Hammond-Kosack, K. E., Di Pietro, A., Spanu, P. D., Rudd, J. J., Dickman, M., Kahmann, R., Ellis, J., & Foster, G. D. (2012). The Top 10 fungal pathogens in molecular plant pathology. *Molecular Plant Pathology*, 13(4), 414–430.
- [2]. Ferentinos, K. P. (2018). Deep learning models for plant disease detection and diagnosis. *Computers and Electronics in Agriculture*, 145, 311–318.
- [3]. Kamilaris, A., & Prenafeta-Boldú, F. X. (2018). Deep learning in agriculture: A survey. *Computers and Electronics in Agriculture*, 147, 70–90.
- [4]. Legg, J. P., et al. (2014). Cassava virus diseases: biology, epidemiology, and management. *Advances in Virus Research*, 91, 85–142.
- [5]. Madden, L. V., Hughes, G., & van den Bosch, F. (2007). *The Study of Plant Disease Epidemics*. APS Press.
- [6]. Mahlein, A.-K. (2016). Plant disease detection by imaging sensors—parallels and specific demands for precision agriculture and plant phenotyping. *Plant Disease*, 100(2), 241–251.
- [7]. Mohanty, S. P., Hughes, D. P., & Salathé, M. (2016). Using deep learning for image-based plant disease detection. *Frontiers in Plant Science*, 7, 1419.
- [8]. Msikita, W., James, B., Nnodu, E., Legg, J., Wydra, K., & Ogbe, F. (2000). Lutte contre les maladies du manioc : guide pratique de lutte intégrée à l'usage des vulgarisateurs. IITA.
- [9]. Mulla, D. J. (2013). Twenty five years of remote sensing in precision agriculture: Key advances and remaining knowledge gaps. *Biosystems Engineering*, 114(4), 358–371.
- [10]. Oerke, E.-C. (2006). Crop losses to pests. *The Journal of Agricultural Science*, 144(1), 31–43.
- [11]. Ramcharan, A., et al. (2017). Deep learning for image-based cassava disease detection. *Frontiers in Plant Science*, 8, 1852.
- [12]. Rouse, J. W., Haas, R. H., Schell, J. A., & Deering, D. W. (1974). Monitoring vegetation systems in the Great Plains with ERTS. *NASA SP-351*, 309–317.
- [13]. Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L.-C. (2018). MobileNetV2: Inverted residuals and linear bottlenecks. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 4510–4520.
- [14]. Savary, S., Willocquet, L., Pethybridge, S. J., Esker, P., McRoberts, N., & Nelson, A. (2019). The global burden of pathogens and pests on major food crops. *Nature Ecology & Evolution*, 3, 430–439.
- [15]. Too, E. C., Yujian, L., Njuki, S., & Yingchun, L. (2019). A comparative study of fine-tuning deep learning models for plant disease identification. *Computers and Electronics in Agriculture*, 161, 272–279.
- [16]. Zhang, N., Wang, M., & Wang, N. (2002). Precision agriculture—a worldwide overview. *Computers and Electronics in Agriculture*, 36(2–3), 113–132.

Annexe A. Formules des métriques

Pour une classe donnée : précision = TP/(TP+FP) ; rappel = TP/(TP+FN) ; F1 = 2 × précision × rappel/(précision + rappel). TP désigne les vrais positifs, FP les faux positifs et FN les faux négatifs. Pour une classification multiclassée, les indicateurs sont calculés par classe puis agrégés, en précisant la méthode d'agrégation.

Annexe B. Fiche minimale de collecte

Champ	Description
ID observation	Identifiant pseudonymisé et unique
Date / heure	Horodatage de la prise de vue et de la collecte météo
Site / parcelle	Code du site ; coordonnées selon consentement
Culture	Variété si connue, âge/stade de développement

Image	Fichier original, qualité, appareil si disponible
Symptômes	Description standardisée et partie de plante concernée
Classe de référence	Saine / CMD / CBSD / autre-indéterminée
Confirmation	Agent formé, expert, test de laboratoire ou statut non confirmé
Agroclimat	Température, humidité, précipitations, source et distance de la station
Suivi	Action recommandée, décision experte, résultat de visite ultérieure