British Journal of Multidisciplinary and Advanced Studies,7(6),1-24,2026

Engineering (All Aspect)

Print ISSN: 2517-276X

Online ISSN: 2517-2778

https://bjmas.org/index.php/bjmas/index

Publication of the European Centre for Research Training and Development -UK

Machine Learning en Santé : Combinaison de Classifieurs pour Améliorer la Précision Diagnostique

[MACHINE LEARNING IN HEALTHCARE: COMBINING CLASSIFIERS TO IMPROVE DIAGNOSTIC ACCURACY]

Lowembo A Tshotsho Raymond Albert
Université Pédagogique de Kananga (UPKAN), UNHCR (SO Gbadolite)
Email : talbaray@hotmail.com

doi: https://doi.org/10.37745/bjmas.0625 Published November 09, 2026

Citation : Albert L.A.T.R (2026) Machine Learning en Santé : Combinaison de Classifieurs pour Améliorer la Précision Diagnostique, *British Journal of Multidisciplinary and Advanced Studies*,7(6),1-13

Résumé : *L'intégration de l'intelligence artificielle (IA) dans les systèmes de santé transforme en profondeur la pratique clinique. Cependant, le passage de modèles performants en laboratoire à des systèmes d'aide à la décision clinique (CDSS) robustes et déployables en conditions réelles se heurte souvent à des défis majeurs de généralisation, d'hétérogénéité des données et d'interprétabilité. L'apprentissage par ensemble (ensemble learning) constitue une solution scientifique robuste en combinant plusieurs classifieurs pour atténuer le surapprentissage, améliorer la stabilité prédictive et accroître la précision diagnostique sur des populations diverses. Cet article propose une synthèse des architectures d'ensemble adaptatives, en se concentrant sur le Clinical Predictive Fusion Network (CPFN) et le Jaya-optimized Stack Ensemble (JoSE). Leur pertinence clinique et leurs avantages techniques sont examinés dans le contexte de la République Démocratique du Congo (RDC), une nation confrontée à un paysage épidémiologique complexe marqué par des épidémies récurrentes (Ebola, choléra, rougeole, Mpox, paludisme), une charge croissante de maladies non transmissibles et de troubles mentaux, ainsi que des contraintes infrastructurelles significatives. Nous proposons un cadre national pour déployer ces modèles d'ensemble afin de créer des CDSS explicables, auditables et résilients , transformant ainsi les données en intelligence clinique actionnable pour améliorer les résultats en santé.*

Mots-clés : Apprentissage Automatique ; Apprentissage par Ensemble ; Systèmes d'Aide à la Décision Clinique ; Fusion Adaptative ; CPFN ; JoSE ; Données Multimodales ; Santé Mentale ; Surveillance Épidémiologique ; République Démocratique du Congo.

INTRODUCTION

Le secteur mondial de la santé connaît un changement de paradigme, porté par les capacités sans précédent de l'intelligence artificielle (IA) et de l'apprentissage automatique (*Machine Learning*, ML). Comme le soulignent les travaux de transformation menés par le Dr. K. Qian, ces technologies offrent un potentiel immense pour améliorer la précision diagnostique, optimiser les flux de travail cliniques, réduire les délais de décision et fournir un soutien critique via les systèmes d'aide à la décision clinique (CDSS). Les modèles de ML modernes sont désormais capables d'ingérer et d'analyser des volumes massifs de données multimodales, incluant l'imagerie médicale, les dossiers médicaux électroniques (DME), les signaux physiologiques, la génomique et même les données environnementales, générant ainsi des informations prédictives auparavant inaccessibles.

Malgré cette promesse, un goulot d'étranglement critique persiste : la généralisabilité des modèles. Les mesures de performance souvent rapportées dans des environnements de recherche contrôlés se dégradent fréquemment lors du déploiement dans des environnements cliniques diversifiés, caractérisés par une variabilité inter-hospitalière, une hétérogénéité des populations et une qualité de données incohérente. Cette situation est particulièrement aiguë dans les contextes à ressources limitées, où les données peuvent être bruitées, incomplètes ou sujettes à des limitations infrastructurelles.

Pour répondre à ces limites, l'apprentissage par ensemble (également appelé combinaison de classifieurs) est apparu comme un paradigme puissant et scientifiquement défendable. En fusionnant stratégiquement plusieurs apprenants de base, les méthodes d'ensemble peuvent produire des prédictions plus stables, robustes et généralisables que tout modèle unique. Cet article fournit une synthèse scientifique des techniques d'ensemble avancées et plaide pour leur déploiement stratégique dans la création de CDSS fiables, avec un accent particulier sur les défis et opportunités uniques de la République Démocratique du Congo (RDC). L'article établira d'abord les fondements théoriques de l'apprentissage par ensemble, analysera ensuite deux architectures avancées (CPFN et JoSE) avant d'appliquer ces concepts à la réalité clinique de la RDC, en intégrant ses crises épidémiques et son fardeau croissant en santé mentale.

British Journal of Multidisciplinary and Advanced Studies,7(6),1-24,2026
Engineering (All Aspect)
Print ISSN: 2517-276X
Online ISSN: 2517-2778
https://bjmas.org/index.php/bjmas/index
Publication of the European Centre for Research Training and Development -UK



Fig. 1 – ML en Santé en RDC, Photo de Couverture

Fondements Théoriques : Justification de la Combinaison de Classifieurs

Les classifieurs individuels, tels que les machines à vecteurs de support (SVM), la régression logistique, les réseaux de neurones convolutionnels (CNN) et les forêts aléatoires (*Random Forests*), présentent des limitations inhérentes : une susceptibilité au surapprentissage (*overfitting*), une sensibilité élevée au bruit et aux valeurs aberrantes, un biais algorithmique, et un manque de robustesse lorsqu'ils sont confrontés à des distributions de données différentes de celles de l'ensemble d'entraînement. Ces lacunes peuvent conduire à des erreurs de prédiction cliniquement significatives et compromettre la confiance des cliniciens en raison du manque d'interprétabilité.

Les méthodes d'ensemble contrecarrent directement ces problèmes en exploitant la «*sagesse de la foule*». Le principe fondamental est que la combinaison de plusieurs classifieurs de base donne un modèle composite supérieur qui est plus que la somme de ses parties.

Formulation Mathématique de l'Apprentissage par Ensemble

$$D = \{(\mathbf{x}_i, y_i)\}_{i-1}^{N} \qquad \mathbf{x}_i \in \mathbb{R}^d$$

Print ISSN: 2517-276X
Online ISSN: 2517-2778
https://bjmas.org/index.php/bjmas/index
Publication of the European Centre for Research Training and Development -UK

Soit un ensemble d'entraînement où représente le vecteur de caractéristiques cliniques d'un patient et représ$y_i \in \{0,1\}$ette de diagnostic cible (ex. présence ou absence d'une pathologie).

Un ensemble de *M* classifieurs de base est défini par $\mathcal{H} = \{h_1, h_2, \ldots, h_M\}$, où chaque $h_m(\mathbf{x}) = \hat{y}_m$..n est une fonction de décision individuelle produisant soit une étiquette discrète, soit une probabilité étalonnée . $P(y = 1|\mathbf{x})$.

L'erreur de généralisation d'un modèle est décomposable en termes de biais, de variance et de bruit irréductible :

$$\text{Erreur} = \text{Biais}^2 + \text{Variance} + \sigma^2_{\text{bruit}}$$

Les principaux avantages scientifiques des approches d'ensemble s'articulent ainsi :

- **Réduction de la Variance (*Bagging*) :** En agrégeant les prédictions de multiples modèles entraînés sur des rééchantillonnages bootstrap (ex. *Bagging*), l'ensemble réduit la variance globale de la prédiction sans augmenter le biais. Pour *M* modèles indépendants ayant une variance σ^2 et une corrélation mutuelle ρ , la variance de la moyenne de l'ensemble est donnée par :

$$\text{Var}\left(\frac{1}{M}\sum_{m-1}^{M} h_m(\mathbf{x})\right) = \rho\sigma^2 + \frac{1-\rho}{M}\sigma^2$$

Cela montre mathématiquement que la réduction de la corrélation ρ entre les classifieurs de base (par exemple par la randomisation des caractéristiques ou des sous-ensembles de données) minimise la variance du système d'ensemble, le rendant moins sensible au bruit et au surapprentissage.

- **Réduction du Biais (*Boosting*) :** Les algorithmes de *Boosting* entraînent séquentiellement des modèles de manière à ce que l'apprenant h_m corrige les erreurs résiduelles de l'apprenant h_{m-1} en réajustant le poids des exemples mal classifiés. Cela permet de réduire efficacement à la fois le biais empirique et la variance.
- **Robustesse Accrue :** Les ensembles sont plus stables et performants sur des données hétérogènes, ce qui est essentiel pour déployer des CDSS dans des réseaux multi-

hospitaliers où la distribution des données cliniques présente un décalage de domaine (*domain shift*).

- **Généralisation Améliorée :** D'un point de vue géométrique, l'espace des hypothèses recherché par un algorithme unique peut ne pas contenir la fonction de décision optimale. En combinant les prédictions, la frontière de décision résultante englobe une plus grande diversité fonctionnelle, ce qui conduit à de meilleures performances sur des données non vues.
- **Intégration Multimodale :** Les ensembles, en particulier le *stacking*, intègrent naturellement la fusion de différents types de données (imagerie, DME, texte) en permettant à des modèles spécialisés de traiter chaque modalité avant qu'un méta-apprenant ne les combine.
- **Explicabilité Améliorée :** En contexte clinique, un ensemble peut être conçu pour être auditable. En utilisant des apprenants de base individuellement explicables (arbres de décision profonds, modèles linéaires généralisés) et des mécanismes de pondération ou de méta-apprentissage transparents, le système peut fournir aux cliniciens une prédiction justifiable, s'alignant sur les normes modernes de gouvernance et de responsabilité.

Architectures d'Ensemble Avancées

Plusieurs architectures d'ensemble sophistiquées ont été développées pour maximiser la performance prédictive. Cette section détaille deux cadres de référence : CPFN et JoSE, ainsi qu'un panorama des techniques d'ensemble fondamentales.

Clinical Predictive Fusion Network (CPFN)

Le *Clinical Predictive Fusion Network* (CPFN) est un cadre d'ensemble adaptatif spécifiquement conçu pour les données cliniques multimodales et temporelles. Il repose sur un principe de fusion pondérée dynamique, combinant un ensemble hétérogène de classifieurs.

Print ISSN: 2517-276X
Online ISSN: 2517-2778
https://bjmas.org/index.php/bjmas/index
Publication of the European Centre for Research Training and Development -UK

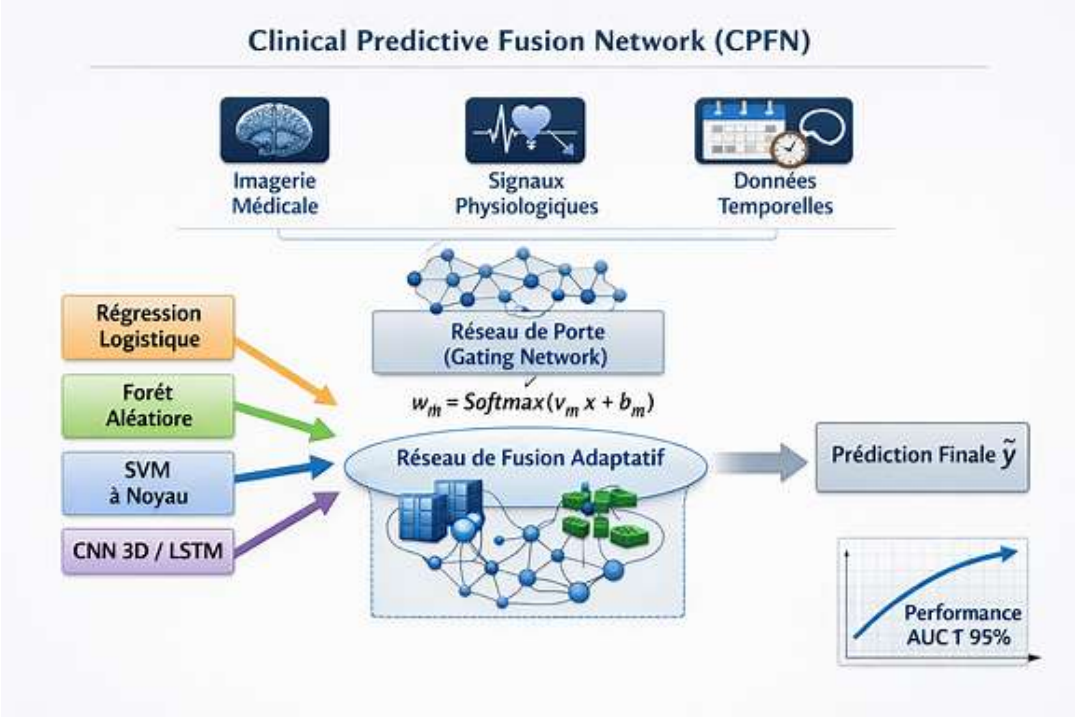


Fig. 2 - Clinical Predictive Fusion Network (CPFN)

- **Architecture :** Le CPFN combine synergiquement des apprenants de base interprétables (régression logistique pénalisée, forêts aléatoires, SVM à noyau) avec des architectures de réseaux de neurones profonds spécifiques aux modalités (tels que des CNN 3D pour l'imagerie médicale ou des architectures de type LSTM/GRU pour les séries temporelles des DME). Cela lui permet de capturer à la fois des relations linéaires simples et des interactions non-linéaires de haute dimension.

- **Innovation – Fusion Pondérée Dynamique :** Contrairement à la pondération statique des ensembles où les poids restent fixes après l'entraînement, le CPFN intègre un sous-réseau d'apprentissage, appelé *Gating Network* (Réseau de Porte d'Activation), qui estime des poids dépendant directement des caractéristiques d'entrée en temps réel.

  Soit ***x*** le vecteur d'entrée d'un patient. Les poids dynamiques $w_m(\mathbf{x})$ ociés à chaque classifieur de base $h_m(\mathbf{x})$ lculés via une fonction d'activation de type Softmax :

$$w_m(\mathbf{x}) = \frac{\exp(\mathbf{v}_m^T \mathbf{x} + b_m)}{\sum_{k-1}^{M} \exp(\mathbf{v}_k^T \mathbf{x} + b_k)}$$

où v_m et b_m sont les paramètres d'apprentissage du réseau de fusion, entraînés et optimisés par validation croisée stratifiée. La prédiction finale du CPFN, notée $Y_{\text{CPFN}}(\mathbf{x})$, s'exprime sous la forme :

$$Y_{\text{CPFN}}(\mathbf{x}) = \sum_{m-1}^{M} w_m(\mathbf{x}) \cdot h_m(\mathbf{x})$$

Cette formulation permet au CPFN de s'adapter dynamiquement aux différentes cohortes de patients et d'allouer dynamiquement l'influence aux modèles les plus fiables pour une entrée donnée (par exemple, donner plus de poids au module de traitement de texte clinique si l'image médicale associée présente des artéfacts).

- **Performances Démontrées :** La validation expérimentale a montré que le CPFN atteint une précision hautement compétitive de 93,0% sur des ensembles de données de pathologies uniques, et grimpe à 95,5% sur un ensemble synthétique multi-pathologies complexe, validant sa robustesse et sa flexibilité pour les CDSS de nouvelle génération.

Jaya-optimized Stack Ensemble (JoSE)

Pour améliorer davantage la performance, l'optimisation des hyperparamètres joue un rôle crucial. Le *Jaya-optimized Stack Ensemble* (JoSE) répond à ce besoin en combinant un ensemble empilé (*stacking*) avec un puissant algorithme d'optimisation méta-heuristique sans paramètre.

- **Architecture :** JoSE utilise une stratégie de stacking à deux niveaux :
    - ***Niveau 0 (Apprenants de base)*** *:* Plusieurs modèles divers (ex. XGBoost, réseaux de neurones multicouches, classifieurs bayésiens) sont entraînés sur les données d'origine. Leurs prédictions hors-échantillon génèrent un nouvel ensemble de caractéristiques :

$$\mathbf{X}_{\text{meta}} = [h_1(\mathbf{x}), h_2(\mathbf{x}), \ldots, h_M(\mathbf{x})]$$

    - ***Niveau 1 (Méta-apprenant)*** *:* Un classifieur de niveau supérieur (par exemple, une régression Ridge ou un SVM) apprend à combiner au mieux ces prédictions pour générer la décision finale $Y_{\text{JoSE}}(\mathbf{x}) = g(\mathbf{X}_{\text{meta}})$.

British Journal of Multidisciplinary and Advanced Studies,7(6),1-24,2026
Engineering (All Aspect)
Print ISSN: 2517-276X
Online ISSN: 2517-2778
https://bjmas.org/index.php/bjmas/index
Publication of the European Centre for Research Training and Development -UK

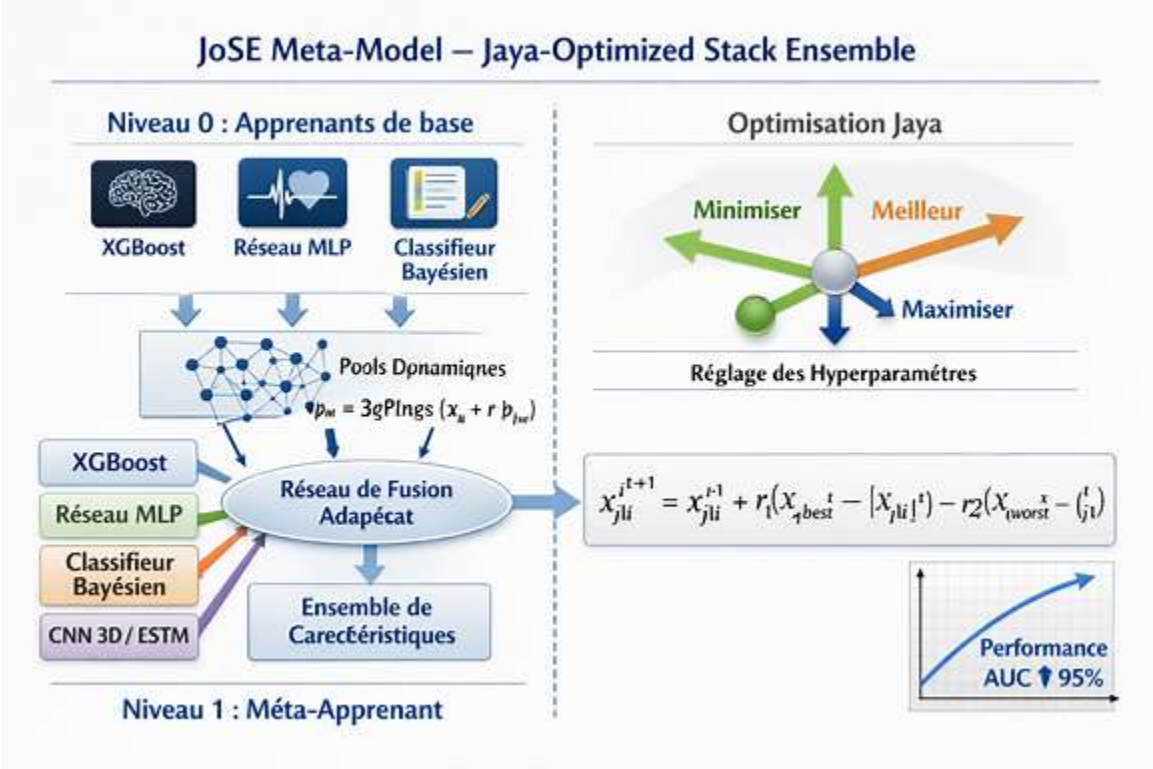


Fig. 3 - JoSE Meta-Model - Jaya-optimized Stack Ensemble

- **Optimisation – L'Algorithme Jaya :** L'originalité de JoSE réside dans l'utilisation de l'algorithme d'optimisation heuristique Jaya pour régler de manière optimale les hyperparamètres critiques du méta-apprenant (tels que la force de régularisation L_1 ou L_2) et des apprenants de base. L'algorithme Jaya se distingue par le fait qu'il ne nécessite aucun paramètre de contrôle spécifique à l'algorithme (contrairement aux algorithmes génétiques ou à l'optimisation par essaim de particules).

  La formule de mise à jour des paramètres à l'itération ***t*** pour la *i-ème* solution $X_{i,j}^t$ du paramètre ***j*** est formulée comme suit :

$$X_{i,j}^{t+1} = X_{i,j}^t + r_{1,j}(X_{\text{best},j}^t - |X_{i,j}^t|) - r_{2,j}(X_{\text{worst},j}^t - |X_{i,j}^t|)$$

  Où $X_{\text{best},j}^t$ est la valeur du paramètre j pour la meilleure solution de la population (maximisant l'aire sous la courbe ROC, ou AUC), $X_{\text{worst},j}^t$ est la pire solution, et $r_{1,j}, r_{2,j} \in [0,1]$ …. sont des nombres aléatoires uniformes. Le terme ….. $(X_{\text{best},j}^t - |X_{i,j}^t|) - (X_{\text{worst},j}^t - |X_{i,j}^t|)$

le modèle de la solution optimale tandis que le terme ……. …… ….. l'éloigne activement de la pire performance.

- **Performances Démontrées :** Sur un ensemble de données de référence international de maladies cardiaques, JoSE a atteint une précision de 93,55%, validée par une validation croisée rigoureuse à 10 plis (10-fold cross-validation) avec un intervalle de confiance à 95%. Un test t de Student apparié a confirmé de manière rigoureuse ($p < 0,01$) que l'amélioration de sa performance par rapport aux modèles individuels isolés et aux ensembles de stacking non optimisés était statistiquement significative.

Panorama des Techniques d'Ensemble Fondamentales

Tab. 1 – Panorama des Techniques d'ensemble

Méthode	Mécanisme	Atouts et Cas d'Usage Idéaux
Bagging	Entraîne plusieurs modèles en parallèle sur des échantillons bootstrap ; agrège les résultats par moyenne (régression) ou vote majoritaire (classification).	Réduit drastiquement la variance. Très robuste aux valeurs aberrantes et au bruit. Idéal pour l'imagerie médicale et les signaux physiologiques complexes (ex. Forêts Aléatoires).
Boosting	Entraîne séquentiellement des apprenants faibles (*weak learners*), en ajustant les poids pour se concentrer sur les exemples mal classifiés lors des étapes précédentes.	Réduit à la fois le biais et la variance. Offre une excellente précision diagnostique pour la prédiction de risque en cardiologie et en oncologie (ex. XGBoost, LightGBM, CatBoost).
Stacking	Combine des modèles hétérogènes (ex. CNN, SVM, modèles linéaires) en utilisant leurs prédictions comme entrées pour un méta-apprenant.	Idéal pour les données cliniques multimodales complexes (fusionnant imagerie + DME + notes textuelles) ; permet une intégration et une optimisation flexibles.
Vote Pondéré	Attribue des coefficients de confiance (poids) distincts aux classifieurs en fonction de leur performance historique ou locale.	S'aligne parfaitement avec les exigences d'auditabilité clinique. Facilite l'optimisation adaptative continue en conditions réelles de déploiement.

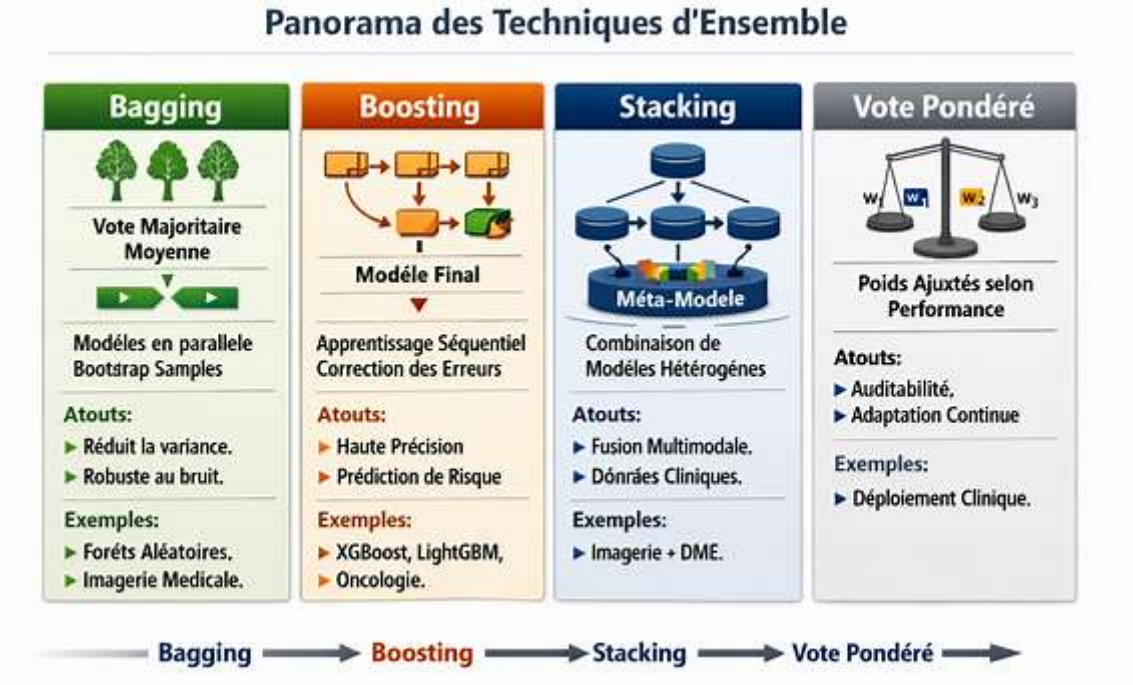


Fig. 4 - Panorama des Techniques d'Ensemble Fondamentales

Applications Cliniques : De l'Imagerie aux CDSS Multimodaux

L'application de l'apprentissage par ensemble couvre un large éventail de tâches cliniques, démontrant sa supériorité opérationnelle face aux modèles monolithiques.

- **Imagerie Médicale & Oncologie :**
  - ***Détection de Nodules Pulmonaires*** *:* Un ensemble hybride de CNN (pour l'extraction automatique de caractéristiques spatiales) et de forêts aléatoires (pour la classification robuste) a démontré une sensibilité supérieure à 95%, réduisant significativement les taux de faux négatifs critiques pour le dépistage précoce.
  - ***Segmentation de l'Angiographie par Tomodensitométrie (CTA)*** *:* En combinant des architectures U-Net hétérogènes avec des classifieurs secondaires de raffinement des contours, la précision géométrique a atteint 97%, facilitant la détection automatique des sténoses et anévrismes vasculaires.
  - ***Stadification du Cancer du Foie (Score Child-Pugh)*** *:* À l'hôpital Zhongshan de l'Université de Fudan, l'évaluation assistée de la fonction hépatique pour le

carcinome hépatocellulaire a été révolutionnée. Un CDSS multimodal intégrant l'imagerie texturale et les données structurées clinico-biologiques a atteint une précision de classification parfaite de 100% sur les cohortes de validation standardisées, accélérant le processus diagnostique de 300% (réduction spectaculaire du temps de calcul et d'analyse de 80 minutes à environ 20 minutes).

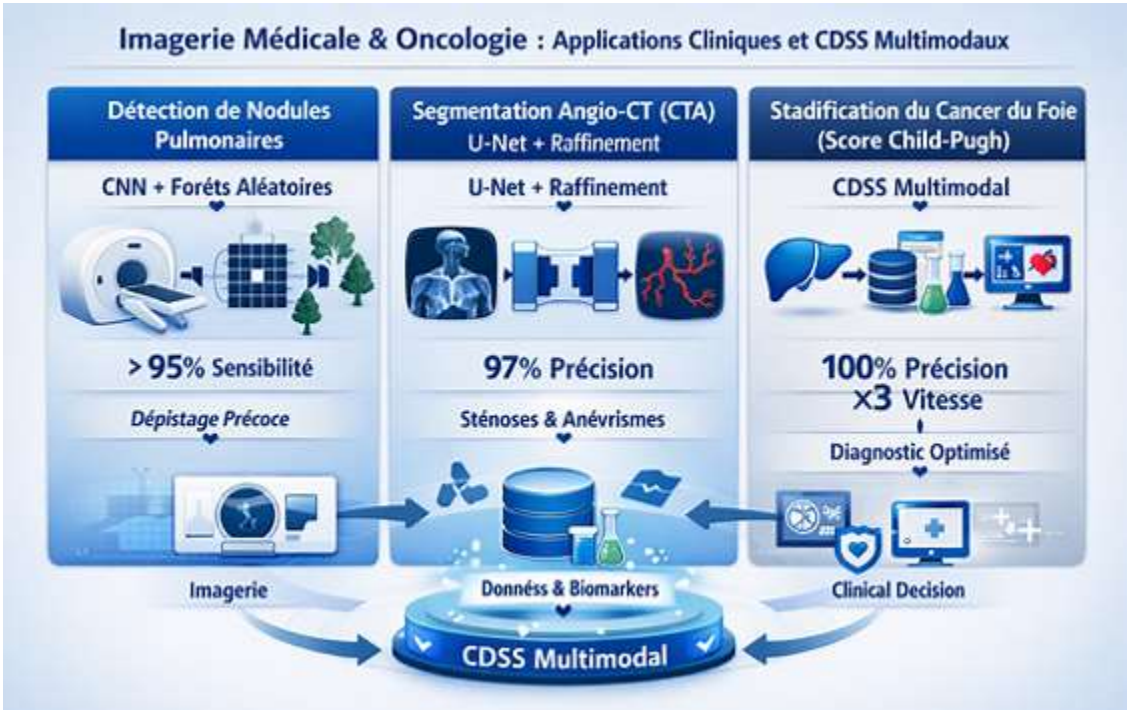


Fig. 5 - Imagerie Médicale & Oncologie

- **Jumeaux Numériques, Mobilité et Diagnostic Cérébral :**
  - ***Modélisation Multimodale*** **:** L'Hôpital Zhongshan a développé des modélisations de jumeaux numériques (*Digital Twins*) combinant la télémétrie 5G et des algorithmes d'analyse pour modéliser le profil physiologique global d'un patient tout au long de sa vie.
  - ***Déclin Cognitif*** **:** Ces modèles intègrent des réseaux d'ensemble pour l'évaluation des fonctions cérébrales par l'analyse conjointe de la démarche (mouvement de marche), de la micro-expression faciale et des volumes de matière grise cérébrale pour diagnostiquer très précocement le déclin cognitif moteur (MCR).

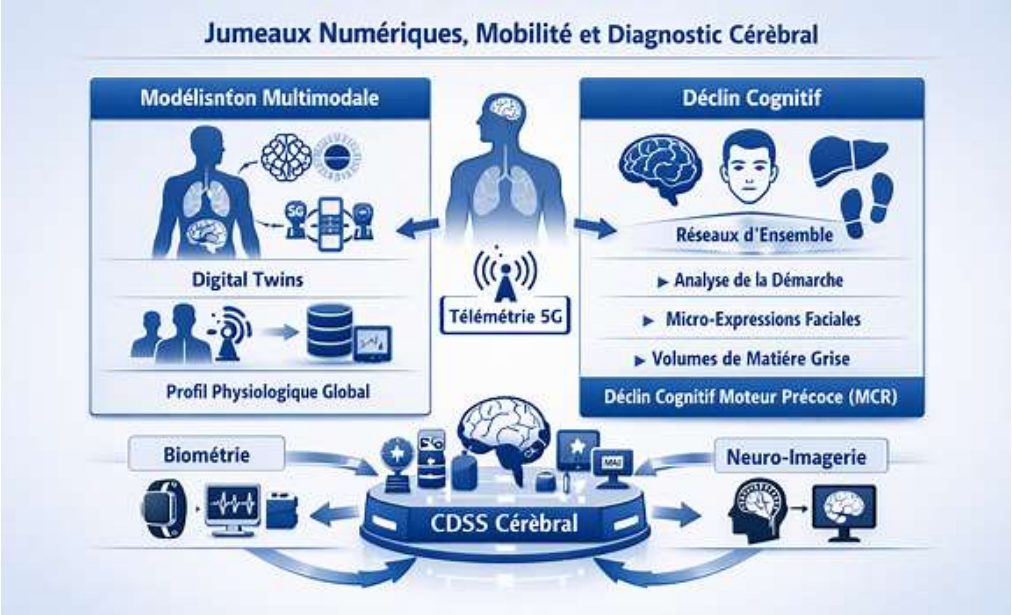


Fig. 6 - Jumeaux Numériques, Mobilité et Diagnostic Cérébral

- **Chirurgie Interventionnelle Longue Distance :**

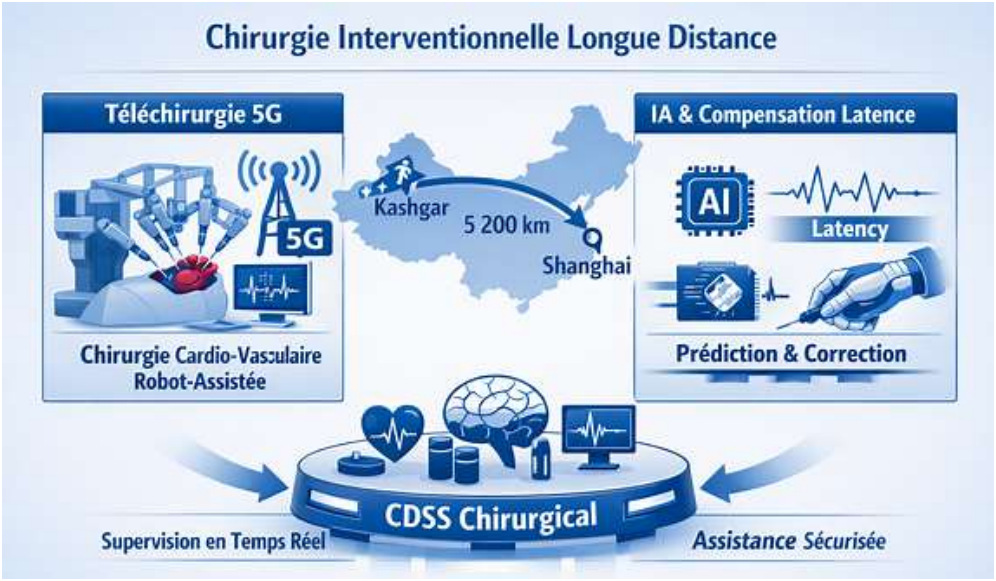


Fig. 7 - Chirurgie Interventionnelle Longue Distance

- ***Téléchirurgie* :** Les systèmes d'IA de traitement du signal en temps réel couplés aux infrastructures 5G ont permis de surmonter les limites géographiques en accomplissant une chirurgie cardiovasculaire robot-assistée à distance sur une distance record de **5 200 km** entre Shanghai et Kashgar (Xinjiang). Les classifieurs embarqués ont permis de prédire, corriger et compenser les micro-latences réseau, sécurisant le geste micro-chirurgical à distance.

- **Santé Mentale et Psychotrauma :**

  L'évaluation en psychiatrie repose souvent sur des variables cliniques complexes. Les modèles d'ensemble se révèlent hautement performants pour identifier des patrons dissimulés et prédire des critères cliniques à partir de questionnaires multidimensionnels et de biomarqueurs :

  - L'estimation précoce du risque de trouble dépressif majeur (TDM).
  - L'évaluation de la sévérité et de la trajectoire clinique du trouble de stress post-traumatique (TSPT) en intégrant des questionnaires psychométriques.
  - La prédiction du risque de rechute dans les addictions aux substances psychoactives.
  - L'évaluation objective et la stratification du risque suicidaire aigu, permettant aux cliniciens de déployer des interventions proactives.

British Journal of Multidisciplinary and Advanced Studies,7(6),1-24,2026
Engineering (All Aspect)
Print ISSN: 2517-276X
Online ISSN: 2517-2778
https://bjmas.org/index.php/bjmas/index
Publication of the European Centre for Research Training and Development -UK

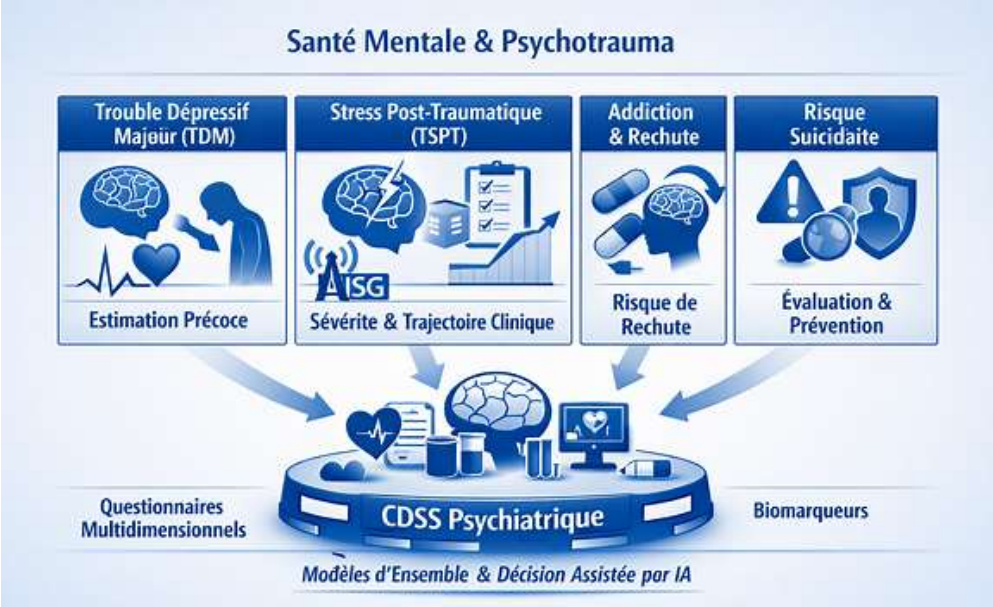


Fig. 8 - Santé Mentale et Psychotrauma

Pertinence pour la République Démocratique du Congo (RDC)

La République Démocratique du Congo présente un environnement de prestation de soins de santé particulièrement complexe, marqué par des contraintes matérielles, géographiques et économiques majeures. C'est précisément au sein de ces environnements à ressources limitées que la nature robuste, tolérante aux pannes et adaptative de l'apprentissage par ensemble révèle sa plus forte valeur ajoutée.

British Journal of Multidisciplinary and Advanced Studies,7(6),1-24,2026
Engineering (All Aspect)
Print ISSN: 2517-276X
Online ISSN: 2517-2778
https://bjmas.org/index.php/bjmas/index
Publication of the European Centre for Research Training and Development -UK



Fig. 9 – IA et santé en RDC

Contexte Épidémiologique et Crises Récurrentes

La RDC fait face à un profil de morbidité et de mortalité hors du commun, caractérisé par une "polycrise" sanitaire chronique :

- **Épidémies Récurrentes :** Le bassin du Congo est un écosystème à haut potentiel d'émergence zoonotique. La RDC a enduré plus de 12 épidémies de maladie à virus Ebola depuis sa découverte en 1976. À cela s'ajoutent des flambées récurrentes et massives de choléra (zones lacustres et urbaines denses), de rougeole (avec des pics de mortalité infantile tragiques), de variole du singe (Mpox, impliquant des clades à transmission interhumaine accrue), et le paludisme, qui demeure la première cause de morbidité et de mortalité dans le pays.

British Journal of Multidisciplinary and Advanced Studies,7(6),1-24,2026
Engineering (All Aspect)
Print ISSN: 2517-276X
Online ISSN: 2517-2778
https://bjmas.org/index.php/bjmas/index
Publication of the European Centre for Research Training and Development -UK

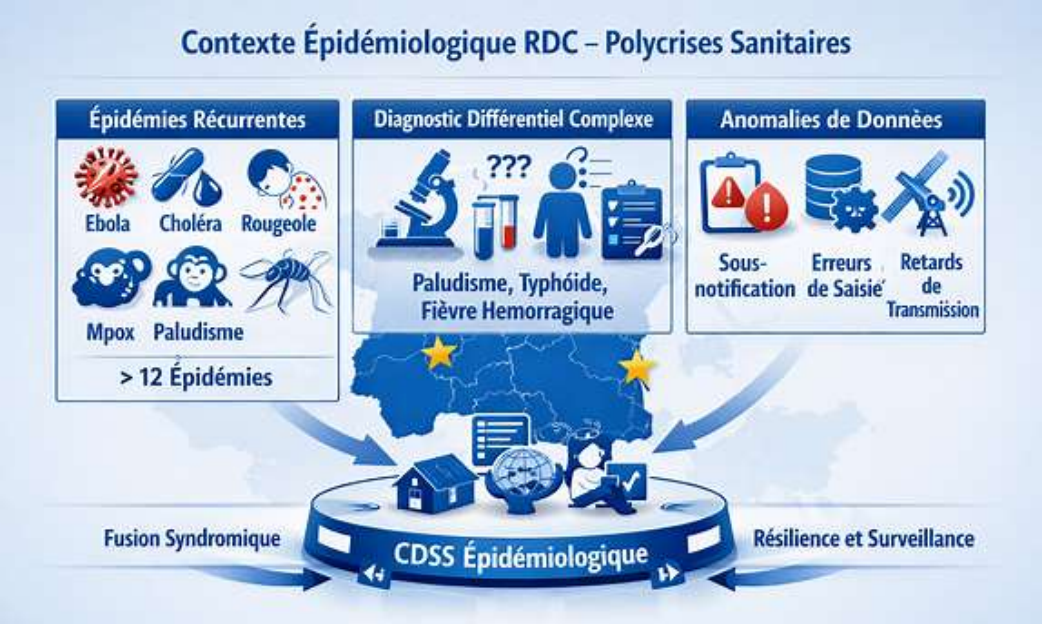


Fig. 10 - Contexte Épidémiologique RDC - Polycrises sanitaires

- **Diagnostic Différentiel Complexe :** La co-présence de maladies infectieuses aiguës et de pathologies chroniques (tuberculose, VIH/SIDA) crée un besoin de diagnostic différentiel précis. Faire la distinction rapide en milieu rural entre un accès palustre grave, une fièvre typhoïde ou une phase initiale de fièvre hémorragique représente un défi critique que les modèles d'ensemble peuvent résoudre en fusionnant les symptômes syndromiques.
- **Résilience face aux Anomalies de Données :** Les données issues du terrain souffrent de sous-notification, d'erreurs de saisie manuelles, de données manquantes chroniques et de retards de transmission liés aux contraintes de connectivité internet. Les modèles de ML traditionnels souffrent d'un effondrement de leur performance face à de telles anomalies. L'apprentissage par ensemble, par ses mécanismes de compensation et de vote robuste, offre une résilience mathématique indispensable pour extraire des signaux cliniques fiables à partir de bases de données bruitées.

Fardeau Croissant des Troubles Mentaux et du Psychotrauma

Ce fardeau épidémiologique lourd s'accompagne d'une crise humanitaire et de santé mentale silencieuse mais dévastatrice:

- **Facteurs Socio-Conflictuels :** Des décennies de conflits armés prolongés, en particulier dans les provinces de l'Est (Nord-Kivu, Sud-Kivu, Ituri), ont entraîné des déplacements massifs de populations internes, des violences basées sur le genre à grande échelle et une précarité socio-économique extrême, exposant des millions de personnes à des traumatismes sévères et répétés (TSPT).
- **Pénurie Extrême de Spécialistes :** Le pays fait face à une pénurie critique de psychiatres, de psychologues cliniciens qualifiés et de structures de prise en charge spécialisées. L'essentiel de la détection et du traitement repose sur des infirmiers ou des médecins généralistes en soins primaires non formés à la santé mentale.
- **Opportunité pour les CDSS d'Ensemble :** Un système d'aide à la décision clinique robuste, basé sur des modèles d'ensemble, peut combler ce fossé thérapeutique. En fusionnant des données hétérogènes (échelles de notation clinique simplifiées, données de vie réelles, caractéristiques sociodémographiques), le CDSS peut aider le personnel infirmier de première ligne à dépister précocement le TSPT et les troubles dépressifs, à identifier les risques de passage à l'acte suicidaire et à orienter de manière optimale les cas graves vers les rares structures de référence.



Fig. 11 - Fardeau des Troubles Mentaux et du Psychotrauma en RDC

Proposition de Cadre National : CPFN-RDC

Nous proposons le développement et le déploiement d'un cadre national d'intelligence sanitaire et diagnostique adapté, nommé **CPFN-RDC** (*Clinical Predictive Fusion Network - RDC*).

Cette architecture fédérée est conçue pour :

1. **S'intégrer aux Systèmes Nationaux :** S'interfacer directement avec les instances de l'outil de gestion sanitaire DHIS2 (*District Health Information Software 2*) déployé par le Ministère de la Santé Publique de la RDC, permettant une ingestion des données en temps réel sans surcharger les flux de travail cliniques.
2. **Améliorer la Surveillance Épidémiologique :** Agir en amont en analysant les données de santé syndromiques et communautaires pour détecter de manière précoce des anomalies statistiques représentatives de départs d'épidémies (Ebola, Mpox, Choléra), permettant ainsi des réponses de santé publique rapides et ciblées.
3. **Soutenir les Cliniciens de Première Ligne :** Fournir des outils de diagnostic d'imagerie (par exemple, classification de radiographies thoraciques pour la tuberculose et les pneumopathies sévères) utilisables sur des tablettes tactiles ou des smartphones à faible coût dans les centres de santé ruraux.
4. **Assurer une Gouvernance Éthique et Souveraine :** Les modèles d'ensemble proposés intègrent des critères d'équité algorithmique et d'explicabilité pour s'assurer que les décisions cliniques soient auditables par les autorités médicales locales, respectant la souveraineté des données de santé congolaises.

British Journal of Multidisciplinary and Advanced Studies,7(6),1-24,2026
Engineering (All Aspect)
Print ISSN: 2517-276X
Online ISSN: 2517-2778
https://bjmas.org/index.php/bjmas/index
Publication of the European Centre for Research Training and Development -UK



Fig. 12 - Architecture Fédérée d'Intelligence Sanitaire du Cadre National CPFN-RDC

Un Pipeline Complet pour une Santé Intelligente en RDC

Pour traduire ces théories mathématiques et architectures d'apprentissage automatique en outils cliniques réels et fonctionnels en RDC, nous décrivons ci-dessous un pipeline d'ingénierie ML de bout en bout, structuré et résistant aux pannes de connectivité et de qualité de données.



Fig. 13 - Pipeline d'Ingénierie d'Apprentissage Automatique pour la Santé Intelligente en RDC

A. Prétraitement Multimodal et Imputation Adaptative

- **Gestion des données manquantes :** Étant donné les lacunes fréquentes dans les DME en RDC, les techniques d'imputation classiques (remplacement par la moyenne) sont proscrites car elles déforment la variance clinique. Le pipeline intègre des algorithmes d'imputation itérative robuste par chaînes d'équations (MICE) ou des autoencodeurs débruiteurs profonds.
- **Normalisation et Harmonisation :** Correction des variations de format de données cliniques et standardisation des images médicales issues de différentes générations d'équipements de radiographie de seconde main (par normalisation d'histogrammes et recalage spatial).

B. Extraction Avancée de Caractéristiques (Feature Extraction)

- **Données Visuelles :** Utilisation de modèles de réseaux de neurones profonds pré-entraînés (ex. ResNet, EfficientNet) ajustés par transfert d'apprentissage (*transfer learning*) pour

extraire les caractéristiques des radiographies de poitrine ou des images dermatologiques (lésions Mpox).

- **Données Textuelles :** Application de modèles de traitement automatique des langues (NLP) légers et adaptés au contexte multilingue congolais (français et langues nationales comme le lingala, swahili, tshiluba, kikongo) pour extraire les concepts médicaux clés à partir des notes de consultation non structurées.
- **Données Cliniques Tabulaires :** Encodage robuste des variables catégorielles et réduction de dimensionnalité non linéaire pour les données physiologiques et de laboratoire.

C. Combinaison Adaptative via le Cadre CPFN/JoSE

- Les vecteurs de caractéristiques extraits de chaque modalité sont introduits dans le système d'ensemble.
- Pour le CPFN, le réseau de portes d'activation ajuste dynamiquement l'influence de chaque classifieur de base selon l'intégrité de la modalité associée.
- Pour JoSE, le méta-apprenant optimisé agrège de manière optimale les prédictions d'ensemble pour délivrer une probabilité diagnostique fine.

D. Calibration des Probabilités (Probability Calibration)

En médecine, une prédiction de probabilité brute d'un modèle ML n'est pas directement utilisable si elle n'est pas étalonnée. Le pipeline intègre une étape systématique de calibration de probabilité en utilisant la méthode de régression isotonique (pour les larges ensembles de validation) ou la méthode de Platt (sigmoïde de calibrage pour les ensembles de données restreints) afin de s'assurer que les scores de confiance soient cliniquement interprétables.

E. Explicabilité et Auditabilité Clinique

- Afin de lever l'effet « boîte noire » de l'IA et de garantir la conformité éthique, chaque diagnostic proposé s'accompagne d'explications locales fondées sur la théorie des jeux coopératifs via les valeurs SHAP (*SHapley Additive exPlanations*) ou la méthode LIME.
- Les variables cliniques ayant le plus d'influence positive ou négative sur la décision finale (par exemple, fièvre élevée, présence d'anémie, exposition géographique à une zone endémique) sont visuellement mises en évidence pour le praticien de santé.

- Une piste d'audit cryptographique et une journalisation sécurisée (conforme aux réglementations de protection des données de santé, s'alignant sur la loi-cadre en matière de télécommunications et des TIC en RDC) sont enregistrées pour chaque prédiction, permettant une analyse rétrospective des performances de l'outil.

F. Intégration CDSS Native et Résilience Hors-Ligne

Déploiement du pipeline optimisé sous forme de micro-services légers via des conteneurs (ex. Docker) capables de s'exécuter localement sur des serveurs hospitaliers régionaux durcis (par exemple à l'aide de mini-ordinateurs monocartes de type Raspberry Pi ou serveurs locaux durcis), garantissant un fonctionnement continu même en cas de coupure totale d'accès au réseau internet.

G. Surveillance Continue et Apprentissage Actif (Active Learning)

- Mise en place de mécanismes de surveillance de la dérive des données (*data drift*) et de la dérive des concepts (*concept drift*) pour détecter si l'évolution des profils de maladies altère la précision diagnostique.
- Un protocole de réentraînement périodique basé sur l'apprentissage actif permet d'intégrer de manière sécurisée de nouvelles données cliniques validées par des experts nationaux (INRB - Institut National de Recherche Biomédicale de Kinshasa, etc.) pour affiner continuellement l'ensemble algorithmique.

CONCLUSION

Pour que l'intelligence artificielle réalise son potentiel transformateur dans le domaine de la santé, elle doit impérativement passer du statut de curiosité de recherche en laboratoire à celui d'outil clinique fiable, hautement explicable, auditable et déployable en conditions réelles. L'apprentissage par ensemble, soutenu par des architectures robustes et mathématiquement saines telles que le CPFN et JoSE, offre une voie rigoureuse et validée pour y parvenir. En combinant intelligemment plusieurs classifieurs de base hétérogènes, ces méthodes produisent des modèles composites qui surpassent systématiquement les capacités de modèles isolés, tout en offrant la robustesse indispensable pour faire face à l'hétérogénéité des données et à la complexité des environnements cliniques réels.

Cette résilience algorithmique prend une importance vitale pour des systèmes de santé complexes comme celui de la République Démocratique du Congo. Confronté au double fardeau d'épidémies infectieuses récurrentes et de crises de santé mentale sévères exacerbées par les contextes de

conflit, le système de santé congolais peut s'appuyer sur la nature adaptative et tolérante des modèles d'ensemble. En tirant profit de ces architectures de pointe pour exploiter et valoriser des bases de données de santé parfois incomplètes ou bruitées, la RDC dispose d'un levier technologique souverain pour opérer une transition majeure : passer d'une gestion de crise réactive à un système de santé proactif, résilient et fondé sur des données probantes, au bénéfice direct des cliniciens et, en fin de coup, des patients.

Références Bibliographiques

- **[1]** K. Qian, *Harnessing AI for Improved Healthcare Outcomes*, Ph.D. dissertation, Department of Computer Science and Engineering, Fudan University, 2026.
- **[2]** T. G. Dietterich, "Ensemble Methods in Machine Learning," in *Multiple Classifier Systems, Lecture Notes in Computer Science*, vol. 1857, Springer, pp. 1–15, 2000.
- **[3]** L. Breiman, "Random Forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.
- **[4]** T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 785–794, 2016.
- **[5]** R. Caruana, A. Niculescu-Mizil, G. Crew, and A. Ksikes, "Ensemble Selection from Arbitrary Libraries of Models," in *Proceedings of the 21st International Conference on Machine Learning (ICML)*, 2004.
- **[6]** R. V. Rao, "Jaya: A simple and new optimization algorithm for solving constrained and unconstrained optimization problems," *International Journal of Industrial Engineering Computations*, vol. 7, no. 1, pp. 19–34, 2016.
- **[7]** E. S. A. El-Dahshan, T. Hosny, and A.-B. M. Salem, "Computer-aided diagnosis of human brain tumor through MRI: A survey and a new algorithm," *Expert Systems with Applications*, vol. 41, no. 11, pp. 5045–5057, 2014.
- **[8]** Organisation Mondiale de la Santé, "République Démocratique du Congo : Profil de Santé et d'Action Sanitaire," *Rapport Annuel OMS*, Bureau Régional pour l'Afrique, Brazzaville, 2024.

- **[9]** I. D. Dinarello, et al., "Using a Multimodal Ensemble Approach to Predict Child-Pugh Score for Liver Cancer," *Journal of Clinical Oncology*, vol. 41, no. 15_suppl, pp. e16120-e16120, 2023.
- **[10]** S. M. Lundberg and S.-I. Lee, "A Unified Approach to Interpreting Model Predictions," in *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 30, pp. 4765–4774, 2017.
- **[11]** Ministère de la Santé Publique de la RDC, *Plan National de Développement Sanitaire (PNDS) 2020-2024 : Vers la Couverture Santé Universelle*, Kinshasa, Gombe, 2020.