

Rapport de KADSI

Contexte:

Dans un contexte mondial marqué par une pression croissante sur les systèmes de santé, la dermatologie illustre de manière frappante les défis d'accessibilité aux soins. Les maladies de la peau figurent parmi les affections les plus répandues au monde, touchant environ **1,9 milliard de personnes** à un moment donné de leur vie, selon l'**Organisation mondiale de la santé (OMS)**. Certaines pathologies cutanées comme le mélanome, le psoriasis, la dermatite atopique ou encore les infections de la peau peuvent avoir des **conséquences graves sur la santé physique, mentale et sociale** des patients.

La **peau**, plus grand organe du corps humain, joue un rôle essentiel dans la **protection immunitaire, la thermorégulation et la perception sensorielle**. En tant que première barrière contre les agressions extérieures, elle est un marqueur direct de l'état de santé général. Reconnaisant cette importance, l'OMS a inscrit les pathologies dermatologiques dans ses **priorités de santé publique**, notamment en appelant à des solutions technologiques pour renforcer le dépistage, l'éducation et l'accès rapide à un diagnostic fiable.

Malgré cela, on observe une **pénurie mondiale de dermatologues**, accentuée dans les zones rurales, les pays en développement et même dans certaines régions urbaines. Les délais d'attente pour une consultation peuvent excéder plusieurs semaines, voire plusieurs mois, entraînant des retards de diagnostic potentiellement critiques, en particulier dans les cas de cancer de la peau.

Présentation du projet:

Technologies utilisées:

Le choix d'utiliser des modèles multimodaux de type Vision-Language Models (VLMs) repose sur plusieurs motivations essentielles :

1. **Dépasser les performances des approches actuelles**, notamment les CNN classiques, en combinant vision et langage.
2. **Améliorer la capacité de généralisation** du modèle, afin de limiter les biais liés à l'origine ethnique ou à la représentation inégale de certaines lésions dans les

bases de données.

3. **Offrir un haut niveau d'explicabilité**, en permettant au modèle de justifier ses prédictions de manière compréhensible pour le public, renforçant ainsi la transparence et la robustesse du diagnostic.

C'est dans cette optique que nous avons choisi de nous orienter vers les VLMs. Ces modèles présentent plusieurs avantages clés :

- **Intégration de données utilisateur** telles que l'âge, la localisation de la lésion, les antécédents médicaux, etc.
- **Utilisation de plusieurs images** prises sous différents angles ou à des moments distincts pour une meilleure compréhension et un meilleur suivi de l'évolution par le modèle.
- **Génération d'une analyse complète** de la lésion, accompagnée d'une explication et d'un diagnostic en langage naturel.
- **Utilisation de méthodes d'entraînement récentes** (Grpo introduit par deepseek fin 2024), qui a démontré son efficacité à améliorer les performances de raisonnement des modèles même avec un nombre limité de données.

Voici donc les spécificités exact de la technologie utilisé :

Modèle utilisé : Qwen2.5-VL 7B avec quantification 4bits, Lora avec r=16.

Méthode d'entraînement : Léger fine tuning sur quelques données formaté aux données attendues puis renforcement learning avec Grpo.

Le Grpo n'est pas encore disponible pour les VLMs, on a donc utilisé une implémentation faite par nos soins : https://github.com/GAD-cell/VLM_GRPO

Résultats:

Classifier	Three-way accuracy
Dermatologist	65 %
SOTA models	72%
Ours - 30 rl steps	53%
Ours - 250 rl steps	70%

Tableau 1 : résultat comparatifs en accuracy de notre méthode par rapport à ce qui existe

On note que :

- Le **modèle SOTA CNN** atteint la meilleure performance (72%), légèrement au-dessus du niveau d'un dermatologue humain (65%).
- Notre VLM, avec **seulement 250 étapes de fine-tuning via reinforcement learning**, approche les performances du modèle SOTA avec **70% de précision**.
- À seulement **30 steps**, la performance chute à 53%, ce qui est attendu compte tenu du faible entraînement.

Les **CNN (SOTA)** sont des modèles visuels spécialisés, entraînés spécifiquement sur des tâches de classification d'images médicales, avec une architecture optimisée pour la vision seule.

Les **VLM (Vision-Language Models)**, comme celui que nous utilisons, intègrent à la fois le traitement d'image et le langage naturel. Ils sont beaucoup plus génériques et permettent un **raisonnement multimodal**.

Cela implique que nos VLM sont **moins optimisés au départ pour la classification stricte**, mais ont **plus de potentiel** pour des tâches de diagnostic explicatif ou d'interprétation clinique avec un entraînement complet et approprié pour ce type de modèle (augmenter significativement le nombre de steps et améliorer le traitement des données notamment en équilibrant le type de peaux ET les types de lésions).

Métriques d'entraînement :

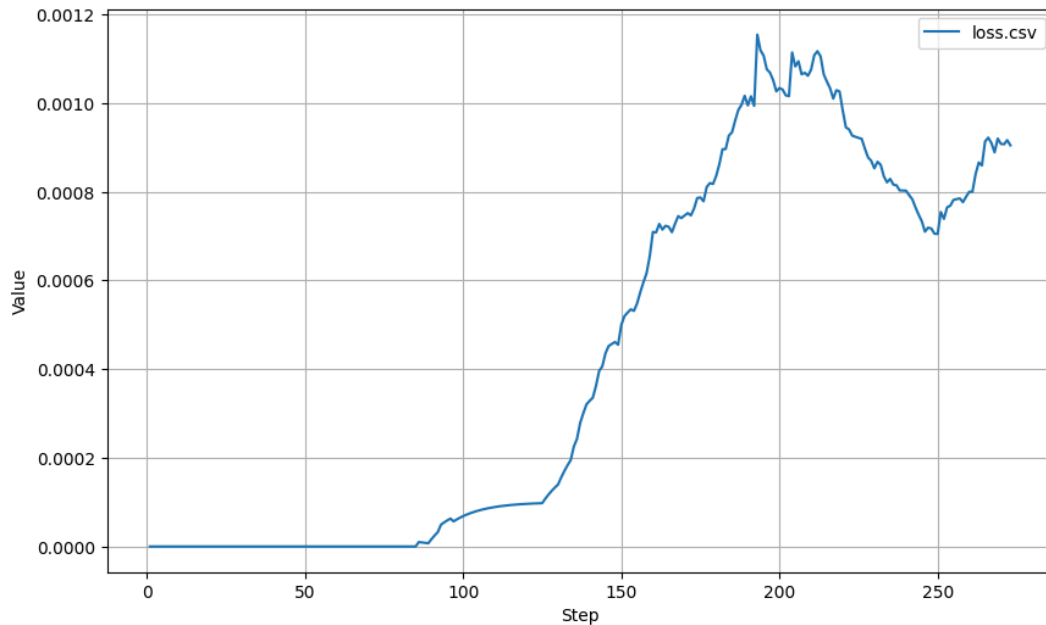


Figure 1 : évolution de la loss en fonction des steps

NB : Il est tout à fait normal d'avoir une loss à 0 au début de l'entraînement. Cela est dû à la manière dont la méthode Grpo de reinforcement learning est implémenté (il dépend de la divergence du modèle entraîné par rapport au modèle de référence, or au début, le modèle de référence est strictement égal au modèle entraîné).

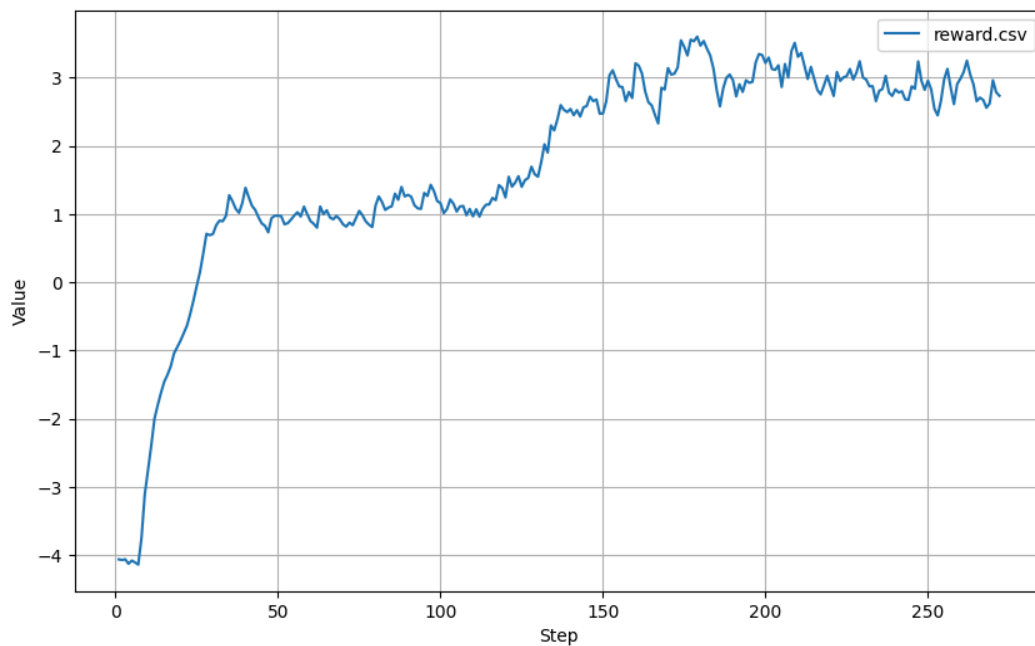


Figure 2 : évolution du reward en fonction des steps

- **Début (0–30)** : Le modèle apprend rapidement, la reward passe de ~ -4 à ~ -1 .
- **Milieu (30–130)** : Phase de stagnation, l'apprentissage ralentit, l'agent explore.
- **Reprise (130–180)** : Nouvelle amélioration significative — meilleure stratégie trouvée.
- **Fin (180–270)** : À nouveau une phase de stagnation

Globalement, l'entraînement est **positif** : la reward augmente, ce qui indique une bonne convergence de l'algorithme.

Discussion sur les résultats :

L'entraînement réalisé jusqu'à présent demeure bien en deçà des standards habituels en apprentissage par renforcement pour ce genre de modèles. Ceci s'explique par le peu d'entraînement qui a été effectué. Un entraînement sur plusieurs milliers d'étapes serait nécessaire pour exploiter pleinement les capacités du modèle et atteindre ses performances maximales. Le fait d'avoir déjà obtenu des résultats comparables à l'état de l'art, tout en dépassant la performance moyenne d'un dermatologue, est particulièrement prometteur. D'autant plus que notre approche, basée sur un VLM, offre un avantage majeur en termes d'explicabilité : elle permet de justifier ses prédictions, contrairement aux CNN classiques qui restent des boîtes noires.

Articles scientifiques et travaux académiques

- **Huang, Leo.** *"Detecting Skin Cancer with Deep Learning"*, AI4Health (2024).
[Lien PDF](#)
- **Cheng, Ting et al.** *"Skin Lesion Classification with Ensembles of Deep Convolutional Neural Networks"*, arXiv (2021).
[arXiv:2104.09957](https://arxiv.org/abs/2104.09957)
- **Adamson, A. S., & Smith, A. (2018).** *Machine Learning and Health Care Disparities in Dermatology*, **JAMA Dermatology**.
→ Cet article souligne les risques de reproduction ou d'amplification des inégalités de santé à travers des biais présents dans les ensembles de données dermatologiques.

- **Daneshjou, R. et al. (2021).** *Lack of diversity in dermatologic datasets*, **Nature Medicine**.
[Lien Nature](#)
→ Étude critique mettant en lumière la sous-représentation des peaux foncées dans les bases de données dermatologiques, avec des implications importantes sur la performance des algorithmes.
- **Mayaki, Karimatou (Mémoire de Master).** *Marketing ethnique : Les marques cosmétiques répondent-elles aux besoins des consommatrices africaines et afro-descendantes en France ?*
Master Management international de la Parfumerie et de la Cosmétique –
Tuteur : Lydie-Marie Lavoisier.
→ Pertinent pour la compréhension des enjeux liés à la représentation des peaux foncées dans les secteurs de la santé et de la beauté.

Bases de données dermatologiques

- **Fitzpatrick17k**
Base de données annotée selon les phototypes de peau (I à VI), permettant l'analyse des biais de performance des modèles selon la couleur de peau.
- **ISIC Archive (International Skin Imaging Collaboration)**
Base de données dermatologique standardisée, couramment utilisée pour l'entraînement et l'évaluation de modèles d'IA en dermatologie.
- **DermNet NZ**
Référentiel riche d'images dermatologiques accessibles publiquement, incluant dans certains cas des annotations relatives au type de peau.

Autres ressources

- **PubMed**
Utilisé comme source de références médicales validées pour l'élaboration du contenu de l'**espace professionnel de santé** intégré dans l'application.

Axes complémentaires explorés

- Étude des **lésions élémentaires et secondaires sur peaux noires**, incluant des cas spécifiques comme le **psoriasis**, l'**eczéma**, l'**acné**, les **chéloïdes**, etc.
→ Un travail d'identification de sources fiables reste en cours pour compléter cette section avec des auteurs médicaux spécialisés.