

DÉTECTION DE LA DISTRACTION AU VOLANT, PAR LA GÉOMÉTRIE DES POSES DU CORPS

Abakar Abbas Abakar et Pr. Yacine Yaddaden



Département de mathématiques, informatique et génie
Université du Québec à Rimouski



Problématique

- ✗ Selon l'OMS, près de **1,35 million** de décès sont causés chaque année par des accidents de la route [1].
 - ✗ La NHTSA estime que **25 %** des accidents impliquent une forme d'inattention du conducteur [2].
 - ✗ Les approches CNN actuelles sont précises mais **coûteuses**, peu interprétables et nécessitent de grandes quantités de données.
- Comment détecter la distraction au volant à l'aide d'un modèle léger, interprétable et rigoureusement évalué ?

Objectifs

- ✓ Concevoir un pipeline **léger et non intrusif** basé sur la géométrie du haut du corps.
- ✓ Reformuler le problème en **classification binaire** (normal vs. distrait) pour répondre aux contraintes réelles de sécurité.
- ✓ Évaluer et valider la **généralisation inter-datasets** selon un protocole **LOSO-CV** (*Leave-One-Subject-Out Cross-Validation*) strict sur *StateFarm* et *AUCDD*.

Jeux de données

Les bases de données utilisées dans ce travail sont :

- ❖ **StateFarm (Kaggle)** : 4 850 images équilibrées réparties en 10 classes de comportements (c0 à c9), issues d'une caméra embarquée en véhicule réel. [3]



FIGURE 1 – c0 — Conduite normale



FIGURE 2 – c1 — SMS main droite



FIGURE 3 – c2 — Téléphone main droite

- ❖ **AUCDD — American University in Cairo Distracted Driver** : 6 480 images équilibrées couvrant les mêmes 10 classes, collectées à partir de deux vues complémentaires (véhicule réel et simulateur), avec 40 conducteurs. [1]



FIGURE 4 – c1 — SMS main droite



FIGURE 5 – c5 — Ajuster la radio

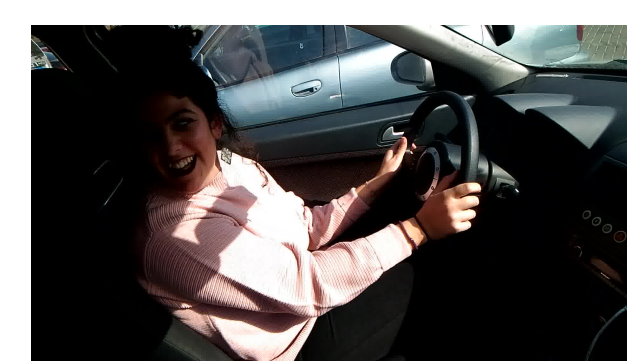


FIGURE 6 – c9 — Parler à un passager

Méthodologie

- **Pré-traitement** : chaque image RGB est traitée par **HRNet** afin de détecter **7 points clés** du haut du corps (nez, épaules, coudes et poignets).
 - Les détections avec un score de confiance $c_i < 0.7$ sont rejetées.
 - Les images contenant moins de 4 points valides sont écartées.
 - Les coordonnées (x, y) des points clés sont extraites pour l'étape suivante.
- **Extraction de caractéristiques** : un vecteur $f \in \mathbb{R}^{31}$ est construit pour chaque image à partir de :
 - **21 distances euclidiennes** entre paires de points, normalisées par la distance inter-épaules pour assurer l'invariance morphologique.
 - **10 angles articulaires** (ex. épaule–coude–poignet, nez–épaule–coude), capturant la posture du conducteur.
- **Classification binaire** : le problème est formulé en **Normal vs. Distrait**.
 - Un **Random Forest** de 500 arbres est entraîné sur les vecteurs de caractéristiques.

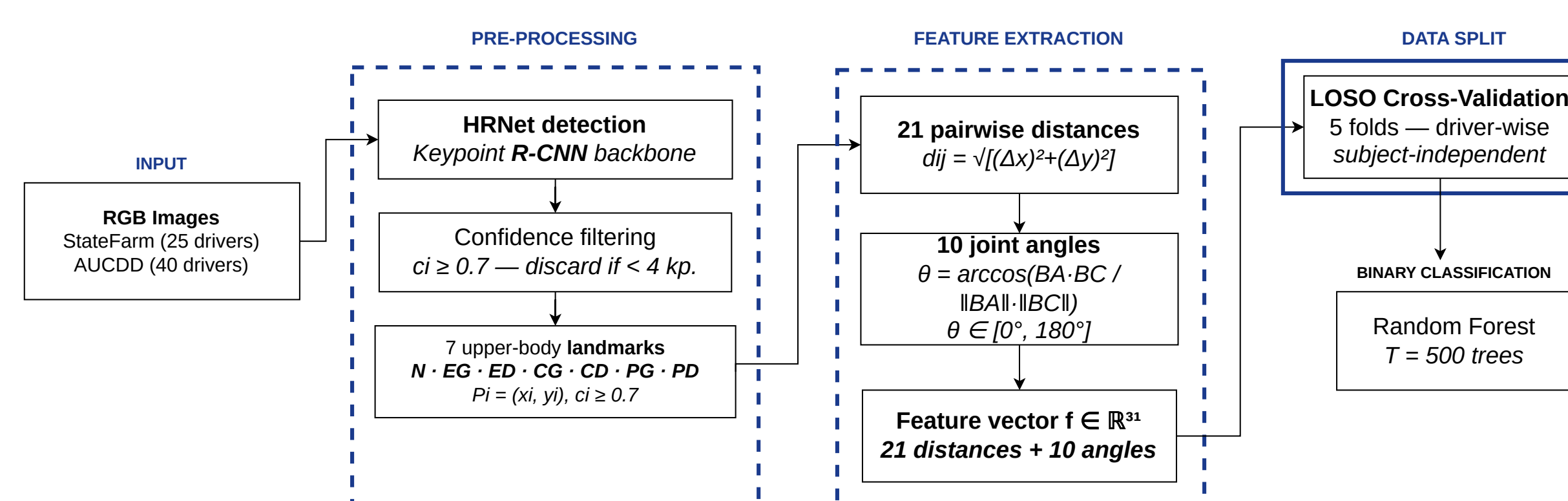


FIGURE 7 – Schéma de la méthodologie proposée

Résultats

- ☛ Sur **StateFarm**, le modèle atteint **94,60 %** d'exactitude et un AUC-ROC de **0,9806**, avec un rappel équilibré entre les classes (Normal = 94,62 %, Distrait = 94,60 %).
- ☛ Sur **AUCDD**, le modèle atteint **83,17 %** d'exactitude et un AUC-ROC de **0,9028**, confirmant la robustesse des caractéristiques géométriques face à la variabilité inter-caméras.
- ☛ En généralisation **inter-datasets**, **AUCDD** → **StateFarm** atteint **92,19 %**, contre **77,94 %** pour **StateFarm** → **AUCDD**, soulignant l'importance de la diversité du domaine d'entraînement.

Évaluation

- **Protocole LOSO-CV à 5 folds** : aucun conducteur commun entre entraînement et test, garantissant une évaluation indépendante du sujet.
- ✗ **StateFarm** : 25 conducteurs → 5 groupes de 5.
- ✗ **AUCDD** : 40 conducteurs → 5 groupes de 8.
- **Métriques utilisées** : Exactitude, Précision, Rappel (Normal et Distrait), F1-score et AUC-ROC.

Matrices de confusion moyennes (5 folds) :

		StateFarm	
		Prédit Normal	Prédit Distrait
Réel Normal		94,6 %	5,4 %
Réel Distrait		5,4 %	94,6 %

		AUCDD	
		Prédit Normal	Prédit Distrait
Réel Normal		83,8 %	16,2 %
Réel Distrait		17,5 %	82,5 %

Conclusion

- ☐ Le pipeline **HRNet + Random Forest** atteint **94,60 %** sur *StateFarm* et **83,17 %** sur *AUCDD* sous un protocole **LOSO-CV** strict, sans nécessiter de GPU.
- ☐ Malgré sa simplicité, l'approche reste **compétitive face aux CNN**, avec une généralisation inter-datasets de **92,19 %** sans réentraînement.
- ☐ Cette approche constitue une alternative légère et non intrusive, adaptée à un déploiement embarqué en temps réel.
- ☐ L'absence d'information temporelle limite la capture des comportements ; l'ajout de **LSTM/Transformer** et de **landmarks faciaux** est une perspective clé.

Références

- [1] M. Moustafa H. Eraqi Y. Abouelnaga M. Saad. Driver distraction identification with an ensemble of convolutional neural networks. *Journal of Advanced Transportation, Machine Learning in Transportation (MLT) Issue*, 2019.
- [2] Fabio Tango. Real-time detection system of driver distraction using machine learning. *IEEE Transactions on Intelligent Transportation Systems*, 2013.
- [3] Xizhong S. Yingjie G. An algorithm for distracted driving recognition based on pose features and an improved knn. 2024.