

**EY Open Science AI & Data Challenge 2025 : Cooling Urban
Heat Islands**

**PRÉSENTATION DES
RESULTATS**



Enzo Sebiane
enzo.sebiane@orange.fr

24/04/2025



SOMMAIRE

1. INTRODUCTION
2. PRÉPARATION ET GÉNÉRATION DES DONNÉES
3. DISCUSSIONS/AMÉLIORATIONS DES RÉSULTATS
4. CONCLUSION

Introduction



Situation Problème :

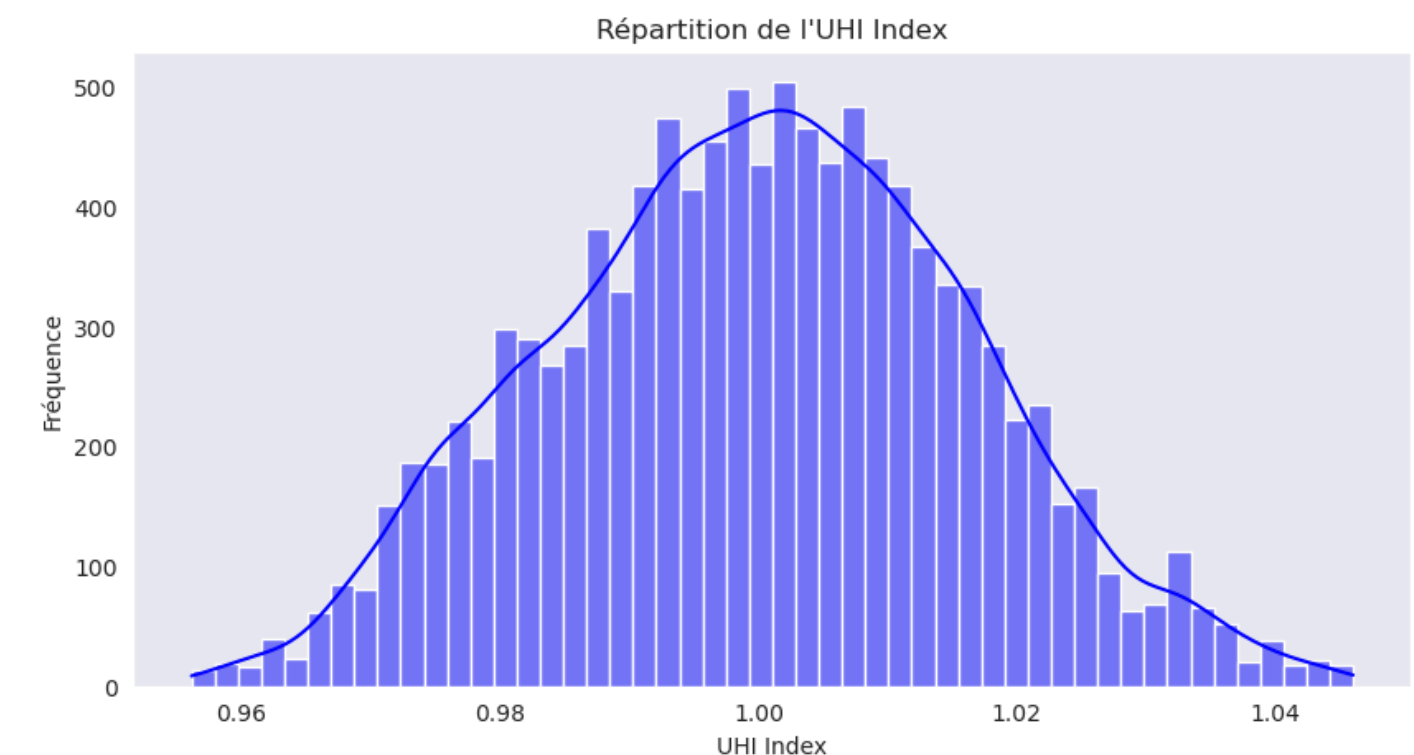
- Le challenge EY 2025, organisé dans le cadre des ODD de l'ONU, vise à mobiliser l'IA pour lutter contre les îlots de chaleur urbains.
- L'objectif est de développer un modèle prédictif à partir de données de température à New York, afin d'identifier les zones les plus exposées et les facteurs de risque, avec pour but d'apporter des solutions durables et innovantes face à l'urgence climatique.

Données fournies :

- Données collectées sur une fenêtre temporelle d'une heure
- **Indice UHI** = Température locale / Température moyenne



*Carte UHI de New York City
(issue de l'énoncé)*



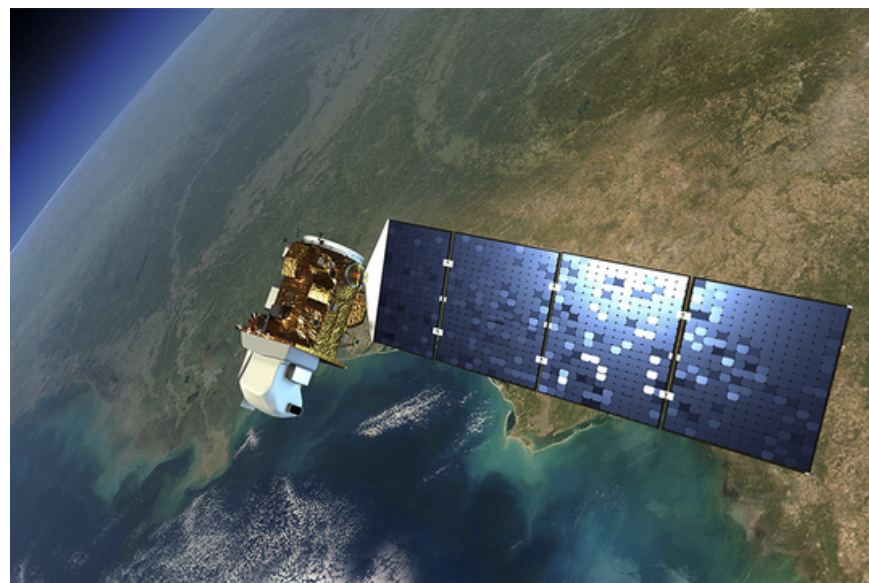
Préparation et génération des Données



Prise en compte des deux satellites indiqués :

Satellite Landsat

- Meilleurs résultats pour moins de données comparé au satellite Sentinel, d'où le choix des données de Landsat comme features “principales”.
- Filtrage des images avec moins de 10% de nuages, moyenne sur les années depuis 2013
- Extraction des bandes différentes bandes (LWR11 + rgb, near infrared).
- Précision sur une RandomForest : **~80 %**



<https://landsat.gsfc.nasa.gov/satellites/>

Satellite Sentinel

- Combinaison avec Landsat → plus de valeurs uniques dans le jeu d'entraînement.
- Sélection limitée à certaines bandes Sentinel : B01, B04, B06, B08 + moyenne des images à moins de 5% de nuages
- Maintien d'un ratio 70% Landsat / 30% Sentinel pour la cohérence des features.
- Précision sur une RandomForest : **~84 %**



https://www.esa.int/Lancement_du_satellite_Sentinelle_2B

Choix du modèle



Premiers essais : Bibliothèque de Scikit-Learn



XGBoost

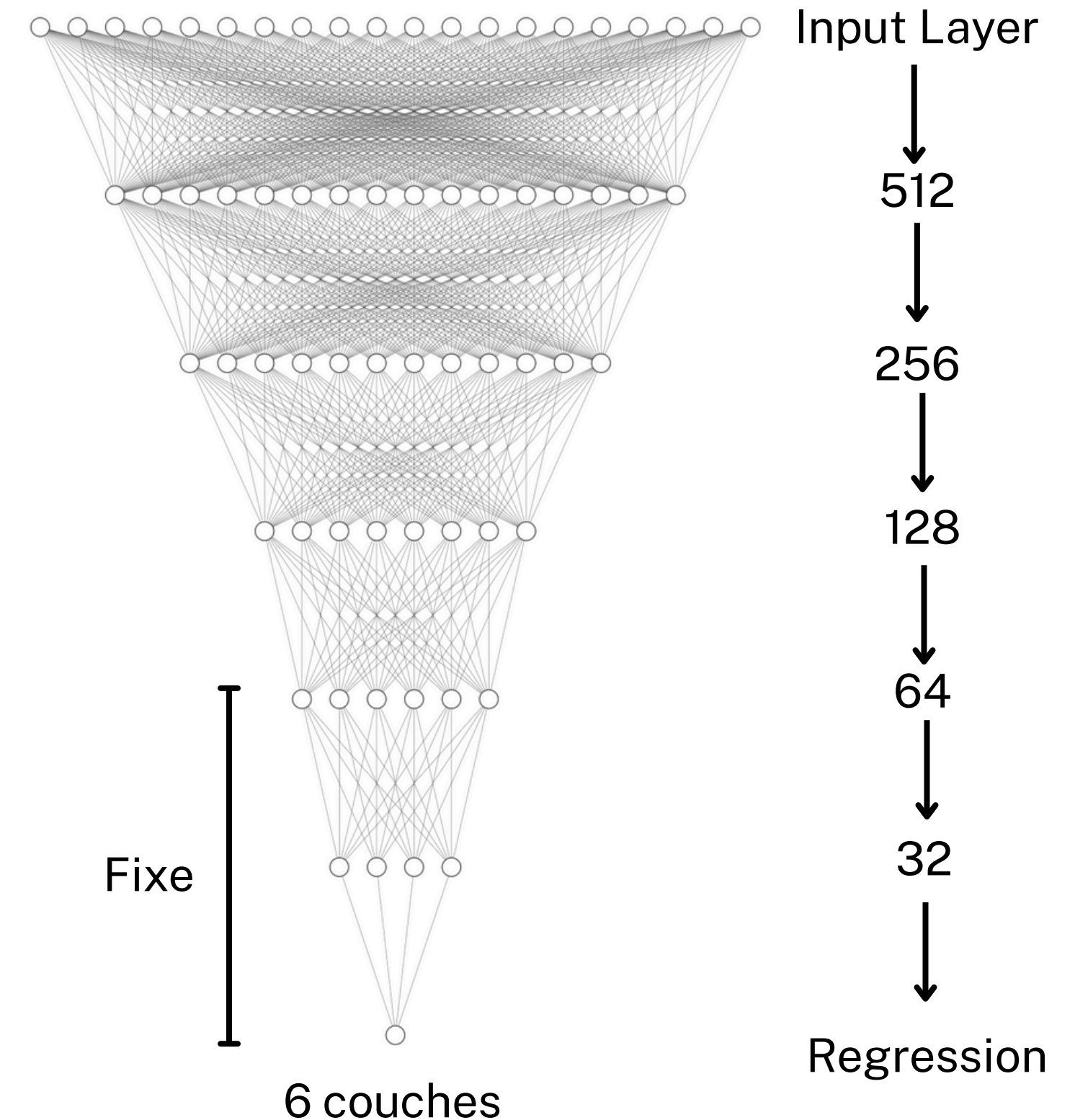
Modèles Rapides donnant de bonnes intuitions sur les features les plus importantes

Modèle personnalisé : Dense Neural Network (DNN).

- Facile à implémenter et rapide
- Editable (nombre de couches, neurones, learning rate..)

Nombre de paramètres : ~ **350 000**

Schéma simplifié :



Analyse des Résultats, Discussions



Précision à ce niveau : ~87 %.

Changement de modèle ? Deux idées d'architecture pour améliorer nos résultats :

	CNN (Convolutional Neural Network)	RNN (Recurrent Neural Network / LSTM)
Idée principale	Traiter les features comme une image 2D	Traiter les features comme une séquence temporelle
Avantage technique	Capte les patterns spatiaux entre bandes (ex: corrélations)	Capte les dépendances entre features dans l'ordre
Complexité du prétraitement	Élevée : reshaping en "pseudo-image"	Élevée : structuration en séquences temporelles
Améliorationn des performances	Pas significativement meilleur que DNN	Trop coûteux pour peu de gain

Conclusion : Le réseau dense semble suffisant et on va chercher à compenser l'absence de relations explicites dans les données par l'ajout de **contexte**.

Ajout de contexte

Prise en compte de la densité des bâtiments :

- Utilisation du fichier BuildingFootprint.kml
- Polygones élargis avec un buffer pour couvrir la totalité Manhattan
- Conversion en image raster : carte d'occupation du sol

Résultat :

- 512 clusters
- Approche technique : apprentissage non supervisé pour identifier automatiquement les structures dans les données
- Impact RSE : faible variabilité des clusters dans les zones vertes (ex. Central Park)

Précision à ce niveau : ~90 %.

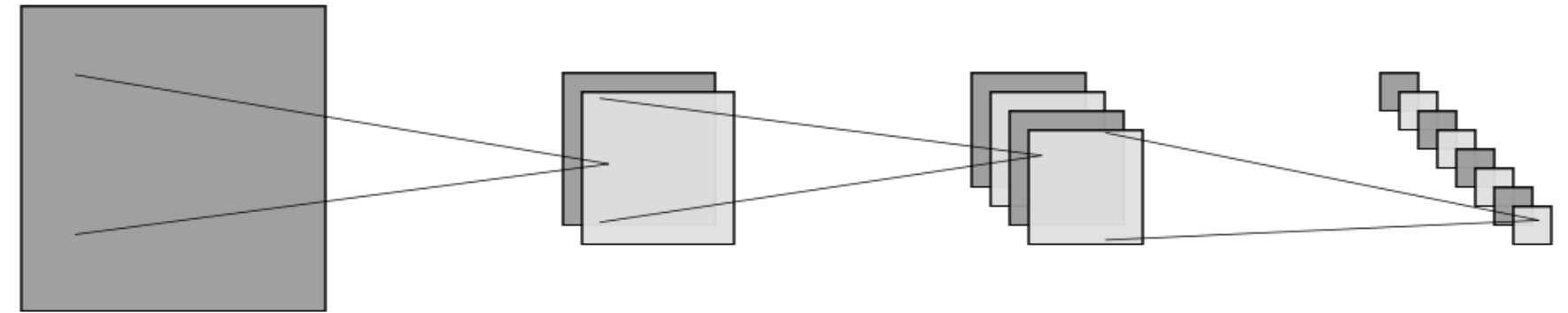


Ajout de contexte



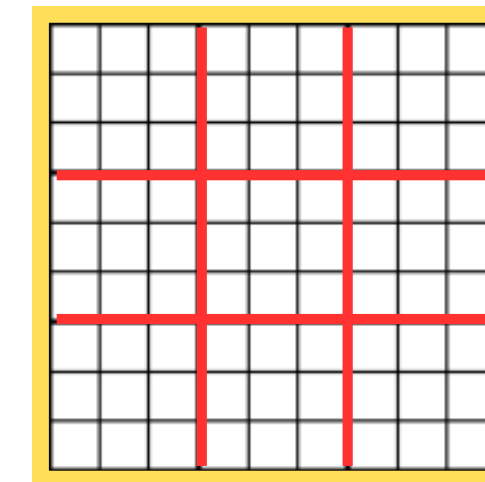
Génération Multi-Résolution : **1, 20, 30, 50 et 80 mètres/pixel**

- Volume de données multiplié par 4
- Répartition 70-30 entre Landsat et Sentinel conservée
- Ordre dans l'implémentation des données



Intérêts :

- Résolutions fines → plus de précision (~9 000 lignes de features uniques pour ~10 000 points d'entraînement)
- Proximité spatiale conservée à haute résolution, fusionnée à basse résolution



Légende :

- : Petite échelle
- : Moyenne échelle
- : Grande échelle

Précision à ce niveau : **~96.5 %.**

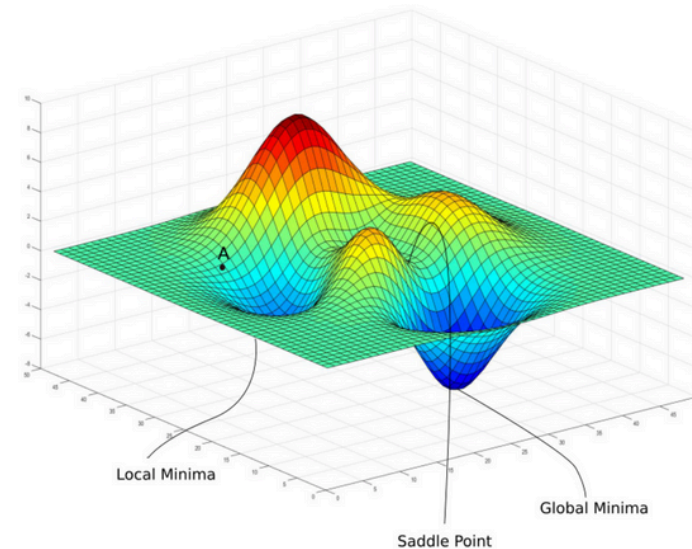
Fine Tuning



Objectif :

- Gagner ~1 % de précision sans changer de modèle

Utilisation de la bibliothèque Optuna :

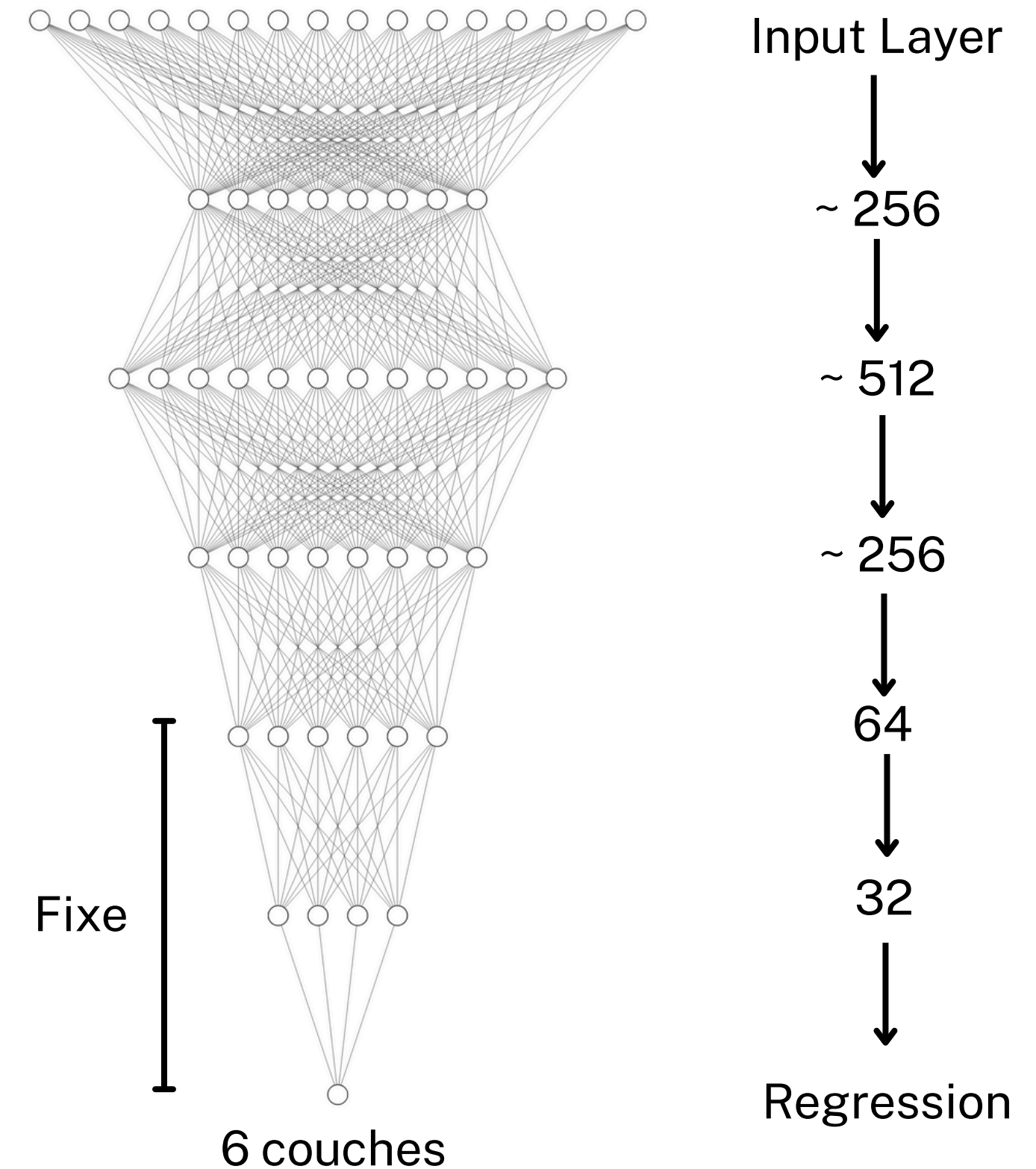


- Recherche bayésienne d'hyperparamètres
- Dans notre cas :

Tailles des 3 premières couches, Learning rate, Dropout

Précision Final : 97,48%.

Schéma simplifié :



CONCLUSION

- **IMPORTANCE DE CONCEVOIR DES FEATURES RICHES, PERMETTANT À UN SIMPLE DNN D'EXTRAIRE UN MODÈLE PERFORMANT.**

BILAN :

- Enrichissement technique : recherche et sélection de données, exploration de modèles
- Enjeux environnementaux : mise en lumière des effets indirects du réchauffement climatique et de l'impact de l'urbanisation

CE CHALLENGE SOULIGNE LE RÔLE POTENTIEL DES MODÈLES D'IA, SOUVENT CRITIQUÉS, COMME OUTILS PRÉCIEUX POUR LA COMPRÉHENSION ET LA LUTTE CONTRE LES DÉFIS ENVIRONNEMENTAUX ACTUELS.



Merci de votre attention !

Des questions ?