

# RAPPORT TECHNIQUE

## NEPHROEXPERT AI

Système Expert d'Analyse Clinique

*Intelligence Artificielle Explicable (XAI)*



**Application de Néphrologie Assistée par IA**

**Hackathon IA - Bootcamp Cohorte 1**

*Détection et Classification des Stades CKD*

13 février 2026



# 1. RÉSUMÉ EXÉCUTIF

NephroExpert AI est une application web intelligente développée pour assister les néphrologues dans le diagnostic et le suivi de la maladie rénale chronique (CKD). Combinant un modèle de machine learning (XGBoost) avec l'intelligence artificielle générative (GPT-4), l'application offre une double interface : un espace médecin pour l'analyse experte et un espace patient pour l'accompagnement personnalisé.

## 1.1. Innovation Majeure

L'application intègre SHAP (SHapley Additive exPlanations), une technologie d'explicabilité permettant aux médecins de comprendre mathématiquement pourquoi l'IA a prédit un stade particulier. Cette transparence est cruciale pour la confiance médicale et la validation clinique.

## 1.2. Objectifs

- Prédire le stade CKD (1 à 5) à partir de données cliniques et biologiques
- Calculer automatiquement le DFGe (Débit de Filtration Glomérulaire estimé)
- Fournir une interprétabilité complète via SHAP
- Générer des rapports techniques pour médecins et vulgarisés pour patients
- Assurer une interface bilingue médecin-patient avec IA générative

## 1.3. Résultats Clés

- Modèle XGBoost entraîné avec pipeline de prétraitement complet
- Analyse SHAP en temps réel pour chaque prédiction
- Génération automatique de rapports PDF double format (médecin/patient)
- Interface Streamlit responsive et ergonomique

## 2. CONTEXTE MÉDICAL ET SCIENTIFIQUE

### 2.1. Maladie Rénale Chronique

La CKD est définie par une altération progressive et irréversible de la fonction rénale, classée en 5 stades selon le DFG<sub>e</sub> (Débit de Filtration Glomérulaire estimé). Un diagnostic précoce permet de ralentir la progression et d'éviter la dialyse ou transplantation.

Stade	DFG <sub>e</sub> (mL/min)	Description
<b>CKD 1</b>	≥ 90	Fonction rénale normale
<b>CKD 2</b>	60-89	Insuffisance rénale légère
<b>CKD 3a</b>	45-59	Insuffisance rénale modérée
<b>CKD 3b</b>	30-44	Insuffisance rénale sévère
<b>CKD 4</b>	15-29	Insuffisance rénale très sévère
<b>CKD 5</b>	< 15	Insuffisance rénale terminale

### 2.2. Besoin d'Outils Intelligents

Face à la charge de travail croissante et au besoin de précision diagnostique, les outils d'aide à la décision basés sur l'IA deviennent indispensables. Cependant, l'opacité des modèles (effet boîte noire) limite leur adoption en milieu clinique. D'où l'importance de l'explicabilité (XAI).

## 3. ARCHITECTURE TECHNIQUE

### 3.1. Stack Technologique

#### 3.1.1. Frontend et Interface Utilisateur

- **Streamlit** - Framework Python pour applications web data science
- **CSS Personnalisé** - Design médical haute visibilité avec contrastes optimisés
- **Interface Responsive** - Adaptée desktop et tablette pour usage clinique

#### 3.1.2. Machine Learning et IA

- **XGBoost** - Modèle de classification multi-classe pour prédiction stades CKD
- **Scikit-learn** - Pipeline de prétraitement (encodage, standardisation)
- **SHAP (SHapley Additive exPlanations)** - Interprétabilité du modèle via valeurs de Shapley
- **OpenAI GPT-4o-mini** - Génération de rapports narratifs médecin et patient
- **Joblib** - Sérialisation du modèle et pipeline

#### 3.1.3. Visualisation et Reporting

- **Plotly** - Graphiques interactifs (SHAP, radar physiologique)
- **Matplotlib** - Visualisations complémentaires
- **ReportLab** - Génération de rapports PDF professionnels

#### 3.1.4. Traitement de Données

- **Pandas** - Manipulation des DataFrames patients
- **NumPy** - Calculs numériques et transformations

### 3.2. Architecture de l'Application

L'application suit une architecture modulaire en 6 composants principaux :

1. **Configuration et Design** - Initialisation Streamlit, CSS médical, gestion clé API OpenAI
2. **Logique Médicale** - Calcul DFGe (formule CKD-EPI), chargement modèle ML
3. **Analyse SHAP** - Interprétabilité en temps réel avec visualisation Plotly
4. **Génération IA** - Rapports narratifs via GPT-4 (médecin expert + patient vulgarisé)
5. **Interface Médecin** - Saisie données, analyse experte, rapport technique PDF
6. **Interface Patient** - Vulgarisation résultats, conseils, rapport simplifié PDF

## 4. MODÈLE DE MACHINE LEARNING

### 4.1. Choix de l'Algorithme : XGBoost

XGBoost (eXtreme Gradient Boosting) a été sélectionné pour ses performances supérieures sur les problèmes de classification médicale, sa robustesse aux données manquantes et sa compatibilité native avec SHAP pour l'explicabilité.

#### 4.1.1. Avantages pour le Diagnostic CKD

- Gestion native des valeurs manquantes (fréquentes en données cliniques)
- Capture des interactions non-linéaires entre biomarqueurs
- Robustesse au déséquilibre de classes (certains stades CKD rares)
- Performance supérieure aux modèles linéaires classiques

### 4.2. Variables d'Entrée (Features)

Le modèle utilise 16 variables cliniques et biologiques :

Variable	Type / Unité
Âge	Numérique (années)
Sexe	Binaire (0=M, 1=F)
Créatinine	mg/dL
Urée	mg/dL
Tension Artérielle Systolique	mmHg
Tension Artérielle Diastolique	mmHg
Hypertension	Binaire (0/1)
Diabète Type 1 / Type 2	Binaire (0/1)
Tabagisme / Alcool	Binaire (0/1)
Consommation AINS	Binaire (0/1)
Phytothérapie	Binaire (0/1)
Hématurie	Binaire (0/1)
Albuminurie	Binaire (0/1)
Diurèse	Encodée (0=Normale, 1=Altérée)

### 4.3. Pipeline de Prétraitement

Un pipeline Scikit-learn assure la cohérence du prétraitement entre entraînement et prédiction :

7. **Imputation** - Remplacement valeurs manquantes par médiane (numériques) ou mode (catégorielles)
8. **Encodage** - One-Hot Encoding pour variables catégorielles
9. **Standardisation** - StandardScaler pour normalisation des variables numériques

### 4.4. Déploiement du Modèle

Le modèle entraîné et le pipeline sont sérialisés via Joblib :

- **xgboost\_ckd\_pipeline.joblib** - Pipeline complet (prétraitement + modèle)
- **features\_list.joblib** - Liste ordonnée des features attendues

## 5. INTERPRÉTABILITÉ VIA SHAP

### 5.1. Principe des Valeurs de Shapley

SHAP (SHapley Additive exPlanations) est une méthode d'explicabilité issue de la théorie des jeux coopératifs. Pour chaque prédiction, SHAP calcule la contribution marginale de chaque variable en simulant toutes les coalitions possibles de features. Cette approche garantit des propriétés mathématiques fortes (additivité, cohérence locale).

### 5.2. Implémentation dans NephroExpert AI

Pour chaque patient analysé, l'application génère en temps réel :

- **TreeExplainer SHAP** - Optimisé pour XGBoost, calcul rapide (<1s)
- **Valeurs SHAP individuelles** - Contribution de chaque variable au stade prédit
- **Visualisation Plotly** - Graphique horizontal interactif des Top 10 features

#### 5.2.1. Flux de Traitement SHAP

10. Transformation des données patient via le pipeline de prétraitement
11. Extraction des noms de features après transformation
12. Initialisation du TreeExplainer avec le modèle XGBoost
13. Calcul des valeurs SHAP pour le patient unique
14. Tri des features par importance absolue
15. Génération du graphique Plotly (Top 10) et export DataFrame complet

### 5.3. Valeur Clinique de l'Explicabilité

L'intégration de SHAP apporte plusieurs bénéfices majeurs :

- **Confiance médicale** - Le médecin comprend le raisonnement de l'IA
- **Validation clinique** - Détection d'incohérences (ex: prédiction dominée par variable secondaire)
- **Apprentissage** - Identification de patterns non évidents (interactions biomarqueurs)
- **Conformité réglementaire** - Répond aux exigences d'explicabilité des dispositifs médicaux IA



## 6. INTÉGRATION GPT-4 POUR GÉNÉRATION DE RAPPORTS

### 6.1. Architecture IA Générative

L'application intègre OpenAI GPT-4o-mini via API pour générer deux types de rapports narratifs distincts, adaptés au niveau d'expertise du destinataire.

#### 6.1.1. Rapport Médecin (Analyse Expert)

Génère une interprétation clinique approfondie incluant :

- Synthèse du profil biologique et hémodynamique
- Justification médicale du stade CKD prédit
- Analyse des facteurs de risque contributifs
- Recommandations de prise en charge et examens complémentaires

#### 6.1.2. Rapport Patient (Vulgarisation)

Traduit les résultats techniques en langage accessible avec :

- Explication simple de la créatinine, urée, tension artérielle
- Signification concrète du DFGe et du stade CKD
- Conseils hygiéno-diététiques personnalisés
- Ton rassurant et pédagogique adapté au grand public

### 6.2. Gestion de la Clé API

L'application permet une configuration sécurisée de la clé OpenAI via la sidebar Streamlit. Deux modes de fonctionnement :

- **Mode IA activé** - Clé API fournie, génération GPT-4 opérationnelle
- **Mode Fallback** - Sans clé API, rapports génériques pré-formatés

### 6.3. Gestion des Erreurs et Robustesse

Mécanismes de sécurité implémentés :

- Try-catch sur appels API pour capturer échecs réseau
- Messages d'avertissement clairs si IA indisponible
- Dégradation gracieuse : contenu fallback si GPT-4 échoue

## 7. FONCTIONNALITÉS PRINCIPALES

### 7.1. Calcul Automatique du DFGe

L'application calcule automatiquement le Débit de Filtration Glomérulaire estimé (DFGe) selon la formule CKD-EPI 2021, référence internationale en néphrologie. Cette formule prend en compte la créatinine, l'âge et le sexe du patient.

#### 7.1.1. Formule CKD-EPI

$$\text{DFGe} = 142 \times \min(\text{Scr}/\kappa, 1)^{\alpha} \times \max(\text{Scr}/\kappa, 1)^{-1.200} \times 0.9938^{\text{âge}} \times [1.012 \text{ si femme}]$$

Où :

- Scr = créatinine sérique en mg/dL
- $\kappa$  = 0.7 (femmes) ou 0.9 (hommes)
- $\alpha$  = -0.241 (femmes) ou -0.302 (hommes)

### 7.2. Prédiction du Stade CKD

Le modèle XGBoost prédit le stade CKD (1 à 5) en analysant l'ensemble des 16 variables cliniques. La prédiction s'affiche avec 3 métriques clés :

- **Stade Prédit** - Classification CKD 1-5
- **DFGe** - Valeur numérique en mL/min
- **État Patient** - Catégorisation qualitative (Stable/Surveillance/Critique)

### 7.3. Analyse SHAP Interactive

Graphique Plotly horizontal montrant les 10 variables les plus influentes avec code couleur (rouge = pousse vers stade élevé, bleu = pousse vers stade faible). Permet au médecin de valider cliniquement la cohérence de la prédiction.

### 7.4. Génération de Rapports PDF

#### 7.4.1. Rapport Médecin Technique

Contenu structuré incluant :

- En-tête avec date, identité médecin et patient
- Bilan biologique complet
- Résultats prédictifs (Stade, DFGe)
- Tableau SHAP des 8 variables les plus importantes
- Analyse narrative générée par GPT-4
- Recommandations médicales

#### 7.4.2. Rapport Patient Vulgarisé

Format simplifié avec :

- Salutation personnalisée
- Explication pédagogique du DFGe et du stade
- Résumé des biomarqueurs en langage simple
- Texte narratif GPT-4 adapté au patient
- Conseils hygiéno-diététiques personnalisés

## **7.5. Interface Double (Médecin/Patient)**

Système d'onglets Streamlit permettant navigation fluide entre espace expert et espace patient. Les résultats de l'analyse médecin sont stockés en `session_state` pour affichage automatique dans l'onglet patient.

## 8. DESIGN ET ERGONOMIE MÉDICALE

### 8.1. Charte Graphique Haute Visibilité

L'interface adopte un design médical professionnel avec contrastes optimisés pour lisibilité en environnement clinique.

#### 8.1.1. Palette de Couleurs

Usage	Couleur	Code Hex
Primaire (Médecin)	Bleu Marine	#1E3A8A
Secondaire (Patient)	Vert Émeraude	#10B981
Alerte Critique	Rouge	#EF4444
Surveillance	Orange	#F59E0B

### 8.2. Sections Visuelles Différenciées

- **Section Médecin** - Bordure gauche bleue marine (#1E3A8A)
- **Section Patient** - Bordure gauche verte émeraude (#10B981)

### 8.3. Accessibilité et Contraste

Respect des normes WCAG 2.1 niveau AA :

- Ratio de contraste minimum 4.5:1 pour texte normal
- Typographie Arial (sans-serif) pour lisibilité optimale
- Taille de police minimum 12pt (24 half-points)

## 9. WORKFLOW UTILISATEUR

### 9.1. Parcours Médecin

- 16. **Configuration** - Saisie clé OpenAI (optionnel) dans sidebar
- 17. **Dossier Patient** - Remplissage identité (médecin, patient, âge, sexe)
- 18. **Habitudes de vie** - Sélection checkboxes (tabac, alcool, phytothérapie, AINS)
- 19. **Données Biologiques** - Saisie créatinine, urée, TA, albuminurie, diurèse
- 20. **Antécédents** - Déclaration HTA, diabète
- 21. **Analyse** - Clic "LANCER L'ANALYSE EXPERTE"
- 22. **Résultats** - Affichage métriques (Stade, DFG<sub>e</sub>, État)
- 23. **SHAP** - Visualisation graphique interprétabilité
- 24. **IA Expert** - Lecture analyse narrative GPT-4
- 25. **Export** - Téléchargement rapport médical PDF

### 9.2. Parcours Patient

- 26. **Accès** - Basculement onglet "ESPACE PATIENT"
- 27. **Accueil** - Message personnalisé avec prénom
- 28. **DFG<sub>e</sub> Vulgarisé** - "Vos reins fonctionnent à X% de leur capacité"
- 29. **Calendrier Suivi** - Tableau actions semaine par semaine
- 30. **Analyse IA Patient** - Texte GPT-4 simplifié avec conseils
- 31. **Export** - Téléchargement rapport patient PDF

## 10. LIMITES ET PERSPECTIVES

### 10.1. Limites Actuelles

#### 10.1.1. Validation Clinique

- Modèle non validé sur cohorte externe indépendante
- Absence d'étude comparative avec néphrologues experts

#### 10.1.2. Dépendances Externes

- Clé API OpenAI requise pour rapports narratifs IA
- Connexion Internet nécessaire pour GPT-4

#### 10.1.3. Scope Fonctionnel

- Pas de stockage persistant des dossiers patients
- Interface monopatient (pas de gestion de cohorte)

### 10.2. Évolutions Futures

#### 10.2.1. Améliorations Techniques

- Intégration base de données (PostgreSQL) pour historique patient
- Suivi évolutif : graphiques temporels DFG<sub>e</sub> et biomarqueurs
- Modèle de prédiction de progression CKD (régression temporelle)
- Export DICOM pour intégration PACS hospitalier

#### 10.2.2. Extensions Fonctionnelles

- Module de calcul de dose médicamenteuse adaptée au DFG<sub>e</sub>
- Intégration imagerie rénale (échographie) via vision AI
- Alerte automatique si dégradation rapide DFG<sub>e</sub>
- Interface mobile (iOS/Android) pour consultation terrain

#### 10.2.3. Déploiement et Scalabilité

- Déploiement cloud sécurisé (Azure Health Data Services, AWS HealthLake)
- Architecture multi-tenant pour usage hospitalier multi-services
- Conformité RGPD et certifications dispositif médical (marquage CE)
- API REST pour intégration SIH (Systèmes d'Information Hospitaliers)

## 11. CONCLUSION

NephroExpert AI représente une avancée significative dans l'application de l'intelligence artificielle explicable au domaine médical. En combinant un modèle de classification performant (XGBoost) avec des technologies d'interprétabilité de pointe (SHAP) et d'intelligence artificielle générative (GPT-4), l'application offre un outil complet pour l'aide au diagnostic et au suivi de la maladie rénale chronique.

La double interface médecin-patient constitue une innovation majeure, permettant à la fois une analyse experte approfondie pour les praticiens et une vulgarisation pédagogique pour les patients. Cette approche favorise l'éducation thérapeutique et l'engagement actif du patient dans sa prise en charge.

L'intégration de SHAP garantit la transparence et l'explicabilité des prédictions, répondant aux exigences croissantes de confiance et de validation clinique des outils d'IA médicale. Chaque décision algorithmique est justifiée mathématiquement, permettant au médecin de conserver son autonomie de jugement.

Développée dans le cadre du Hackathon IA - Bootcamp Cohorte 1, cette application démontre le potentiel transformateur de l'IA dans la pratique néphrologique. Elle ouvre la voie à une médecine plus précise, plus accessible et plus centrée sur le patient, tout en préservant l'expertise humaine indispensable à la décision clinique.

**Au-delà du hackathon, NephroExpert AI constitue une base solide pour un développement clinique approfondi, une validation hospitalière et un déploiement à large échelle au service de la santé rénale.**

## ANNEXES

### Annexe A : Dépendances Python

#### Liste complète des bibliothèques requises :

- streamlit - Framework web pour data science
- pandas - Manipulation de données
- numpy - Calculs numériques
- xgboost - Modèle de machine learning
- scikit-learn - Pipeline de prétraitement
- joblib - Sérialisation modèle
- shap - Interprétabilité
- plotly - Visualisations interactives
- matplotlib - Graphiques complémentaires
- reportlab - Génération PDF
- openai - API GPT-4

### Annexe B : Guide d'Installation

#### B.1. Prérequis

- Python 3.8 ou supérieur
- Clé API OpenAI (optionnel pour rapports IA)
- Fichiers modèle : xgboost\_ckd\_pipeline.joblib, features\_list.joblib

#### B.2. Installation

32. Installer les dépendances : `pip install -r requirements.txt`
33. Placer les fichiers .joblib dans le répertoire racine
34. Lancer l'application : `streamlit run app.py`

*L'application sera accessible à : <http://localhost:8501>*

### Annexe C : Contact

- **Projet** : Hackathon IA - Bootcamp Cohorte 1
- **Application** : NephroExpert AI
- **Stack** : Python, Streamlit, XGBoost, SHAP, GPT-4
- **Domaine** : Néphrologie - Maladie Rénale Chronique (CKD)

---

*Fin du Rapport Technique*