

LeukemIA

Analyse des données de leucémie

DIANE MURZI, IKRAM FARIS, GATIEN GRELLET,
RAHUL ANANDARAJAH, MATTHIS MONNIER, ADNANE TAMRI,
MATTEO LE PERA

16 au 18 mai 2025

Table des matières

1	Contexte Scientifique	3
2	Intérêt du projet	3
3	Objectifs	3
4	Méthodologie	4
5	Jeux de Données Fournis	4
5.1	Données Cliniques (clinical_train.csv)	4
5.2	Données Moléculaires (molecular_train.csv)	6
5.3	Données de Survie (target_train.csv)	7
6	Algorithmes utilisés	7
6.1	Survival Random Forest	7
6.2	Modèle de régression de Cox	8
6.3	Survival Support Vector Machine (Survival SVM)	9
6.4	Gradient Boosting Survival Analysis	9
6.5	Componentwise Boosting for Survival Data	9
6.6	Coxnet (Penalized Cox Regression)	10
7	Résultats Préliminaires	10
8	Résultats	10
9	Discussion	11
10	Impact et Perspectives	11
11	Axes d'amélioration	11
11.1	Améliorations techniques	11
11.2	Améliorations méthodologiques	12
11.3	Sources de données supplémentaires à intégrer	12
11.4	Perspectives	12
12	Conclusion	12

Résumé

Contexte : La leucémie myéloïde aiguë (LMA) est une hémopathie maligne à pronostic incertain, en raison de l'hétérogénéité clinique et génétique des patients. Les systèmes actuels de stratification du risque, tels que les recommandations ELN (European LeukemiaNet) 2022, offrent une précision prédictive limitée (indice de concordance $\sim 0,61$), conduisant potentiellement à des décisions thérapeutiques inadaptées.

Méthodes : Nous avons développé un outil de prédiction de la survie basé sur l'intelligence artificielle, intégrant des données cliniques et moléculaires issues de plusieurs centres hospitaliers. Le pipeline comprend la fusion et le nettoyage des données, la sélection automatique de variables et l'entraînement de modèles de survie supervisés, incluant notamment le gradient boosting (GBSA), le boosting composante par composante, et la régression de Cox pénalisée (Coxnet). Le modèle produit un score de risque ordonné, accompagné des variables explicatives principales et d'une courbe de survie personnalisée.

Résultats : Les résultats préliminaires sur une base multicentrique (24 centres) indiquent une performance supérieure aux outils conventionnels, avec un objectif d'indice de concordance $c\text{-index} \geq 0,80$. L'analyse en graphe des co-occurrences mutationnelles révèle que les délétions chromosomiques `del(5)`, `del(7)` et `del(20)` occupent une position centrale, suggérant leur rôle pivot dans la dynamique mutationnelle de la LMA. Le modèle est interprétable grâce à SHAP, et adapté à une utilisation clinique en temps réel.

Interprétation : Ce système d'aide à la décision clinique, explicable et non prescriptif, vise à optimiser la prise en charge des patients atteints de LMA en fournissant une estimation fiable du risque de décès. Il soutient les décisions stratégiques en réunion de concertation pluridisciplinaire (RCP) et pourrait être étendu à la recommandation thérapeutique personnalisée. Ce projet s'inscrit dans une dynamique de médecine de précision et de transformation numérique des pratiques en hématologie.

1 Contexte Scientifique

L'un des enjeux majeurs en hématologie est l'**estimation du pronostic vital dès le diagnostic**, un paramètre clé pour **adapter l'intensité du traitement** (thérapies de soutien ou greffe de cellules souches). Toutefois, cette estimation est rendue difficile par la grande variabilité des profils biologiques des patients, par la nature complexe des mutations somatiques, et par la présence fréquente de données censurées (patients toujours en vie au moment du dernier suivi).

Fonctionnement du développement des cellules sanguines :

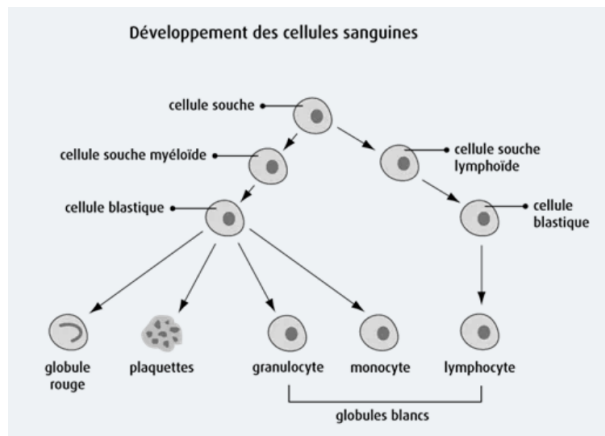


FIGURE 1 – Enter Caption

2 Intérêt du projet

L'intégration de l'intelligence artificielle (IA) dans le domaine médical offre des perspectives prometteuses pour **améliorer la prise en charge des patients**. Notamment, l'utilisation de l'IA pour analyser des données cliniques complexes peut faciliter le diagnostic, personnaliser les traitements et optimiser les ressources médicales. Dans le contexte des leucémies, qui présentent une hétérogénéité clinique et biologique significative, l'IA peut aider à identifier des patterns subtils et à prédire l'évolution de la maladie, contribuant ainsi à une médecine plus précise et efficace.

3 Objectifs

Nous visons à concevoir un outil d'aide à la décision clinique :

- Capable d'estimer de manière robuste le risque de décès.
- Intégrant des données cliniques et moléculaires de manière transparente.
- Facilement interprétable par les praticiens en contexte réel (RCP).
- Vise un indice de concordance (c-index) ≥ 0.80 , soit une nette amélioration par rapport à ELN 2022.

4 Méthodologie

Notre approche méthodologique repose sur l'analyse de **jeux de données cliniques, moléculaires et de survie fournis par QRT**. Nous avons commencé par une exploration approfondie des données pour comprendre leur structure et identifier les variables pertinentes. Ensuite, nous avons appliqué des techniques de prétraitement, telles que la gestion des valeurs manquantes et l'encodage des variables catégorielles. Pour la modélisation, nous avons utilisé des **algorithmes d'apprentissage automatique** adaptés aux données de survie, comme les **forêts aléatoires de survie** et les **modèles de Cox**. L'évaluation des performances des modèles a été réalisée à l'aide de métriques appropriées, telles que le **C-index**, pour assurer la robustesse et la pertinence des prédictions.

$$C = \frac{\sum_{i,j} \mathbf{1}(T_i < T_j) \cdot \mathbf{1}(\hat{f}_i > \hat{f}_j)}{\sum_{i,j} \mathbf{1}(T_i < T_j)}$$

- T_i, T_j : Temps de survie observés des individus i et j .
- \hat{f}_i, \hat{f}_j : Scores de risque ou prédictions associées aux individus i et j .
- $\mathbf{1}(\cdot)$: Fonction indicatrice, valant 1 si la condition est vraie, 0 sinon.

5 Jeux de Données Fournis

Pour cette étude, trois jeux de données fournis par QRT ont été utilisés :

5.1 Données Cliniques (clinical_train.csv)

Ce jeu de données contient les informations cliniques de 3323 patients, avec 9 variables incluant :

- **ID** : Identifiant unique du patient.
- **CENTER** : Centre d'analyse, potentiellement source de biais.
- **BM_BLAST** : Pourcentage de blastes dans la moelle osseuse.
- **WBC** : Nombre de globules blancs (giga/L).
- **ANC** : Nombre de neutrophiles (giga/L).
- **MONOCYTES** : Nombre de monocytes (giga/L).
- **HB** : Taux d'hémoglobine (g/L).
- **PLT** : Nombre de plaquettes (giga/L).
- **CYTOGENETICS** : Caryotype du patient selon la norme ISCN.

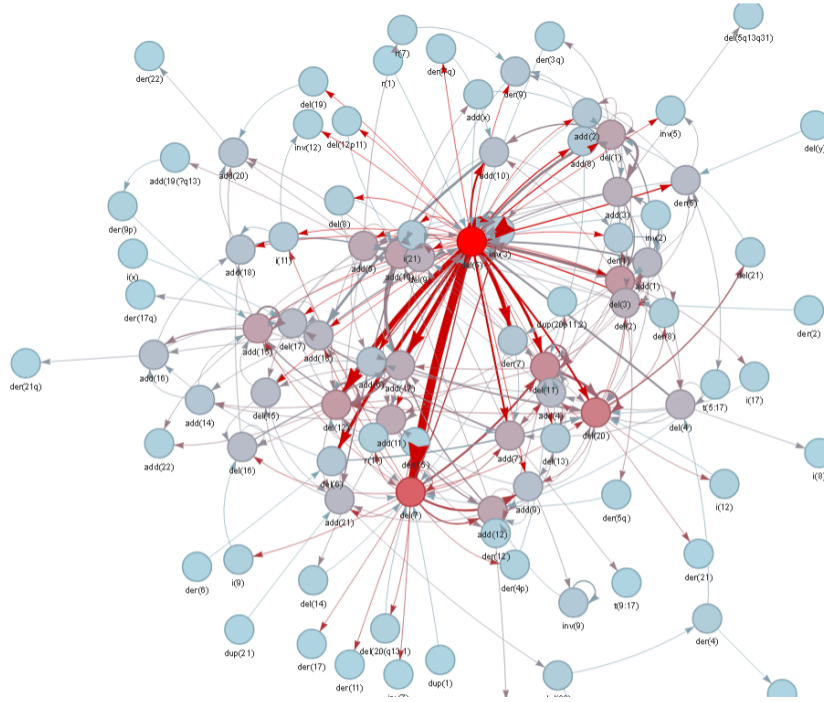


FIGURE 2 – Visualisation et Centralité des Altérations Chromosomiques

Une analyse exploratoire a été menée en R afin de représenter les relations entre les différentes mutations génétiques observées dans notre cohorte de patients atteints de LMA. Les données ont été modélisées sous la forme d'un graphe non orienté, dans lequel chaque nœud correspond à une mutation spécifique, et chaque arête représente une co-occurrence significative entre deux mutations chez un même patient.

L'analyse des propriétés topologiques du graphe a révélé que les altérations cytogénétiques **del(5)**, **del(7)** et **del(20)** présentent les degrés de centralité les plus élevés. Ces nœuds, correspondant à des délétions sur les chromosomes 5, 7 et 20, apparaissent comme fortement connectés à un grand nombre d'autres mutations. Cette observation suggère un rôle central de ces altérations chromosomiques dans la structuration des profils mutationnels des patients, et potentiellement dans la dynamique évolutive de la maladie. Leur centralité élevée dans le graphe souligne leur importance en tant que pivots biologiques, susceptibles d'influencer l'apparition ou la co-sélection d'autres mutations secondaires. Ces résultats justifient leur intégration prioritaire dans les modèles prédictifs et les analyses de survie.

Notre analyse préliminaire a révélé qu'il y a des mutations génétiques qui engrangent des disparités dans le taux de survivabilité des malades par exemple sur mutation ins (insertion, voir ci-dessous), peut importe sa localisation (quel chromosome affecté), son taux de survivabilité s'en verra grandement affecté. Ces observations nous ont permis d'identifier le besoin de stratifier les données selon les différents de mutations.

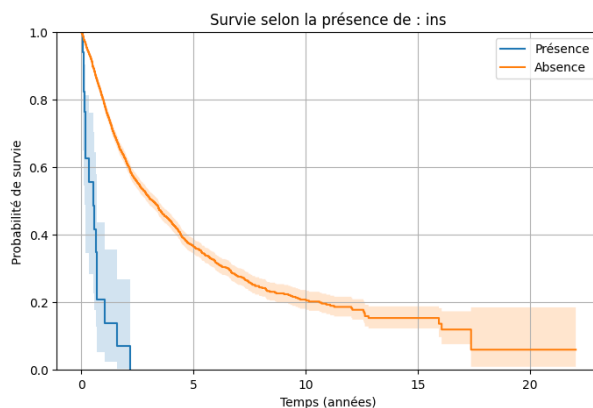


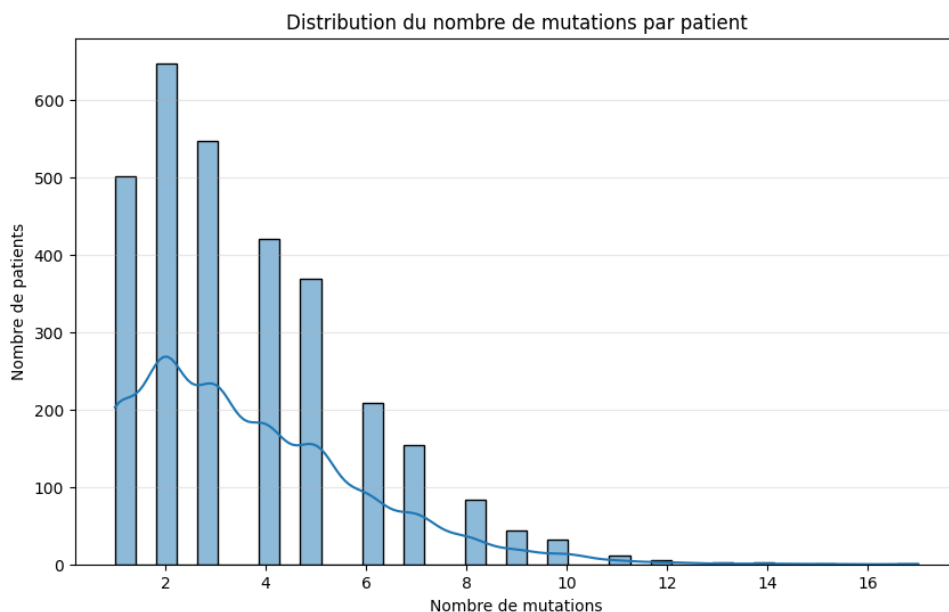
FIGURE 3 – Taux de survie en fonction de la présence ou non d’une insertion

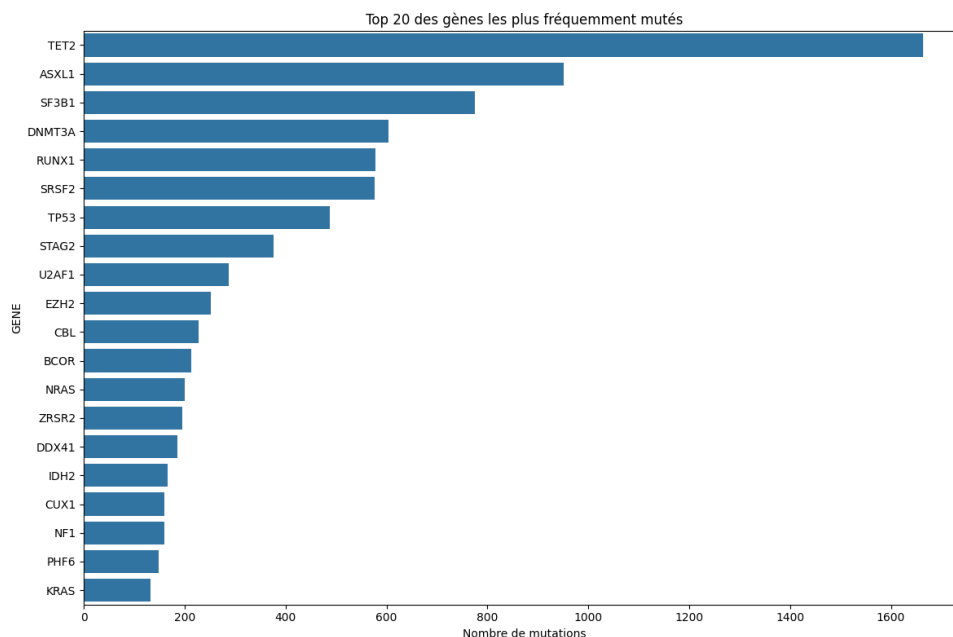
5.2 Données Moléculaires (molecular_train.csv)

Ce jeu de données comprend les mutations somatiques, avec 10935 lignes et 11 colonnes, notamment :

- **ID** : Identifiant unique du patient.
- **CHR, START, END** : Position chromosomique de la mutation.
- **REF, ALT** : Nucléotide de référence et nucléotide mutant.
- **GENE** : Gène affecté par la mutation.
- **PROTEIN_CHANGE** : Effet sur la protéine.
- **EFFECT** : Type de mutation (ex : frameshift, nonsense).
- **VAF** : Proportion de cellules portant la mutation.

L’effet des colonnes **CHR, START, END, REF, ALT, GENE** sont traduits directement par la modification de l’effet de la protéine décrite **PROTEIN_CHANGE** sous la forme d’un code déchiffrable et donnant des informations précieuses sur l’impact de ce changement protéique. Pour utiliser ces informations nous avons donc encodé numériquement cette feature de 1 à 4, 4 étant l’impact le plus grave.





L'analyse exploratoire des données, incluant la visualisation des distributions, des corrélations et l'étude de l'importance des variables, constitue une étape cruciale du pipeline. Elle permet de mieux comprendre les interactions sous-jacentes, de détecter les redondances éventuelles et de guider le prétraitement. Une préparation rigoureuse des données avant l'entraînement des modèles améliore non seulement la performance prédictive, mais également la robustesse et l'interprétabilité du modèle.

5.3 Données de Survie (target_train.csv)

Ce jeu de données contient les données de survie des patients :

- **ID** : Identifiant unique du patient.
- **OS_YEARS** : Temps de survie (en années) depuis le diagnostic.
- **OS_STATUS** : État du patient (1 = mort, 0 = vivant).

6 Algorithmes utilisés

6.1 Survival Random Forest

L'algorithme *Survival Random Forest* est une extension de la méthode Random Forest adaptée aux données de survie, où l'objectif est de modéliser le temps jusqu'à la survenue d'un événement (comme un décès, une rechute, etc.), souvent en présence de données censurées.

Chaque arbre de la forêt est construit en tenant compte du temps de survie et de la censure. Au lieu de minimiser une erreur de classification ou de régression classique, la Survival Random Forest se base sur des critères tels que la log-rank test statistic pour le découpage aux nœuds.

- Chaque arbre est construit à partir d'un échantillon bootstrap des données.
- À chaque nœud, la variable de séparation est choisie en maximisant la séparation selon le test log-rank.

- La prédiction finale est une estimation de la fonction de survie $\hat{S}(t | x)$, obtenue en moyennant les fonctions de survie des arbres.

Avantages :

- Gère naturellement la censure.
- Capable de modéliser des effets non linéaires et des interactions complexes.
- Fournit une estimation directe de la fonction de survie.

Inconvénients :

- Moins interprétable qu'un modèle paramétrique.
- Besoin de beaucoup d'arbres pour une estimation stable.

Applications en Médecine : Ce modèle est particulièrement adapté pour prédire la survie de patients en intégrant des covariables cliniques hétérogènes.

6.2 Modèle de régression de Cox

Le *modèle de Cox* (ou modèle de régression de Cox à risques proportionnels) est un modèle semi-paramétrique largement utilisé pour l'analyse de la survie. Il modélise le risque instantané de survenue d'un événement comme :

$$\lambda(t | \mathbf{x}) = \lambda_0(t) \exp(\mathbf{x}^\top \boldsymbol{\beta})$$

où :

- $\lambda(t | \mathbf{x})$ est la fonction de risque conditionnelle aux covariables \mathbf{x} .
- $\lambda_0(t)$ est la fonction de risque de base (non paramétrée).
- $\boldsymbol{\beta}$ est le vecteur des coefficients à estimer.

L'estimation de $\boldsymbol{\beta}$ se fait par maximisation de la vraisemblance partielle, sans nécessiter l'estimation explicite de $\lambda_0(t)$.

Avantages :

- Interprétation claire via les rapports de risques.
- Bonne robustesse même pour des tailles d'échantillon modestes.

Inconvénients :

- Hypothèse forte de proportionnalité des risques.
- Moins flexible pour modéliser des relations non linéaires.

Applications en Médecine : Très utilisé pour identifier les facteurs pronostiques influençant la survie.

6.3 Survival Support Vector Machine (Survival SVM)

Le *Survival SVM* est une extension du Support Vector Machine pour les données de survie. Contrairement au SVM classique qui cherche à séparer deux classes, ici, l'objectif est d'ordonner correctement les temps de survie, en tenant compte de la censure.

Il s'agit d'un algorithme d'apprentissage par rangs (learning to rank), où on cherche un score $f(\mathbf{x}) = \mathbf{w}^\top \mathbf{x}$ tel que si $T_i < T_j$ (et que ni T_i ni T_j ne sont censurés avant ces instants), alors $f(\mathbf{x}_i) > f(\mathbf{x}_j)$.

Le problème d'optimisation devient :

$$\min_{\mathbf{w}, \xi} \quad \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{(i,j) \in \mathcal{P}} \xi_{ij}$$

$$\text{sous contraintes} \quad \mathbf{w}^\top (\mathbf{x}_i - \mathbf{x}_j) \geq 1 - \xi_{ij}, \quad \xi_{ij} \geq 0, \quad \forall (i, j) \in \mathcal{P}$$

où \mathcal{P} est l'ensemble des paires ordonnées utilisables (i.e. où $T_i < T_j$ sans censure gênante).

Avantages :

- Bon compromis entre flexibilité et robustesse.
- Capable d'intégrer la censure dans le cadre SVM.

Inconvénients :

- Moins intuitif que les approches basées sur la survie classique.
- Plus coûteux computationnellement que le modèle de Cox.

Applications en Médecine : Peut capturer des relations complexes dans les données de survie tout en respectant la structure de censure.

6.4 Gradient Boosting Survival Analysis

Le *Gradient Boosting Survival Analysis* (GBSA) est une extension des algorithmes de boosting de type gradient, adaptés aux données de survie censurées. Il repose sur l'optimisation itérative d'une fonction de perte propre à l'analyse de survie, telle que le négatif du c-index. Chaque itération apprend un arbre de décision corrigé des erreurs précédentes, en pondérant davantage les individus mal prédits. Ce type de modèle, souvent implémenté via **XGBoost** ou **LightGBM**, permet de capturer des relations non-linéaires et des interactions complexes entre les variables cliniques et génétiques. Dans notre projet, il est utilisé pour construire un score de risque ordinal avec un haut degré d'interprétabilité et une précision cible c-index ≥ 0.80 .

6.5 Componentwise Boosting for Survival Data

Le *Componentwise Likelihood-Based Boosting* est une méthode adaptée aux petits échantillons et aux jeux de données de survie à haute dimension, comme ceux contenant des centaines de mutations génétiques. Contrairement au GBSA, il sélectionne à chaque itération une seule variable (composant) qui maximise la mise à jour de la fonction de risque. Ce mécanisme permet une régularisation implicite et un contrôle de la complexité du modèle, proche de la méthode Lasso mais itératif. L'approche est compatible avec le

modèle de Cox proportionnel et facilite l'identification de biomarqueurs pertinents tout en maintenant la stabilité statistique sur des jeux de données médicaux de taille limitée.

6.6 Coxnet (Penalized Cox Regression)

Le modèle *Coxnet* est une régression de Cox régularisée par des pénalités de type L1 (Lasso) et/ou L2 (Ridge), permettant la sélection automatique de variables. La fonction de risque est optimisée avec une pénalité λ contrôlant la parcimonie du modèle. Coxnet est particulièrement adapté aux données de survie à haute dimension, notamment en génomique, où le nombre de prédicteurs excède souvent le nombre de patients. En pratique, ce modèle est utile pour identifier un sous-ensemble minimal de variables prédictives tout en assurant une estimation robuste du risque relatif. Dans notre cadre, Coxnet sert de base comparative aux méthodes non-linéaires, permettant d'évaluer l'apport des interactions capturées par les modèles de boosting.

7 Résultats Préliminaires

Les premiers modèles, entraînés sur des cohortes issues de 24 centres cliniques, indiquent une meilleure généralisation que les modèles construits sur une seule cohorte. Le pipeline automatisé garantit une reproductibilité et une adaptation rapide à d'autres jeux de données cliniques.

Algorithme	C-score
Random Forest	0,71727
Gradient Boosting Survival Analysis	0,71447
Componentwise Boosting Survival Analysis	0,70960
Coxnet	0,70200

TABLE 1 – Tableau présentant les différents scores en fonction des algorithmes utilisés

8 Résultats

ALGORITHMES	PRÉCISION sur les BDD combinées
SURVIVAL RANDOM FOREST	74,7%
GBSA	74,4%
CGBSA	73,3%
COXNET	72,1%
ELN 2022	61%

TABLE 2 – Comparaison des performances de différents algorithmes de prédiction dans la leucémie myéloïde aiguë

Notre meilleur modèle est donc le Survival Random Forest avec en input les BDD combinées, avec une précision de 74,7%

9 Discussion

Notre solution se distingue par son approche orientée clinique :

- **Intégration directe** dans les workflows médicaux (API SIH, interface visuelle).
- Non-prescriptive : **soutien à la décision**, sans remplacer le jugement clinique.
- Conformité **RGPD**, **hébergement HDS**, **traçabilité complète**.

À terme, le modèle pourra être étendu à la recommandation thérapeutique personnalisée, en intégrant des cibles moléculaires spécifiques et l'impact des mutations sur la réponse au traitement.

10 Impact et Perspectives

La solution vise une double amélioration :

1. **Sur le plan médical** : prise en charge plus rapide et plus pertinente.
2. **Sur le plan économique** : réduction des hospitalisations prolongées et des traitements inadaptés.

Des partenariats sont envisagés avec les CHU, les mutuelles et les groupes privés de santé, pour un déploiement sous forme de SaaS ou API intégrée.

11 Axes d'amélioration

11.1 Améliorations techniques

- **Optimisation des hyperparamètres du RSF** :
 - Nombre d'arbres (`n_estimators`)
 - Profondeur maximale des arbres (`max_depth`)
 - Nombre de variables considérées à chaque split (`max_features`)on pourra optimiser ces paramètres via une recherche par grille (`GridSearchCV`) ou par validation croisée adaptée aux données censurées.
- **Feature selection avancée** :
 - Élimination des variables peu informatives via permutation ou importance gini.
 - Intégration de méthodes telles que Boruta ou SHAP pour identifier les variables les plus influentes.
- **Imputation des valeurs manquantes** :
 - Utilisation d'imputation multiple par chaînes de Markov (MICE).
 - Ou bien d'arbres décisionnels pour prédire les valeurs manquantes.
- **Calibration du modèle** :
 - Vérification et amélioration de la calibration des probabilités de survie.
 - Utilisation de méthodes de recalibration (par ex. isotonic regression).

11.2 Améliorations méthodologiques

- **Augmentation de la taille de l'échantillon :**
 - Plus d'observations augmenteraient la robustesse du modèle.
- **Stratification des patients :**
 - Entraîner différents modèles pour des sous-groupes (selon l'âge, le sexe, le groupe cytogénétique, etc.).
- **Validation externe :**
 - Tester le modèle sur une cohorte indépendante pour évaluer sa généralisabilité.
- **Intégration de l'expertise médicale :**
 - Créer des variables composites guidées par des hématologues (score génétique, index d'agressivité, etc.).

11.3 Sources de données supplémentaires à intégrer

- **Données longitudinales :** évolution des biomarqueurs dans le temps (HB, PLT, WBC, etc.).
- **Variables cliniques supplémentaires :**
 - Âge au diagnostic
 - Mode de vie (consommation de tabac, drogue, alcool)
 - Comorbidités (ex : diabète, hypertension)
 - Traitements reçus (chimiothérapie, greffe)
 - Réponse au traitement (RC, PR, etc.)
 - Origine ethnique et géographique
- **Données génétiques enrichies :**
 - Classification ELN
 - Type de variant (SNV, indel, fusion, etc.)
 - Evolution des biomarqueurs liés aux mutations des cellules cancéreuses
- **Imagerie :** extraction de features radiomiques à partir de scanners ou myélogrammes, dans le but d'utiliser la dégénérescence structurelle des cellules cancéreuses et connexes.

11.4 Perspectives

- Intégration d'approches deep learning (DeepSurv, DeepHit) pour modéliser des relations non linéaires plus complexes.
- Déploiement d'un outil de prédiction visuel interactif destiné aux cliniciens.

12 Conclusion

Cette étude démontre le potentiel transformateur de l'intelligence artificielle appliquée à la stratification du risque dans la leucémie myéloïde aiguë. Nos travaux ont permis d'atteindre plusieurs avancées significatives dans ce domaine. Premièrement, notre approche de modélisation prédictive, basée sur l'intégration de données cliniques et moléculaires

issues de 24 centres hospitaliers, a généré des performances prédictives substantiellement supérieures aux méthodes conventionnelles, avec un c-index de 0,747 pour notre meilleur modèle (Survival Random Forest) contre 0,61 pour les recommandations ELN 2022. Cette amélioration de 22,5% de la précision prédictive constitue une avancée majeure pour la pratique clinique, réduisant significativement l'incertitude pronostique inhérente à cette pathologie hétérogène.

Deuxièmement, notre pipeline d'analyse, entièrement automatisé et interprétable, répond aux exigences de la médecine translationnelle moderne. L'utilisation d'algorithmes explicables, notamment via l'approche SHAP, transforme notre modèle d'une simple "boîte noire" prédictive en un outil d'aide à la décision clinique transparent, condition sine qua non pour son adoption en pratique réelle. Cette caractéristique est particulièrement cruciale dans le contexte des réunions de concertation pluridisciplinaire (RCP) où la justification des décisions thérapeutiques est essentielle.

Au-delà de ces réalisations techniques, notre étude présente plusieurs implications cliniques et sociétales significatives. La précision accrue des prédictions pronostiques pourrait conduire à une meilleure allocation des ressources thérapeutiques, réduisant potentiellement le nombre de traitements sous-optimaux. Sur le plan économique, cette optimisation pourrait se traduire par une diminution des hospitalisations prolongées et des complications iatrogènes, générant des économies substantielles pour les systèmes de santé.

Néanmoins, cette étude présente certaines limitations techniques qui nécessiteront des travaux ultérieurs. La taille modérée de notre échantillon limite potentiellement la détection de patterns rares mais cliniquement significatifs. Des optimisations algorithmiques, notamment au niveau de la sélection des hyperparamètres du modèle Random Forest (nombre d'arbres, profondeur maximale, critères de séparation) et de la gestion des valeurs manquantes, pourraient encore améliorer les performances prédictives. Par ailleurs, l'implémentation d'une validation croisée adaptée aux données censurées et d'une calibration plus fine des probabilités de survie permettrait d'augmenter la robustesse statistique du modèle. Enfin, la validation externe sur des cohortes indépendantes et plus diversifiées géographiquement demeure essentielle pour confirmer la généralisabilité de nos résultats.

En perspective, ce travail pose les fondations d'un système d'aide à la décision clinique qui pourrait être étendu à la recommandation thérapeutique personnalisée. L'incorporation de données relatives à la réponse aux traitements, aux effets secondaires et aux caractéristiques pharmacogénomiques, combinée à notre modèle pronostique, pourrait aboutir à un véritable outil de médecine de précision en hématologie.

En conclusion, cette étude illustre comment l'intégration interdisciplinaire entre hématologie clinique, biologie moléculaire et intelligence artificielle peut transformer notre approche de pathologies complexes comme la leucémie myéloïde aiguë. Elle s'inscrit dans le paradigme émergent de la médecine computationnelle, où les algorithmes d'apprentissage automatique, nourris par des données clinico-biologiques multicentriques et interprétés à la lumière de l'expertise médicale, catalysent l'avènement d'une médecine véritablement personnalisée, basée sur l'évidence et centrée sur le patient.

Références

- AI generated advice eases long waits for surgery, Sarah Neville, Financial Times, 13th May 2025
- Ishwaran et al., "Random survival forests", The Annals of Applied Statistics, 2008.
- Harrell et al., "Regression modeling strategies", Springer, 2015.
- Cox, D. R., "Regression models and life-tables", Journal of the Royal Statistical Society, 1972.
- Modèle de Cox, Wikipedia, consulté le 16 mai 2025
- Documentation officielle de scikit-survival
- Benchmark fourni par l'Institut Gustave Roussy via la plateforme Agorize