

IA pour la sûreté des systèmes industriels critiques

Elkhalfi Ali, Iberoualene Melissa

Résumé

Ce document examine comment l'intelligence artificielle (IA) peut contribuer à renforcer la sécurité des systèmes industriels essentiels. Il met en évidence les limites des méthodes existantes et suggère des alternatives pour les améliorer en utilisant des techniques contemporaines de machine learning et d'ingénierie des systèmes critiques.

Mots clé : Intelligence Artificielle, Systèmes critiques, Sûreté, Détection d'anomalies, Résilience

1 Introduction

Les systèmes industriels critiques, tels que les infrastructures énergétiques, les réseaux de transport ou les industries pétrochimiques, sont fondamentaux pour la société moderne. En raison de leur complexité et des risques qu'ils impliquent, toute défaillance peut entraîner des conséquences graves sur la sécurité humaine, la continuité des services et l'économie. Garantir leur fiabilité et leur sécurité nécessite l'intégration de technologies avancées comme l'intelligence artificielle (IA), capable de surpasser les approches traditionnelles souvent limitées dans la surveillance et la prédiction des anomalies.

Ces systèmes s'appuient sur des réseaux de capteurs collectant en continu des données essentielles, telles que la pression, la température ou les vibrations. Cependant, l'analyse de ces volumes massifs dépasse les capacités des approches conventionnelles. L'IA, grâce à ses algorithmes avancés, permet d'exploiter ces données en temps réel pour détecter des anomalies, prédire des pannes et optimiser les performances, renforçant ainsi la résilience des infrastructures critiques.

Ce projet s'inscrit dans cette démarche en explorant et en analysant les solutions d'IA existantes pour améliorer la sûreté des systèmes critiques. En identifiant leurs forces et leurs limites, cette étude vise à développer des approches innovantes pour répondre aux besoins spécifiques de ces infrastructures dans des contextes complexes et en évolution constante.

2 État de l'art

Cette section est structurée pour explorer les principales contributions de l'intelligence artificielle à la sûreté des systèmes industriels critiques. Les sous-sections sont organisées de manière à établir une progression logique : des modèles fondamentaux aux méthodologies d'entraînement et de validation, en passant par leur adaptation à des contraintes spécifiques. Les thématiques transversales, telles que la cybersécurité et la collaboration homme-machine, viennent compléter cette analyse.

2.1 Modèles d'intelligence artificielle pour la sûreté des systèmes industriels critiques

Les systèmes industriels critiques utilisent divers modèles d'intelligence artificielle (IA) pour renforcer leur sûreté et leur résilience face aux anomalies et menaces potentielles. Ces modèles incluent notamment les réseaux neuronaux spiking standards (SNN) et bayésiens (BSNN), les systèmes flous adaptatifs (ANFIS), les réseaux neuronaux artificiels (ANN), ainsi que les systèmes experts. Chaque approche est conçue pour répondre à des besoins spécifiques tels que la gestion des incertitudes, la précision dans des environnements bruités, ou encore la rapidité de convergence pour des scénarios critiques.

Les réseaux neuronaux spiking standards (SNN), inspirés des neurones biologiques, traitent les données sous forme d'événements discrets ou "spikes", offrant ainsi une gestion efficace des flux de données en temps réel [KK24]. Cependant, les SNN rencontrent des limitations lorsqu'il s'agit de quantifier les incertitudes dans des environnements complexes.

Les réseaux neuronaux bayésiens de type spiking (BSNN), en revanche, intègrent des probabilités bayésiennes, ce qui leur permet de mieux gérer les incertitudes épistémiques (liées au modèle) et aléatoires (liées aux données). Cela améliore leur robustesse face aux environnements industriels imprévisibles, mais au prix d'une complexité computationnelle plus élevée [KK24].

Parallèlement, des modèles tels que les systèmes flous adaptatifs (ANFIS) et les réseaux neuronaux artificiels (ANN) sont souvent utilisés pour l'évaluation des risques et la détection des anomalies. Les systèmes flous adaptatifs combinent des règles floues et des capacités d'apprentissage automatique, ce qui leur confère une grande flexibilité pour gérer les incertitudes et les scénarios où les relations entre variables sont non linéaires [GICA16]. Les ANN, quant à eux, sont largement utilisés pour leur capacité à traiter de grandes quantités de données, mais leur manque d'explicabilité peut limiter leur adoption dans des environnements critiques [AKSM23]. Les systèmes experts, basés sur des règles prédéfinies (if-then), offrent une alternative simple mais souvent limitée à des scénarios spécifiques où les règles sont bien définies [GICA16].

Deux études ont été menées pour comparer les performances de ces modèles dans des scénarios industriels critiques :

- ****SNN vs BSNN**** : L'étude vise à évaluer l'efficacité des réseaux neuronaux spiking standards (SNN) et bayésiens (BSNN) pour la détection des anomalies dans les systèmes de contrôle industriel (ICS). Ces modèles ont été testés sur un ensemble de données contenant plus de 2 millions d'échantillons, collectés via des capteurs industriels tels que des gyroscopes et des capteurs de pression. Les résultats montrent que le modèle BSNN surpasse le SNN en termes de précision (95,93 % contre 90,93 %), avec des taux d'erreur et de faux positifs significativement réduits. Cette comparaison met en évidence l'importance d'intégrer la gestion des incertitudes dans des environnements où la sécurité et la fiabilité sont essentielles [KK24].

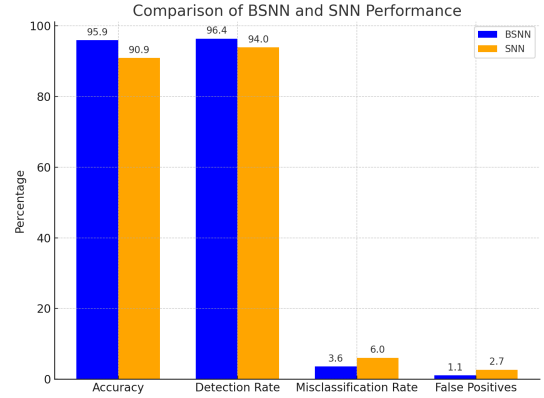


Figure 1: Performances des modèles BSNN et SNN en termes de précision, taux de détection, taux d'erreur et faux positifs.

- ****ANFIS, ANN et Système Expert**** : Cette analyse compare les modèles ANFIS, ANN et un système expert pour l'évaluation des risques dans des infrastructures critiques, en simulant des scénarios de défaillances de pipelines. Les modèles ont été évalués en termes de précision, taux d'erreur et rapidité de convergence. ANFIS a montré une meilleure performance (précision de 94,5 %, MAPE de 5,5 %) tout en nécessitant moins d'itérations pour converger par rapport aux autres techniques. Cette étude souligne l'avantage des systèmes flous dans des contextes où les relations entre variables sont complexes et où une grande précision est nécessaire [GICA16].

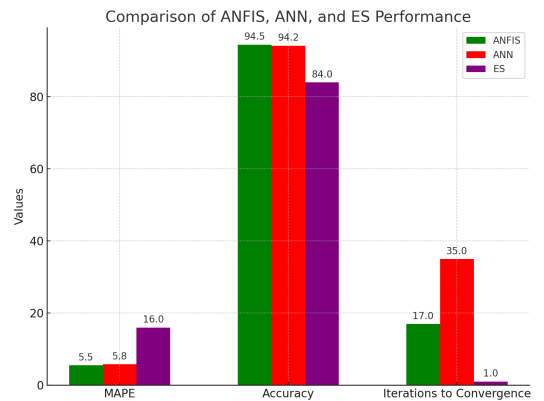


Figure 2: Performances comparées des modèles ANFIS, ANN et système expert pour l'évaluation des risques.

2.2 Entraînement et validation des modèles d'IA dans des systèmes critiques

- **Adaptation aux environnements industriels** : Des outils comme AIMOS (AI Metamorphism Observing Software) permettent de tester la robustesse des modèles en simulant divers scénarios, tels que des perturbations environnementales. Par exemple, chez Renault, AIMOS a validé des modèles dédiés au contrôle qualité des soudures dans des conditions variées.

Cependant, l'approche AIMOS repose sur des scénarios prédéfinis, ce qui limite sa capacité à s'adapter à des conditions imprévues. Bien qu'elle fonctionne efficacement dans des cas spécifiques, son application à d'autres secteurs est difficile.

- **Spécificité des données d'entraînement** : Dans le secteur pétrochimique, des capteurs mesurent des paramètres tels que température, pression ou vibrations, fournissant des données critiques pour entraîner des modèles capables de détecter des anomalies en temps réel. Cependant, la variabilité des données reste un défi majeur. [AKSM23]

- **Critères et outils de validation** : Des métriques comme la précision, le rappel et l'incertitude épistémique sont utilisées pour évaluer les modèles. Ces validations garantissent leur robustesse dans des environnements bruités et leur conformité aux normes de sûreté, comme l'IEC 61508. [AKSM23]

En résumé, l'entraînement et la validation des modèles d'IA dans des systèmes critiques nécessitent des approches adaptées aux contraintes industrielles, garantissant leur performance et leur intégration sûre.

2.3 Adaptation de l'IA dans des systèmes critiques à contraintes spécifiques

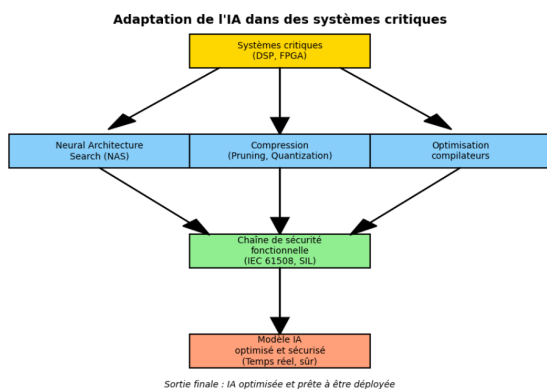


Figure 3: Schéma explicatif d'optimisation et de sécurité fonctionnelle pour l'intégration de l'IA dans des systèmes critiques

L'adaptation de l'intelligence artificielle (IA) dans des systèmes critiques à contraintes spécifiques représente un défi majeur, notamment dans des environnements matériels contraints tels que les DSP et les FPGA. L'article [LES⁺24] met en avant une chaîne d'outils complète pour optimiser les modèles d'IA dans ces environnements tout en respectant les exigences en temps réel et en garantissant la sécurité fonctionnelle. Des techniques telles que la recherche d'architectures neuronales (Neural Architecture Search), la compression de modèles (pruning et quantization), et l'optimisation des compilateurs sont utilisées pour réduire les besoins en ressources et maximiser l'efficacité des modèles. Par ailleurs, l'intégration des fonctions d'IA dans des chaînes de sécurité fonctionnelle, comme décrit dans [KI23], combine les approches traditionnelles de la sûreté avec les capacités prédictives des modèles IA. Ces systèmes sont validés pour s'assurer qu'ils respectent les standards industriels comme IEC 61508, et sont testés avec des méthodologies rigoureuses pour garantir leur conformité aux niveaux d'intégrité de sécurité (SIL).

2.4 Anomalies et cybersécurité dans les systèmes industriels critiques

Parmi les anomalies qui menacent les systèmes industriels critiques, celles liées à la cybersécurité sont particulièrement préoccupantes, car elles peuvent cibler directement les systèmes de contrôle industriels (IACS). Les cyberattaques, telles que les tentatives d'intrusion, les malwares ou les menaces internes, compromettent à la fois la sécurité fonctionnelle et la sécurité des informations en détournant les commandes critiques ou en perturbant les dispositifs de protection. Par exemple, les attaques exploitant des kits d'intrusion ou des menaces internes ont été identifiées comme des sources probables de défaillances dangereuses, en particulier dans les infrastructures critiques telles que les réseaux ferroviaires [TMT⁺23].

Pour détecter et contrer ces menaces, des systèmes multi-agents auto-apprenants combinés à l'apprentissage par renforcement et à la logique floue ont démontré une efficacité accrue, mais leur dépendance à des scénarios spécifiques et à des données de haute qualité limite leur généralisation dans des environnements variés [TMT⁺23]. De même, les méthodes basées sur l'analyse logique des données (LAD) sont limitées par leur scalabilité et leur forte dépendance à des données étiquetées de qualité, ce qui peut réduire leur efficacité dans des systèmes industriels complexes. [SYA22].

2.5 Collaboration homme-machine et sécurité harmonisée

La collaboration entre l'homme et la machine est un élément clé pour renforcer la sûreté des systèmes industriels critiques, notamment grâce à l'émergence de technologies avancées comme l'IoT, l'IA et les robots collaboratifs. Le concept de "sécurité harmonisée" repose sur une approche intégrée où les interactions entre les travailleurs, les machines et l'environnement sont ajustées pour garantir des conditions de travail sûres et efficaces. Par exemple, les robots collaboratifs, équipés de capteurs avancés, peuvent adapter leur vitesse et leur comportement en fonction de la proximité et des compétences des travailleurs, assurant ainsi une coexistence sécurisée dans des espaces partagés [KO19].

Cette approche dépasse les modèles traditionnels de sécurité basés sur l'isolement ("stop and isolation") en proposant des solutions intelligentes qui permettent aux machines de s'adapter dynamiquement aux mouvements et aux conditions des opérateurs. Elle met également l'accent sur la coordination entre les machines et l'environnement, facilitée par des outils de suivi en temps réel et des dispositifs de protection capables de prévenir les risques de manière proactive. En intégrant ces innovations, la "sécurité harmonisée" offre un cadre où les technologies modernes améliorent à la fois la productivité et la sécurité des environnements industriels critiques [KO19].

3 Étude expérimentale : Prédiction de la RUL dans les moteurs d'avion

Dans cette section, nous présentons une étude expérimentale visant à comparer deux modèles d'apprentissage supervisé appliqués à la prédiction de la durée de vie restante (Remaining Useful Life, RUL) de moteurs d'avion à partir de données capteurs.

3.1 Objectifs et contexte industriel

Dans le secteur aéronautique, la maintenance prédictive constitue un levier stratégique pour garantir la sécurité, minimiser les interruptions de service et optimiser les coûts opérationnels. La prédiction de la durée de vie utile restante (Remaining Useful Life, RUL) est au cœur de cette démarche, notamment face à la complexité croissante des équipements et à l'augmentation des données collectées par les capteurs.

Grâce à l'article de Salvador et al. [SYA22], nous avons pu mieux comprendre le fonctionnement des systèmes industriels aéronautiques ainsi que les défis liés à l'exploitation des données pour renforcer la fiabilité. Ce travail met en évidence l'intérêt du *machine learning* pour analyser des scénarios de risque complexes, souvent difficilement traitables par

les approches traditionnelles. Parmi les techniques évoquées, les réseaux de neurones artificiels (ANN) sont présentés comme capables de modéliser des relations non linéaires à partir de données capteurs hétérogènes. C'est dans cette logique que nous avons retenu l'ANN comme point de départ de notre expérimentation : un modèle puissant, flexible, et crédible dans un contexte industriel réel.

Cependant, les ANN présentent également des limites bien connues dans le cadre des systèmes critiques. Leur nature de "boîte noire" rend difficile l'interprétation des décisions produites, ce qui peut poser problème dans des environnements où chaque recommandation de maintenance doit être traçable et justifiable. L'article de Salvador et al. souligne précisément l'importance de la transparence dans les outils d'aide à la décision, en mettant en avant des approches comme LAD pour leur explicabilité. De plus, comme l'a montré notre état de l'art, les ANN sont souvent coûteux en ressources de calcul et complexes à paramétrer, ce qui peut limiter leur déploiement dans des systèmes embarqués.

Dans cette perspective, nous avons choisi d'explorer une alternative avec l'algorithme **LightGBM**, connu pour sa rapidité d'entraînement, sa robustesse sur des données tabulaires et, surtout, sa compatibilité avec l'outil **SHAP** (*SHapley Additive Explanations*). Ce dernier permet de quantifier l'impact de chaque variable sur les prédictions, apportant ainsi un niveau d'explicabilité comparable à celui de LAD, tout en s'intégrant facilement dans un flux de traitement automatisé. Cette transparence est essentielle pour garantir la confiance et l'acceptabilité des modèles dans les environnements critiques soumis à des exigences élevées en matière de sûreté.

Dans ce cadre, nous avons défini les objectifs suivants pour guider notre étude comparative.

L'objectif de cette étude est d'évaluer dans quelle mesure l'algorithme **LightGBM** peut constituer une alternative efficace aux réseaux de neurones artificiels (**ANN**) pour la prédiction de la durée de vie utile restante (*Remaining Useful Life*, RUL) dans un contexte industriel critique.

Plus précisément, nous cherchons à comparer ces deux approches selon trois axes principaux :

1. la précision des prédictions, évaluée à travers plusieurs métriques ($RMSE$, MAE , R^2) ;
2. le coût computationnel lié à l'entraînement des modèles ;
3. la capacité d'explication des résultats, en particulier via l'intégration de l'analyse **SHAP** pour **LightGBM**.

Cette évaluation vise à orienter le choix de modèles d'intelligence artificielle adaptés aux exigences des

systèmes industriels critiques, où la fiabilité, la transparence et l'efficacité sont des critères essentiels.

3.2 Données et prétraitement

Des essais sont effectués sur le jeu de données FD001 issu de la base de données NASA C-MAPSS (Commercial Modular Aero-Propulsion System Simulation), qui est largement employée dans les recherches sur la prédiction de la durée de vie résiduelle des systèmes complexes. Ce jeu reproduit le fonctionnement des moteurs d'avion à travers diverses phases, de leur entrée en service jusqu'à leur défaillance.

Structure des données : Chaque ligne correspond à un instant de fonctionnement d'un moteur, identifié par un ID. On y retrouve :

- 3 paramètres de configuration opérationnelle (op_setting_1 à op_setting_3) ;
- 21 capteurs mesurant des grandeurs physiques telles que la température, la pression, les vibrations, etc. ;
- Le nombre de cycles depuis la mise en service.

Création de la variable cible (RUL) : Comme le jeu ne fournit pas directement la Remaining Useful Life, nous la recalculons pour chaque moteur comme :

$$RUL_i = N_i^{max} - t_i$$

où N_i^{max} est le nombre total de cycles avant panne du moteur i , et t_i le cycle courant. Cela permet de générer une cible supervisée pour l'apprentissage.

Prétraitement appliqué :

- Suppression des colonnes redondantes ou peu informatives (si applicable) ;
- Normalisation des variables continues (standardisation centrée-réduite) afin de faciliter l'entraînement du réseau de neurones ;
- Vérification de la complétude des données et traitement des valeurs manquantes (aucune dans FD001) ;
- Fusion avec les données de test pour appliquer les mêmes transformations ;
- Séparation des jeux d'entraînement et de test en respectant la logique moteur par moteur.

Résultat final : Après traitement, le jeu d'entraînement contient **20 631 observations** issues de plusieurs moteurs simulés, avec **17 variables explicatives** retenues après nettoyage, et une cible continue représentant la RUL.

L'ensemble des transformations a été implémenté via des scripts Python, en utilisant Pandas, NumPy et Scikit-learn. Les fichiers finaux ("train_FD001_preprocessed.csv" et "test_FD001_preprocessed.csv") sont ensuite utilisés pour entraîner les modèles ANN et LightGBM.

3.3 Méthodes de prédiction utilisées

Cette recherche a vu l'élaboration et l'entraînement de deux modèles d'apprentissage supervisé pour prédire la Remaining Useful Life (RUL) : un réseau de neurones artificiels (ANN) et un modèle de gradient boosting basé sur LightGBM. Ces deux méthodes possèdent des attributs complémentaires, facilitant une évaluation tant sur la performance de prévision que sur l'efficacité computationnelle.

3.3.0.1 Modèle 1 : Réseau de Neurones Artificiels (ANN) Le réseau de neurones a été élaboré pour saisir les relations complexes non linéaires entre les variables des capteurs et la durée de vie restante (RUL). Le modèle met en œuvre une architecture entièrement connectée (dense) qui comprend :

- Une couche d'entrée de 17 neurones (correspondant aux features retenues) ;
- Deux couches cachées de 64 et 32 neurones, activées par ReLU ;
- Une couche de sortie linéaire (régression) ;
- Optimisation avec l'algorithme Adam, fonction de coût MSE ;
- Entraînement sur 30 époques, batch size = 128, avec early stopping sur la validation.

Ce modèle a été implémenté en PyTorch, et présente une courbe de convergence stable dès les premières itérations. C'est une référence de haute performance dans les études récentes de la maintenance prédictive.

3.3.0.2 Modèle 2 : Light Gradient Boosting Machine (LightGBM) LightGBM est un algorithme de boosting d'arbres de décision particulièrement efficace sur les données tabulaires. Il est ici configuré comme suit :

- Nombre d'estimateurs : 100 ;
- Learning rate : 0.05 ;
- Max depth : 7 ;
- Objectif : régression (L2 loss) ;
- Fonction d'évaluation : RMSE.

En plus de sa capacité d'entraînement rapide, LightGBM offre une analyse détaillée de l'importance des variables grâce à la technique SHAP (SHapley Additive exPlanations), qui sera utilisée dans la section suivante. Ce modèle a été réalisé en utilisant la librairie `lightgbm` de Python, avec une validation croisée à cinq volets afin d'éviter le surajustement.

3.3.0.3 Comparaison des approches Le réseau de neurones propose une capacité d'approximation robuste, cependant à un coût d'entraînement considérable et avec une explicabilité limitée. En re-

vanche, LightGBM est plus rapide et transparent, ce qui le rend plus approprié pour une mise en œuvre industrielle si les performances sont similaires. L'analyse détaillée de ces deux aspects sera effectuée lors de la prochaine étude expérimentale.

3.4 Résultats obtenus

Dans cette section, nous comparons les performances de deux modèles sur le jeu de données test FD001, en considérant plusieurs éléments : exactitude des prévisions, temps d'apprentissage et constance de la convergence. Une synthèse est présentée dans le tableau 1.

3.4.0.1 Précision prédictive Les deux modèles montrent des performances comparables en ce qui concerne le RMSE et le MAE. LightGBM montre un RMSE un peu plus bas, reflétant une précision supérieure, de même qu'une erreur moyenne inférieure (MAE), attestant de sa solidité.

3.4.0.2 Stabilité de l'entraînement du modèle ANN Comme le montre la courbe de perte figurant en figure 4, le modèle ANN converge rapidement, atteignant une stabilité après la cinquième époque. Cette constance suggère une formation efficace, mais qui reste gourmande en temps de calcul.

3.4.0.3 Temps de calcul LightGBM se démarque clairement par sa vitesse : moins de 2 secondes comparé à plus de 20 secondes pour l'ANN. La vitesse d'exécution de LightGBM en fait un choix privilégié pour les environnements intégrés ou aux ressources limitées.

3.4.0.4 Discussion intégrée Pour résumer, même si les deux modèles peuvent prédire la RUL de manière efficace, LightGBM se démarque par une balance remarquable entre précision, vitesse et capacité d'explicabilité. Ces avantages en font une option prometteuse pour les systèmes industriels sensibles, où la transparence et la rapidité de réaction sont essentiels.

3.5 Explicabilité et analyse SHAP

Un des atouts majeurs du modèle LightGBM réside dans sa capacité à fournir une interprétation directe des prédictions. Pour exploiter pleinement cette caractéristique, nous utilisons la méthode SHAP (SHapley Additive exPlanations), qui permet d'attribuer à chaque variable d'entrée une contribution quantitative, tant au niveau global (importance moyenne) que local (influence sur une prédiction donnée).

3.5.0.1 Importance des variables La figure 5 illustre les résultats de l'analyse SHAP. On y observe que les capteurs `sensor_11`, `sensor_4` et `sensor_7` figurent parmi les plus influents, soulignant leur rôle essentiel dans le processus de vieillissement moteur. Ce

type de visualisation offre à la fois une vue d'ensemble de l'impact moyen des variables et une explication locale des prédictions individuelles. Ces insights sont précieux pour comprendre les mécanismes internes du modèle et justifier ses décisions.

3.5.0.2 Comparaison avec l'ANN À l'opposé de LightGBM, le réseau de neurones artificiel (ANN) fonctionne comme une boîte noire. Il n'offre pas une explication limpide des décisions prises sur la base des variables d'entrée. Dans des contextes industriels sensibles, où l'importance de la transparence des modèles est cruciale pour assurer la traçabilité et la sûreté, cette absence de clarté représente un défaut significatif. L'approche SHAP offre ici une solution rigoureuse et pratique pour satisfaire ce besoin d'explication.

3.5.0.3 Apports pour la maintenance prédictive En plus de l'interprétation des prédictions, l'analyse SHAP souligne les capteurs les plus influents pour la prévision de la RUL. Cela permet de diriger les stratégies de maintenance vers les sources d'information les plus pertinentes, de favoriser les inspections sur certaines parties essentielles du moteur, et d'optimiser la disposition des capteurs. Ainsi, la possibilité d'expliciter les décisions représente un atout crucial pour LightGBM, augmentant la confiance des opérateurs et simplifiant l'incorporation du modèle dans un processus de prise de décision effectif.

4 Visualisation des résultats

Les figures ci-dessous présentent une synthèse visuelle des résultats obtenus pour les deux modèles étudiés. Elles permettent d'apprécier à la fois la performance prédictive des modèles (via le tableau comparatif), la dynamique d'apprentissage de l'ANN (à travers la courbe de perte), ainsi que l'interprétabilité des prédictions du modèle LightGBM grâce à l'analyse SHAP.

Critère	LightGBM	ANN	Commentaires
RMSE	33.89	34.70	Léger avantage pour LightGBM
MAE	24.50	26.10	LightGBM génère des erreurs moyennes plus faibles
R^2 Score	0.33	0.30	LightGBM explique mieux la variance du RUL
Temps d'entraînement	1.68 sec	20.26 sec	LightGBM est environ 12 fois plus rapide
Courbe d'apprentissage	N/A	Stable dès epoch 5	Convergence rapide pour l'ANN mais coût élevé

Table 1: Comparaison des performances entre *LightGBM* et *ANN* sur le jeu de test *FD001*

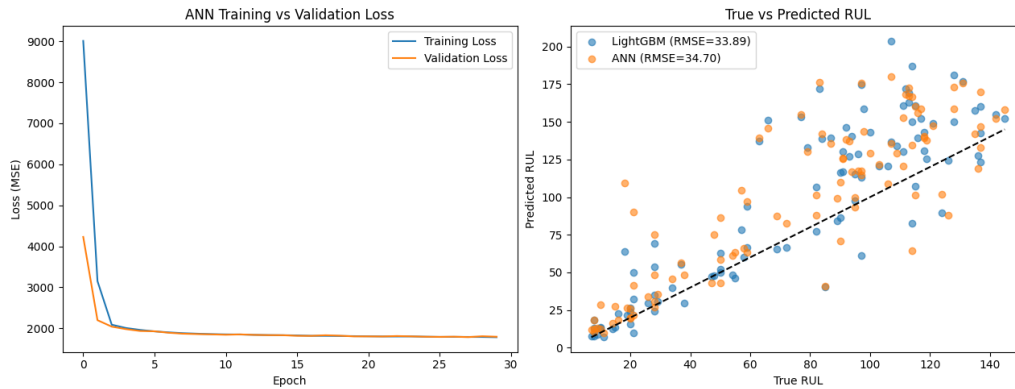


Figure 4: Évolution de la perte (MSE) pour le modèle ANN sur les jeux d'entraînement et de validation

```
[LightGBM] [Info] Auto-choosing row-wise multi-threading, the overhead of testing was 0.001319 seconds.
You can set 'force_row_wisetrue' to remove the overhead.
And if memory is not enough, you can set 'force_col_wisetrue'.
[LightGBM] [Info] Total Bins 2832
[LightGBM] [Info] Number of data points in the train set: 28631, number of used features: 17
[LightGBM] [Info] Start training from score 107.807862
```

Figure 5: Importance et contribution des variables selon l'analyse SHAP

5 Discussion

Les résultats montrent que LightGBM offre des performances proches de l'ANN testé, avec un temps d'entraînement bien inférieur. Ce constat reste à relativiser, l'architecture de l'ANN étant simple et non optimisée.

L'analyse SHAP apporte une première forme d'interprétabilité utile, mais limitée à une lecture globale. Enfin, l'expérimentation étant restreinte au sous-ensemble FD001, la généralisation des résultats reste à confirmer sur des scénarios plus complexes.

6 Conclusion et perspectives

La première phase de ce travail a porté sur l'état de l'art des méthodes d'intelligence artificielle appliquées à la sûreté des systèmes industriels critiques. Nous avons identifié les limites de plusieurs approches avancées (BSNN, ANFIS, VAE), notamment en termes de complexité, d'explicabilité et de compatibilité avec des environnements certifiés.

Ces constats ont mis en évidence le besoin de modèles à la fois robustes, explicables et adaptés aux contraintes industrielles. Ils ont orienté la phase expérimentale menée au second semestre, centrée sur la prédiction de la Remaining Useful Life (RUL) de moteurs d'avion.

À partir du jeu de données NASA C-MAPSS (FD001), nous avons comparé deux modèles : un réseau de neurones artificiels (ANN) et un modèle LightGBM associé à une analyse SHAP. Les résultats montrent que LightGBM atteint une performance comparable, voire supérieure, tout en réduisant le temps d'entraînement et en apportant une meilleure lisibilité des décisions.

Notre démarche propose ainsi une alternative efficace et interprétable à l'usage de modèles complexes dans les systèmes industriels critiques. Elle relie analyse théorique et mise en œuvre concrète, dans un cadre proche des contraintes réelles du terrain.

Perspectives : des pistes d'amélioration incluent l'évaluation sur des données plus complexes, l'ajout de méthodes de détection d'anomalies temps réel, ou l'extension à d'autres sous-systèmes critiques. Le déploiement embarqué et certifiable de LightGBM, couplé à une surveillance dynamique des variables clés, constitue également un axe prometteur pour l'industrie 4.0.

Références

- [AKSM23] S.R. Arun, S. Karthik, Vigneshwaran. S, and Sasikumar. M. Using Artificial Intelligence to Enhance Hazard Assessment and Safety Measures in Petrochemical Industries : Development and Analysis . In *2023 3rd International Conference on Pervasive Computing and Social Networking (ICPCSN)*, pages 823–830, Los Alamitos, CA, USA, June 2023. IEEE Computer Society.
- [GICA16] A. Guzman, S. Ishida, E. Choi, and A. Aoyama. Artificial intelligence improving safety and risk analysis : A comparative analysis for critical infrastructure. In *2016 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM)*, pages 471–475, 2016.
- [KI23] Michael Kieviet and Padma Iyengar. Integration of machine learning safety functions in the ontology of functional safety. In *2023 IEEE 21st International Conference on Industrial Informatics (INDIN)*, pages 1–5, 2023.
- [KK24] Tala Talaei Khoei and Naima Kaabouch. A safe and reliable bayesian spiking neural network in industrial control systems. In *2024 IEEE 3rd International Conference on Computing and Machine Intelligence (ICMI)*, pages 1–7, 2024.
- [KO19] Hiroo Kanamaru and Hiroyuki Ogihara. Harmonious safety - collaboration, cooperation and coordination between workers, machines and environments. In *2019 58th Annual Conference of the Society of Instrument and Control Engineers of Japan (SICE)*, pages 1766–1771, 2019.
- [LES⁺24] Joakim Lindén, Andreas Ermedahl, Hans Salomonsson, Masoud Daneshmand, Björn Forsberg, and Paris Carbone. Autonomous realization of safety- and time-critical embedded artificial intelligence. In *2024 Design, Automation Test in Europe Conference Exhibition (DATE)*, pages 1–4, 2024.
- [SYA22] Marcos Salvador, Soumaya Yacout, and Ayman AboElHassan. Using big data and machine learning to improve aircraft reliability and safety. In *2022 Annual Reliability and Maintainability Symposium (RAMS)*, pages 1–9, 2022.
- [TMT⁺23] Roumen Trifonov, Slavcho Manolov, Georgi Tsochev, Galya Pavlova, and Kamelia Raynova. On the choice of effective artificial intelligence methods for cyber defense of industrial systems. In *2023 International Scientific Conference on Computer Science (COMSCI)*, pages 1–4, 2023.