JOURNAL AFRICAIN DES SCIENCES
VOL. 3, **3** (2026) 136-145

https://jafrisci.org

ISSN (online) : 3107-0949 ; ISSN (print) : 3107-0930

OPEN ACCESS

ARTICLE DE RECHERCHE

Article Info.:
Reçu : le 10/03/2026
Accepté: le 13/04/2026
Publié : le 14/04/2026

OPTIMISATION BI-CRITERE DES HYPER PARAMETRES D'UN MODELE SVM ASYMETRIQUE BINAIRE PAR NSGA-II POUR LA PREDICTION DU RISQUE D'ACCIDENT VASCULAIRE CEREBRAL

Dieu merci Mamboko Mamboko [1,*], Gradi Lubwele Kamingu [2], Eli Ndunga Mbangu [2], Carmi Ndeke [2], Jean-Pierre Mulenda Lueteta [1]. Pierre Katalav Kafunda [2]

https://doi.org/10.70237/jafrisci.2026.v3.i3.15

Résumé

Notre article traite de la problématique de l'évaluation binaire du risque d'accident vasculaire cérébral (AVC) en utilisant une approche d'apprentissage supervisée optimisée. Le diagnostic médical, par intelligence artificielle, est souvent confronté au défi de la classification asymétrique où le cout d'un faux négatif (patient malade non détecté) est nettement plus préjudiciable qu'un faux positif. Face à cet enjeu, nous proposons une méthodologie d'optimisation bi-critère des hyper paramètres γ, C+ et C- respectivement paramètre du noyau et coefficients de régulation de deux classes de prédiction du modèle support vector machine (SVM), et reposant sur l'application de la méta-heuristique évolutionnaire NSGA-II (Non Dominated Sorting Genetic Algorithm II) qui permet d'obtenir un front de Pareto composé de plusieurs modèles SVM de compromis entre la minimisation du taux de faux positifs (NFA) et celle du taux de faux négatifs (NFR) afin d'y naviguer pour trouver le SVM de meilleur compromis, au lieu d'un modèle unique optimisé par des méthodes classiques, à l'occurrence GridSearchCV. L'expérimentation menée sur les données cliniques, démontre que cette approche bi-critère surpasse les méthodes des réglages classiques. En ajustant finement les paramètres de pénalité de manière asymétrique, le modèle devient capable de mieux identifier les sujets à risque d'AVC, garantissant ainsi une plus grande fiabilité pour l'aide à la décision médicale. Les résultats obtenus confirment l'efficacité du couplage SVM/NSGA-II pour traiter des bases de données médicales déséquilibrées.

Mots-clés: SVM, NSGA-II, optimisation bi-critère, SVM asymétrique binaire, Accident Vasculaire Cérébral (AVC)

Abstract

Our article addresses the issue of binary risk assessment for stroke (CVA) using an optimized supervised learning approach. Medical diagnosis via artificial intelligence is often faced with the challenge of asymetric classification, where the cost of a false negative (an undetected sick patient) is significantly more detrimental than a false positive. To address this challenge, we propose a bi-criterion optimization methodology for the hyperparameters γ, C+ and C- representing the kernel parameter and the regulation coefficients for two prediction classes of the support vector machine (SVM) model. This methodolgy relies on the application of the evolutionary meta-heuristic NSGA-II (Non Dominated Sorting Genetic Algorithm II), wich generates a pareto front composed of several SVM models. These models offer various trade –offs between minimizing the false positive rate (NFA) and the false negative rate (NFR). This allows for navigation to find the SVM with the best compromise, rather than relying on a single model optimized by traditional methods, such as GridSearchCV. Experiments conducted on clinical data demonstrate that bi-criterion approach outperforms traditional tuning methods. By finely adjusting the penalty parameters asymmetrically, the model becomesb bette rat identifying subjects at risk of stroke, thereby ensuring greater reliability for medical decision support. The results confirm the effectiveness of the SVM/NSGA-II coupling for processing unbalanced medical databases.

Keywords: SVM, NSGA-II, Bi-criterion optimization, binary asymmetric SVM, Stroke (CVA)

1. Introduction

v Selon la publication de la Haute Autorité de Santé (HAS) d'octobre 2025, Les accidents vasculaires cérébraux (AVC) constituent aujourd'hui une urgence médicale majeure et l'une des principales causes de handicap et de mortalité à l'échelle mondiale. Face à la complexité des facteurs de risque, l'intégration de l'intelligence artificielle, et particulièrement

Correspondance : *dieumerci.mamboko@ucc.ac.cd* ***(D. M. Mamboko)***

Copyright : © The Author(s) Published under a Creative Commons Attribution 4.0 International Licence (CC BY 4.0)

[1] Faculté de sciences Informatiques, Université Catholique du Congo (UCC), République Démocratique du Congo (RDC)
[2] Faculté de sciences et technologies, Mention Mathématiques, Statistique et informatique, Université de Kinshasa (Unikin), République Démocratique du Congo (RDC)

des machines à vecteurs de support (SVM), est devenue un levier crucial pour le diagnostic précoce [1] [2]. Cependant, l'application du SVM en milieu médical se heurte souvent au problème de classification asymétrique où le cout lié à un faux négatif (patient malade non détecté) est infiniment plus lourd que celui d'un faux positif. [3] [4]

L'efficacité d'un modèle SVM repose intrinsèquement sur le réglage de ses hyper paramètres γ, C_+ et C_- désignant respectivement paramètre du noyau et coefficients de régulation de deux classes à prédire [3] [4].

La littérature récente souligne une transition des méthodes de recherche classiques vers des approches méta-heuristiques plus robustes. Par exemple, Bhardwaj et Gupta (2023), dans leur article intitulé « Multi-Objective Optimization of SVM Hyperparameters using NSGA-II for Imbalanced Health Data », démontrent que l'optimisation des hyper paramètres par des algorithmes évolutionnaires permet de mieux capturer les structures de données complexes dans les Datasets de santé déséquilibrés [3]. Dans la même lignée, Moradpour et al. (2024), dans leur article intitulé « Multi-Objective Performance Optimization of Machine Learning Models in Healthcare », ont introduit le cadre MOOF, utilisant l'algorithme NSGA-II pour équilibrer la sensibilité et la spécificité, offrant ainsi aux cliniciens un front de Pareto de solutions optimales plutôt qu'un modèle unique potentiellement biaisé [5]. Par ailleurs, les travaux de Rodriguez et Chen (2022), dans leur article intitulé « Adressing class imbalance in stroke prediction : A cost-sensitive SVM approach », insistent sur la nécessité d'intégrer des fonctions de cout asymétrique directement dans e processus d'optimisation pour minimiser les risques cliniques [6]. Enfin, M. Sekkal, A. Benzina et L.Benkrelifa (2023), ont démontré comment l'optimisation multi objectif basée sur NSGA-II pour les réseaux de neurones , permet d'atteindre des taux de classification de 100% sur certains jeux de données médicaux en gérant mieux la complexité des hyper paramètres tout en soulignant que cette méthodologie est directement transposable aux autres modèles[7].

Le présent article s'inscrit dans cette dynamique en proposant une optimisation bi-critère des hyper paramètres du SVM par la méta-heuristique NSGA-II, avec comme particularité, une rigueur mathématique sur la formulation mathématique de notre problématique.

Dans le contexte de classification binaire asymétrique, la performance du modèle SVM est relative aux valeurs prises par ses hyper paramètres γ, C_+ et C_- afin de minimiser les erreurs de prédiction NFA et NFR respectivement, Erreur type I et Erreur type II, dont les expressions mathématiques sont explicitement formulées dans la section 3.2., de notre article. C'est ainsi que notre attention est principalement retenue par la question suivante : Comment choisir les hyper-paramètres du modèle SVM de manière à réduire simultanément l'erreur du type I et l'erreur du type II pour la prédiction du risque d'AVC dans un contexte d'asymétrie des données d'entrainement ? Le choix des hyper-paramètres du modèle SVM ainsi posé relève du problème d'optimisation bi-critère avec comme fonctions objectif, NFA et NFR pour enfin trouver un front de Pareto composé de plusieurs modèles SVM non dominés en appliquant l'algorithme NSGA II comparativement plus efficace que le modèle unique optimisé par GridSearchCV, une approche d'optimisation monocritère des hyper paramètres de référence dont le critère de sélection est AUC [8]. Il est à noter que la prediction du risque dont il est question dans notre article, ne tient pas compte du degré du risque et elle est binaire soit il y'a risque (classe positive), soit pas de risque classe (négative).

Ainsi pour capter l'attention de nos lecteurs futurs, nous avons structuré notre article en plusieurs sections dont une section préliminaire pour expliquer de manière brève avec rigueur mathématique les concepts théoriques de notre étude, suivie de la méthodologie de l'étude, puis formulation et enfin les résultats obtenus avec discussion pour clôturer par une conclusion ainsi que des perspectives.

2. Préliminaires

2.1. SVM asymétrique binaire

2.1.1. Formulation mathématique

Un classifieur SVM binaire cherche une frontière appelée hyperplan séparateur, optimale entre deux classes à partir des données d'apprentissage. L'objectif est de choisir un hyperplan séparateur qui maximise la marge. La marge étant la distance entre l'hyperplan et les points de données des classes distincts les plus proches de lui, appelés vecteurs de support [2], [9], [10]. L'hyperplan optimal est celui qui satisfait les conditions suivantes :

- La distance entre les vecteurs bien classés et l'hyperplan optimal doit être maximale.
- La distance entre les vecteurs mal classés et l'hyperplan optimal doit être minimale

Étant donné un ensemble de m exemples d'apprentissage $x_i \in R^n$ appartenant à la classe y_i:

$$(x_i,y_i)\dots(x_m,y_m),\ x_i \in R^n,\ y_i \in \{-1,+1\}$$

Pour traiter l'asymétrie, on n'utilise plus un seul coefficient C mais deux coefficients différents C_+ (pour la classe positive) et C_- (pour la classe négative). L'idée est d'imposer une pénalité plus forte sur les erreurs de la classe minoritaire. Dans ce cas, le programme mathématique s'écrit alors [6], [11], [12]:

$$\min_{w,b,\xi} \frac{1}{2}\|w\|^2 + c_+ \sum_{y_i=1} \xi_i + c_- \sum_{y_i=-1} \xi_i$$

$$y_i(\langle w, x_i\rangle + b) \geq 1 - \xi_i \qquad i = 1,2 \dots, m \quad (1)$$

$$\xi_i \geq 0 \quad i = 1,2 \dots . m$$

Avec

- $w \in R^n$: Le vecteur de poids
- $b \in R$: Le biais
- ξ_i : les variables de ressort déterminant les erreurs de classification.

Partant du programme (1), La maximisation du lagrangien dual par rapport aux α_i s'écrit :

$$\text{Max}_\alpha \left\{\sum_{i=1}^{m} \alpha_i - 1/2 \sum_{i,j=1}^{m} \alpha_i \alpha_j y_i y_j \langle x_i, x_j \rangle\right\}$$

$$\text{sc.} \quad \begin{cases} \alpha_i \leq C_+ \text{ pour } y_i = +1 \\ \alpha_i \leq C_- \text{ pour } y_i = -1 \\ \sum_{i=1}^{m} \alpha_i y_i = 0 \\ \alpha_i \geq 0 \end{cases} \quad (2)$$

Une fois la solution optimale $\alpha^* = (\alpha_1^*, \alpha_2^* \dots \alpha_m^*)$ du probleme obtenue, le vecteur des poids de l'hyperplan à marge maximale s'écrit :

$$w^* = \sum_{i=1}^{m} \alpha_i^* y_i x_i \quad (3)$$

Une fois que α^*, b^* calculés, la règle de classification d'une nouvelle observation x basée sur l'hyperplan à marge maximale est donnée par :

$$h(x) = \text{sign}\left(\sum_{i=1}^{m} y_i \alpha_i^* \langle x_i, x \rangle + b^*\right) \quad (4)$$

Avec $sign(.) \in \{1, -1\}$: fonction signe

Dans la plupart de problèmes réels, les données d'apprentissage ne sont pas toujours linéairement séparables. Dans ce cas, projeter les m exemples d'apprentissage dans un espace T de dimension $q > n$ grâce à une fonction non linéaire $\emptyset$ appelé fonction noyau choisie à priori et d'appliquer la même méthode d'optimisation dans l'espace T appelé espace des caractéristiques ou aussi espace transformé. [6], [11], [12]

En remplaçant $\langle x_i, x_j \rangle$ par $\langle \emptyset(x_i), \emptyset(x_j) \rangle$ dans le programme (2), l'hyperplan séparateur obtenu dans l'espace T est appelé hyperplan optimal généralisé. Ainsi on peut facilement calculer $\langle \emptyset(x_i), \emptyset(x_j) \rangle$ à l'aide d'une fonction symétrique K, dite noyau, définie par :

$$K(x_i, x_j) = \langle \emptyset(x_i), \emptyset(x_j) \rangle \quad (5)$$

Le noyau $K(x, y)$ est une fonction de deux variables, symétrique et positive.

D'où le programme (2) devient :

$$\text{Max}_\alpha \left\{\sum_{i=1}^{m} \alpha_i - 1/2 \sum_{i,j=1}^{m} \alpha_i \alpha_j y_i y_j K(x_i, x_j)\right\}$$

$$\text{sc.} \quad \begin{cases} \alpha_i \leq C_+ \text{ pour } y_i = +1 \\ \alpha_i \leq C_- \text{ pour } y_i = -1 \\ \sum_{i=1}^{m} \alpha_i y_i = 0 \\ \alpha_i \geq 0 \end{cases} \quad (6)$$

Où les α_i représentent les multiplicateurs de Lagrange. Dans le cas d'un noyau gaussien RBF, la fonction noyau est :

$$K(x_i, x_j) = e^{\left(\frac{\|x_i - x_j\|^2}{-2\gamma}\right)} \quad (7)$$

Où γ : permet de régler la largeur de la gaussienne. Quand γ est grand, la similarité d'un exemple par rapport à ceux qui l'entourent est élevée, alors que γ tend vers 0, l'exemple ne sera similaire à aucune autre.

Ainsi, la frontière de décision devient :

$$h(x) = \text{sign}\left(\sum_{i=1}^{m} y_i \alpha_j^* K(x_i, x) + b^*\right) \quad (8)$$

2.1.2. Optimisation des hyper paramètres

L'optimisation des hyper paramètres est une étape essentielle dans le processus de machine Learning. Une bonne sélection des hyper paramètres peut considérablement améliorer les performances d'un modèle.

L'optimisation des hyper paramètres garantit La performance, La généralisation et L'efficacité du modèle SVM. Plusieurs techniques d'optimisation des hyper paramètres sont utilisées notamment GridSearchCV qui essaie toutes les combinaisons possibles d'hyper paramètres dans une grille définie.

Dans de nombreux cas (déséquilibre de classes, absence de connaissance des coûts d'erreur), un critère unique ne suffit pas à évaluer la performance. C'est là qu'intervient l'optimisation multicritère, précisément l'optimisation bi-critère dans notre article. [7], [13]

2.2. Problème d'optimisation multicritère

2.2.1. Définitions

Le problème d'optimisation multicritère peut être défini formellement comme suit [14], [15]:

$$\text{optimiser } \overrightarrow{F(\vec{x})} = \left(f_1(\vec{x}), f_2(\vec{x}) \dots f_m(\vec{x})\right) \text{ tel que } \vec{x} \in X$$

$$X = \left\{ \vec{x} \in R^n \middle| \begin{matrix} g_i(\vec{x}) \leq 0 \text{ ou } g_i(\vec{x}) \geq 0 \quad i = 1 \dots q \\ h_j(\vec{x}) = 0 \quad j = 1 \dots l \\ \vec{x} = [x_1 \dots x_n]^T \end{matrix} \right\} (9)$$

- $\overrightarrow{F(\vec{x})} \in R^m, g_i(\vec{x}), \quad h_j(\vec{x})$ sont les fonctions des contraintes déterminant le domaine réalisable X de solutions.
- R^n: Espace de décision et R^m: espace de critères

Lorsque $m = 2$, le problème d'optimisation est dit bi-critère, c.à.d. deux fonctions à optimiser.

Un problème d'optimisation multicritère consiste à trouver le vecteur de décision idéal $\vec{x}^*$ tel que les contraintes $g_i(\vec{x})$ et $h_j(\overrightarrow{x})$ soient satisfaites et dont $\vec{F}(\vec{x}^*)$

Dans l'optimisation multicritère, souvent les fonctions optimiser étant conflictuelles, la solution optimale n'est pas unique comme le cas de l'optimisation monocritère mais un

ensemble de solutions non dominées au sens de Pareto ou de compromis optimal. [14], [15]

2.2.2. Dominance au sens de Pareto

- Soient $\vec{u}, \vec{v} \in X$, $\vec{u}$ domine $\vec{v} \Leftrightarrow \exists\, i'\ tel\ que\ f_{i'}(\vec{u}) < f_{i'}(\vec{v})\ et\ \forall i \neq i', f_i(\vec{u}) \leq f_i(\vec{v})$ avec $i = 1,2 \ldots m$
- $\vec{u} \in X$ est dit non dominé $\Leftrightarrow \nexists\, \vec{v} \in X$ qui domine $\vec{u}$
- $S = \{\vec{u} \in X : \vec{u}\ est\ non\ dominé\}$ est appelé ensemble de solutions non dominés ou de compromis optimal entre $f_i(\vec{x})$
- On appelle front de Pareto $\mathcal{F} = \vec{F}(S) \in R^m$. Donc l'image de S dans l'espace des critères

Résoudre un problème d'optimisation multicritère revient à définir S ou $\mathcal{F}$.[14], [15]

2.2.3. Algorithme NSGA-II

L'algorithme NSGA-II (Non-dominated Sorting Genetic Algorithm II), un algorithme génétique élitiste parmi les Meta heuristiques basées sur une population de solutions, conçu pour les problèmes multi-objectifs. Il permet d'obtenir une population de solutions équilibrée en termes de qualité et de diversité [14], [16] [17].

Etapes [16], [17], [18]

1. Initialisation
   1.1. Créer une population parente P_0 de taille N aléatoirement
   1.2. Générer une population enfant Q_0 (croisement + mutation) de taille N
   1.3. $t \leftarrow 0$
2. Boucle principale (*Tant que t<M)*
   2.1. Combiner les populations : $R_t = P_t \cup Q_t$ (taille 2N)
   2.2. Tri par non dominance : diviser R_t en différents fronts de Pareto ($F_1, F_2 \ldots F_n$)
3. Remplissage de la nouvelle population P_{t+1} :
   3.1. Initialiser $P_{t+1} = \emptyset$
   3.2. $i = 1$
   3.3. Tant que $|P_{t+1}| + |F_i| \geq N$:
   Calculer crowding_distance (F_i)
   $P_{t+1}=P_{t+1} \cup F_i$
   i=i+1
   3.4. Fin tant que
   3.5. Si $|P_{t+1}|<N$:
   Calculer crowding_distance (F_i)
   Trier F_i par distance décroissante
   Ajouter les N-$|P_{t+1}|$ meilleurs individus dans P_{t+1}
   3.6. Fin si
4. Création de la population enfant Q_{t+1} :
   4.1. Sélection par tournoi encombré sur P_{t+1}
   4.2. Application du croisement et de la mutation
5. $t = t + 1$
6. Aller à 2
7. Retourner F_1 de la dernière population P_t

Il est à noter que F_1 contient les solutions non dominées par aucune autre, F_2 contient les solutions dominées que par les solutions de F_1, ainsi de suite.

Description et Analyse de la complexité de NSGA-II [14], [18]

La complexité de la méthode NSGA-II est principalement dictée par son mécanisme de tri non dominé conçu pour être plus efficace que la première version.

Décomposition de la complexité

Sot *N* la taille de la population et *M* le nombre d'objectifs. Le cout de la méthode se divise en 3 parties majeures :

Tri non dominé rapide : consiste à affecter chaque individu à un front de Pareto.

Les variables clés sont : n_p nombre d'individus dominant l'individu p et S_p ensemble d'individus dominés par l'individu p.

1. Pour chaque p ∈ P :
1.1. $S_p = \emptyset$
1.2. $n_p = 0$
2. Pour chaque q ∈ P:
2.1. Si p domine q : $S_p = S_p \cup \{q\}$
2.2. Sinon si q domine p : $n_p = n_p + 1$
3. Si $n_p = 0$: $p.rank = 1$ et $F_1 = F_1 \cup \{p\}$
4. i = 1
5. Tant que $F_i \neq \emptyset$:
5.1. $H = \emptyset$ (Ensemble pour le prochain front)
5.2. Pour chaque p ∈ F_i :
Pour chaque q∈S_p : $n_q=n_q$-1
Si n_q=0 : q.rank=i+1 et H=H∪{q}
i=i+1
F_i=H

Commentaire :

Pour la comparaison de dominance : pour chaque paire d'individus, on compare leurs M objectifs. Il y'a $\frac{N(N-1)}{2}$ paires. D'où le cout est $O(MN^2)$

Crowding_distance_assignement : calcule la distance de regroupement pour maintenir la diversité.

Entrée : un ensemble de solutions I appartenant à un même front de Pareto

Sortie : l'ensemble I où chaque individu possède une valeur de distance

Crowding_distance_assignement (I)

n =Nombre d'individus dans I

1. Initialisation
1.1. Pour chaque i ∈ I : I[i].distance=0
2. Pour chaque objectif m :
2.1. Trier les individus selon l'objectif m
3. I = Trier_par_objectif (I, m)
3.1. I[1].distance=∞
3.2. I[n].distance=∞
3.3. f_{max}=I[n].m
3.4. f_{min}=I[1].m
3.5. Pour i allant de 2 à n-1 :

$$I[i].distance=I[i].distance+(\frac{I[i+1].m-I[i-1].m}{f_{max}-f_{min}})$$

Commentaire :

Pour chaque objectif, on trie la population de taille *N* selon cet objectif. D'où le cout du tri est $O(NlogN)$. Et pour tous les objectifs, le coût est O(MNlogN)

Operateur de comparaison encombrée ($<_n$) : utilisé lors de la sélection pour le tournoi ou pour remplir le dernier front de Pareto. Une solution *x* est préférée à une solution *y* si :

8. rank(x)<rank(y)
9. ou rank(x)=rank(y) et distance (x)>distance(y)

Commentaire : les opérateurs génétiques de sélection, croisement et mutation dépendent linéairement de la taille de la population et le nombre d'objectifs. D'où le cout est O(MN).

Ainsi, la complexité totale par génération est dominée par le terme le plus élevé : $O(MN^2)$

3. Formulation du problème posé

Considérons un problème de classification asymétrique binaire d'un modèle SVM.

Soient

- un ensemble de n données d'entrainement E= $\{(x_i, y_i): x_i \in R^n, y_i \in \{-1,+1\}, i = 1,2 \dots n\}$ avec n_+ x_i telles que $y_i = +1$ et n_- x_i telles que $y_i = -1$ et $n_+ + n_- = n$
- un ensemble de p données de validation V= $\{(x_i, y_i): x_i \in R^n, y_i \in \{-1,+1\}, i = 1,2 \dots n\}$ avec p_+ x_i telles que $y_i = +1$ et p_- x_i telles que $y_i = -1$ et $p_+ + p_- = n$
- un ensemble de m données test T= $\{(x_i, y_i): x_i \in R^n, y_i \in \{-1,+1\}, i = 1,2 \dots m\}$ avec m_+ x_i telles que $y_i = +1$ et m_- x_i telles que $y_i = -1$ et $m_+ + m_- = m$

Le problème posé dans notre article est de trouver les valeurs des hyper paramètres γ, C_+, C_-du modèle SVM minimisant simultanément le taux de faux négatifs et celui de faux positifs, tenant compte des ensembles de données E, V et T ci-dessus définis. En d'autres termes, il s'agit de trouver un front de Pareto composé de plusieurs modèles SVM de compromis entre la minimisation du taux de faux négatifs et celle du taux de faux positifs.

3.1. Définition des variables

- γ: Paramètre du noyau gaussien
- C_+ : Coefficient de régulation de la classe positive
- C_- : Coefficient de régulation de la classe négative

3.2. Modélisation mathématique explicite des fonctions-objectif

Nous avons deux fonctions-objectif à minimiser simultanément dont le taux de faux positifs NFA et le taux de faux négatifs NFR. Mathématiquement, ces deux fonctions peuvent s'écrire comme suit :

En effet, pour classer un individu x_i, nous nous fions au signe de la fonction de décision $h(\cdot)$, donc

$$\begin{aligned}\widehat{y_i} &= \text{sign}(h(x)) \\ &= \text{sign}\left(\sum_{i=1}^{m} \alpha_i^* y_i K(x_i, x) + b^*\right) \quad (10)\end{aligned}$$

Ecrivons maintenant de manière explicite ce problème.

Regroupons Les hyper paramètres du SVM dans le vecteur suivant :

$$\theta = (\gamma, C_+, C_-),$$

La fonction de décision peut alors s'écrire de la manière suivante :

$$h_\theta(x) = \sum_{i=1}^{m} \alpha_i^*(\theta) y_i K_\gamma(x_i, x) + b^*(\theta) \quad (11)$$

Cependant, les deux fonctions à minimiser simultanément sont :

- NFA, c'est-à-dire $y_i = -1$ et $\widehat{y_i} = +1$,

$$NFA(\theta) = \frac{1}{p_-} \sum_{i:y_i=-1}^{m_-} \mathbb{1}(h_\theta(x_i) \geq 0) \quad (12)$$

- NFR, c'est-à-dire $y_i = +1$ et $\widehat{y_i} = -1$,

$$NFR(\theta) = \frac{1}{p_+} \sum_{i:y_i=+1}^{m_+} \mathbb{1}(h_\theta(x_i) < 0) \quad (13)$$

Où :

- y_i est la classe réelle de x_i et $\widehat{y_i}$ est la classe prédite de x_i
- p_+ et p_- désignent respectivement le nombre des individus de la classe +1 et le nombre des individus de la classe -1 dans V.

- $\mathbb{1}(\cdot)$ est la fonction indicatrice (ou fonction caractéristique), qui vaut 1 si la condition entre parenthèses est vraie, 0 sinon.

D'où le problème d'optimisation bi-critère peut s'écrire explicitement

$$\min_{\theta \in D}\big(NFA(\theta), NFR(\theta)\big) \quad (14)$$

Avec $\big(\alpha^*(\theta), b^*(\theta)\big)$ la solution optimale du dual de SVM.

Dans la pratique, Les deux quantités NFA et NFR sont données dans la matrice de confusion obtenue après avoir testé le modèle SVM entrainé.

4. Méthodologie de l'étude

Pour bien atteindre la finalité relative à la question principale de notre étude de trouver un front de Pareto composé de plusieurs modèles SVM de compromis entre la minimisation du taux de faux négatifs et celle du taux de faux positifs, La méthodologie utilisée dans notre étude se résume à travers la figure suivante :

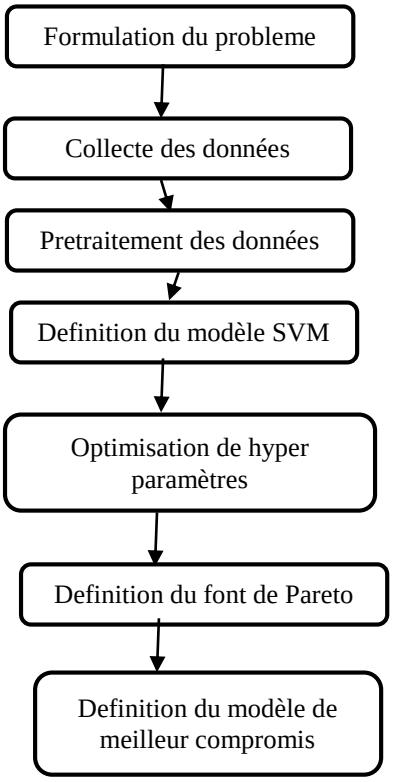

Fig.1 : *Méthodologie de l'étude*

4.1. Collecte des données

4.1.1. Description du jeu des données

Les données cliniques utilisées sont téléchargées en ligne sous le nom Healthcare Dataset-Stroke-Data.xls dont la source est Kaggle — Stroke Prediction Dataset (fedesoriano) et le lien de téléchargement est https://www.kaggle.com/datasets/fedesoriano/stroke-prediction-dataset. Apres nettoyage, ce jeu de données contient 5109 individus catégorisés en deux classes dont :

- Les individus avec risque d'AVC (classe positive)
- Les individus sans risque d'AVC (classe négative)

Les 10 caractéristiques retenues après nettoyage sont : gender (le sexe), age (l'âge), hypertension (l'hypertension), heart_disease (les antécédents de maladies cardiaques), ever_married (la situation matrimoniale (marié(e) ou célibataire)), work_type (la profession), Residence_type (le type de logement), avg_glucose_level (la glycémie moyenne), bmi (l'IMC), smoking_status (le statut tabagique (si l'individu fume ou pas))., sans tenir compte du degré de ce risque.

4.1.2. Prétraitement des données

(1) Suppression de la colonne id ;
(2) Suppression de l'unique individu de genre « Other »;
(3) Imputation de l'IMC manquant par la médiane ;
(4) Encodage des variables catégorielles (gender → binaire, ever_married → binaire, work_type → ordinal 0–4, Residence_type → binaire, smoking_status → ordinal 0–3) ;
(5) Normalisation *StandardScaler* ajustée sur l'entraînement uniquement.

4.1.3. Nature des données

Chaque individu dans le jeu de données peut être formalisé par : $(x_i = (x_{ij}), y_i)$ *où* $x_i \in R^{10}$ *et* $y_i \in \{0,1\}$

Avec :

- $j \in \{1 \dots 10\}$
- 1 : étiquette de la classe positive, c.à.d. individu avec risque d'AVC
- 0 : étiquette de la classe négative c.à.d. individu sans risque d'AVC

4.1.4. Partitionnement du jeu de données

Le jeu de données est composé de 4 861 (95,13 %) de la classe négative et 249 cas (4,87 %) de la classe positive pour rester dans le contexte de la classification asymétrique binaire. Le déséquilibre significatif (4,87 % de cas positifs) justifie une pénalité asymétrique $C_+ > C_-$. Ce jeu de données est reparti comme suit :

- 70% ou 3575 individus pour l'entraînement dont 3400 de classe négative et 175 de classe positive ;
- 15% ou 766 individus pour la validation dont 729 de classe négative et 37 de classe positive
- 15% ou 766 individus pour le test dont 729 de classe négative et 37 de classe positive

4.2. Définition du modèle SVM

Chaque modèle SVM est défini par le noyau gaussien exprimé par la relation (7) de paramètre γ et les coefficients de

régulation C_+, C_- respectivement pour la classe 1 et la classe 0 ci-dessus définies, qui sont considérés comme les hyper paramètres du modèle SVM.

En posant par $\theta = (\gamma, C_+, C_-)$, les paramètres internes $\alpha_i^*(\theta)$ et $b^*(\theta)$, respectivement multiplicateurs optimaux de Lagrange et le biais optimal de l'hyperplan à marge maximale sont déterminés pendant l'entrainement du modèle [11].

4.3. Résultats, Analyse et discussion
4.3.1. Optimisation des hyper paramètres

Baseline

Le modèle NSGA-II-SVM est évalué en comparaison avec la méthode GridSearchCV entraînée sur le même jeu d'entraînement, validée et testée sur le même jeu de test afin de garantir une comparaison équitable.

SVM monocritère

Le SVM de référence est optimisé à l'aide de GridSearchCV, avec l'AUC comme critère de sélection et une validation croisée à 3 plis. La grille d'hyper paramètres explorée est définie par $C_+, C_- \in \{0{,}1; 1; 10; 100; 500\}$ et $\gamma \in \{0{,}001; 0{,}01; 0{,}1; 1\}$.

Ce modèle repose sur le même classifieur de base que NSGA-II-SVM, garantissant une comparaison directe et équitable. Il constitue ainsi un point de référence représentatif des résultats qu'un praticien obtiendrait sans stratégie d'optimisation bi-critère.

Matrice de confusion du modèle optimisé par GridSearchCV

Table 1 : matrice de confusion 1

	Avec risque(1)	Sans risque(0)	Total réel
Avec risque d'AVC (1)	0	37	37
Sans risque d'AVC (0)	729	0	729

Sur 37 cas positifs, le modèle optimisé par GridSearchCV n'a prédit aucun cas positif. Alors que sur 729 cas négatifs, le modèle a prédit, tous les cas réels négatifs, négatifs. L'exactitude du modèle est donnée $=\frac{729}{766} \cong 0{,}95$ ou 95%

- **Métriques de performance**

Les indicateurs de performance du modèle optimisé par GridSearchCV se présentent comme suit :

(1) Spécificité= 0%
(2) Sensibilité=100%
(3) F1-score=0%

Convergence de NSGA-II

L'initialisation de la taille de la population N et du nombre de génération G est cruciale car elle définit l'équilibre entre la qualité de l'approximation du front de Pareto et le temps de calcul [16].

Initialisation des paramètres [17]

- Taille de population : $N = 40$
- Nombre de générations : $M = 60$
- Borne $C_+ = [1; 5000]$ car favorise la sensibilité quand C_+ élevé
- Borne $C_- = [1; 500]$ Pénalité plus faible pour les faux positifs
- Borne $\gamma = [10^{-4}; 1]$ Plage standard pour données normalisées

La fonction d'évaluation à deux critères calcule les faux positifs (**NFA**) et les faux négatifs (**NFR**) sur le jeu de validation (15 % des données). Le jeu de test (15 %) reste strictement séparé jusqu'à l'évaluation finale afin d'éviter un biais d'optimisme. La répartition en jeux d'entraînement, de validation et de test est stratifiée (les proportions de chaque classe sont conservées).

Tenant compte de la taille et du partitionnement de notre jeu de données de validation, de la population initiale à la population finale ou front de Pareto complet obtenu par NSGA-II, les deux fonctions objectifs de chaque solution non dominée (ici modèle SVM) vérifient les conditions suivantes :

- $0 \leq NFA \leq 37$
- $0 \leq NFR \leq 729$

Ces deux conditions sont aussi vérifiées dans la table 1 du SVM Baseline validé par le même jeu de données et illustrées dans la figure 2 ci-dessus.

NSGA-II atteint un front de Pareto complet composé de 40 solutions non dominées dès la 10e génération, et ce front maintient sa stabilité jusqu'à la 60e génération.

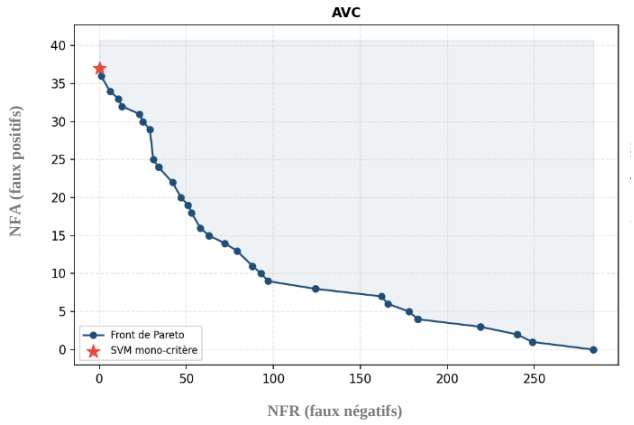

Fig.2 : *Front de Pareto complet des modèles SVM*

Au total, 2 287 SVM distincts ont été entraînés (sur un maximum théorique de 2 400 = 40 × 60), le cache évitant 113 évaluations redondantes (environ 5 %).

Les 40 solutions (SVM) du front de Pareto sont non dominés les uns aux autres au sens de Pareto, c.à.d.

- $\forall\, \theta, \theta'$, θ domine $\theta' \Leftrightarrow (\mathrm{NFA}(\theta) < \mathrm{NFA}(\theta')$ et $\mathrm{NFR}(\theta) \leq \mathrm{NFR}(\theta'))$ ou $(\mathrm{NFA}(\theta) \leq \mathrm{NFA}(\theta')$et $\mathrm{NFR}(\theta) < \mathrm{NFR}(\theta'))$
- $\forall\, \theta$, θ est non dominé si la condition ci-dessus n'est pas vérifiée.

Avec $\theta = (C_+, C_-, \gamma)$

Il est à noter que la solution $\theta = (C_+, C_-, \gamma)$ du front complet est de meilleur compromis si et seulement $d(\theta, \theta')$ est minimale. [16]

Avec

- $\theta' = (C'_+, C'_-, \gamma')$ Correspondant à $(\mathrm{NFA}_{\min}, \mathrm{NFR}_{\min}), \forall\, \theta$ appartenant au front de pareto complet, la solution idéale du front de Pareto complet.
- $d(\theta, \theta')$: Distance euclidienne entre θ et θ'

Partant de la figure 2, le SVM idéale est atteint au point (NFA_{min}=0, NFR_{min}=0).

Apres navigation sur le front de Pareto complet obtenu, le modèle SVM de meilleur compromis est défini par NFA_{min}=17, NFR_{min}=54.

Matrice de confusion du modèle SVM optimisé par NSGA-II

La matrice de confusion observée pour le modèle SVM de meilleur compromis est donnée par le tableau ci-dessous dont en ligne, les classes réelles et en colonne, les classes prédites :

Table 2 : matrice de confusion 2

	Avec risque(1)	Sans risque(0)	Total réel
Avec risque d'AVC (1)	20	17	37
Sans risque d'AVC (0)	54	675	729

Sur 37 cas positifs, le modèle optimisé a prédit correctement 20 avec une erreur de 17 cas ou 46%. Alors que sur 729 cas négatifs, le modèle a prédit correctement 675 cas avec une erreur de 54 cas ou 7%. L'exactitude du modèle est donnée $=\frac{20+675}{766} \cong$ 0,91 ou 91

Métriques de performance par classe du modèle SVM optimisé par NSGA-II

Les indicateurs de performance du modèle optimisé par NSGA-II sont présentés comme suit:

(1) Spécificité= 0,926 ou ≅ 93%
(2) Sensibilité=0,540 ou 54%
(3) F1-score=0,360 ou 36%

4.3.2. Discussion des résultats

GridSearchCV Vs NSGA-II

Dans le contexte de la classification asymétrique, l'exactitude du modèle SVM est une métrique non appropriée pour évaluer la performance du modèle avec risque de favoriser la classe majoritaire au détriment de la classe minoritaire. Par contre, le F1-score qui exprime la moyenne harmonique de la précision et de rappel, indique si le modèle est réellement performant sur la classe d'intérêt et détermine l'équilibre entre la précision et la sensibilité du modèle. Il permet d'assurer que le SVM ne sacrifie pas l'un pour l'autre. Un F1-score élevé signifie que le modèle est robuste et équilibré. [4], [19]

Les métriques de performance sont synthétisées par la figure suivante :

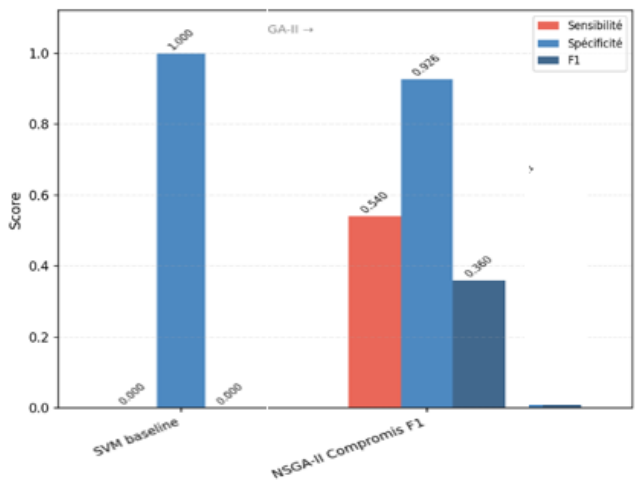

Fig.3 : *comparaison des métriques de performance entre SVM Baseline et SVM/NSGA-II*

Table 3 : SVM Baseline vs SVM/NSGA-II

Approche	Sensibilité	Spécificité	F1 score
SVM Baseline	0,000	1,000	0,000
SVM/ NSGA-II	0,540	0,926	0,360

Premièrement, le SVM optimisé par un seul critère (approche de référence), bien qu'avec une exactitude de 95%, obtient un score F1 de 0,000. Le modèle ne détecte aucun cas positif mais classe tous les individus négatifs, négatifs, ce qui peut être cohérent avec l'optimisation de l'AUC dans un contexte de déséquilibre important (4,87 % de cas positifs), mais s'avère totalement inefficace du point de vue clinique. Ce comportement illustre clairement les limites des stratégies d'optimisation monocritère sur des ensembles de données fortement déséquilibrés.

Deuxièmement, la solution NSGA-II-SVM de meilleur compromis dans le front de Pareto complet, également appelée « Compromis F1 », bien qu'avec une exactitude inférieure à celle du SVM Baseline de 93%, a obtenu le meilleur score F1 parmi toutes les méthodes évaluées (F1 = 0,360, sensibilité = 0,540, spécificité = 0,926). Du point de vue clinique, Ce résultat confirme que l'optimisation bi-critère surpasse la méthode de référence monocritère tout en améliorant simultanément les deux objectifs NFA et NFR.

Les expériences confirment l'efficacité de l'approche par optimisation bi-critère NSGA-II-SVM pour la prediction du risque d'AVC. Alors que le SVM monocritère, malgré sa réputation de performance, échoue sur un jeu de données fortement déséquilibré (F1 = 0), le NSGA-II-SVM génère des modèles pertinents, offrant un large éventail de compromis cliniquement exploitables. Le front de Pareto dépasse la simple optimisation algorithmique : il constitue un véritable outil d'aide à la décision médicale, permettant à l'expert de gérer explicitement le compromis entre faux positifs et maladies non détectées.

5. Conclusion et perspectives

Dans notre article, nous avons développé une approche bi-critère basée sur la Meta-heuristique NSGA-II pour améliorer la performance ou l'efficacité du modèle SVM dans le contexte de classification asymétrique binaire pour la prediction du risque d'AVC, qui repose sur le choix des hyper paramètres du modèle. [20], [21]

Au regard des résultats obtenus par notre expérimentation , réalisée en Python, sur des données cliniques déséquilibrées pour la prediction du risque d'AVC, l'optimisation de référence des hyper paramètres GridSearchCV, une approche monocritère dont le critère de sélection est AUC, nécessite une connaissance préalable des couts de classification ainsi que le paramètre du noyau, offre un modèle unique SVM moins performant cliniquement avec un F1 score de 0% sacrifiant ainsi la classe minoritaire (cas positifs) au profit de la classe majoritaire(cas négatifs) bien que son exactitude est de 95%. Contrairement à l'approche bi-critère basée sur l'algorithme évolutionnaire NSGA-II, visant à trouver un compromis optimal entre le taux ou nombre de faux positifs et le taux ou nombre de faux négatifs considères comme les deux fonctions objectif, ne nécessite aucune connaissance préalable des valeurs des hyper paramètres et offre un front de Pareto complet composé de plusieurs modèles SVM non dominés dont le modèle de meilleur compromis, bien qu'avec une exactitude inférieure au SVM Baseline de 93% mais un F1 score élevé de 0,360 , offre une performance cliniquement efficace au SVM Baseline. Ce résultat confirme notre hypothèse selon laquelle le modèle optimisé par NSGA-II est plus performant et plus efficace que le modèle optimisé par GridSearchCV [8].

Plus tard, nous envisageons continuer cette recherche sur plusieurs axes, notamment sur la classification SVM multiclasse probabiliste avec plus avec plus de deux critères sur de données cliniques de plusieurs natures pour la prediction d'autres pathologies comme le cancer ou autres.

Références bibliographiques

[1] Collectif de recherche en IA médicale, « Machine Learning for Stroke prediction using Imbalanced data », *PubMed Central (PMC)/NIH*, 2025
[2] Tazin T. et al. « Stroke Prediction using Machine Learning and Explainable AI : A Binary classification approach », International journal of Advanced computer science and application, 12(1), 2021
[3] Bhardwaj A. & Gupta S., « Multi-Objective Optimization of SVM Hyperparameters using NSGA-II for Imbalanced Healthcare Data », *Expert Systems with Applications*, 2023
[4] Chen H. & Wang X., « Asymetric Classification of Cerebrovascular Diseases : A Bi-objective Genetic Algorithm Perspective », *IEEE Transactions on Nanobioscience*, Vol 21, n°2, pp 245-256,2022

[5] Moradpour M. et al. « Multi-Objective Performance Optimization of Machine Learning Models in Healthcare », *Studies in Health Technology and Informatics*, 2024
[6] Rodriguez J.L. & Chen Y., « Addressing Class imbalance in Stroke prediction : A Cost-sensitive SVM Approach », *Journal of Biomedical Informatics*, 2022
[7] M. Sekkal, A. Benzina, L. Benkrelifa , « Multi-Objective Evolutionary Algorithm Based on NSGA-II for Neural Network Optimization application to the prediction of severe diseases », *Informatica*, 2023
[8] Garcia L. & Martinez P., « Hyperparameters Tuning for Support Vector Machine in Stroke prediction : Comparaison between GridSearch and Meta-heuristics », *Informatics in Medicine Unlocked*, vol 24, 100582, 2021
[9] Shoer S. et al. « Machine Learning Architectures for Binary Stroke Prediction : A comparative Analysis of neural networks and Ensemble Methods », *Journal of Medical Systems*, 48(2), pp. 112-128, 2024
[10] Zhang L. et al, « Predicting Acute Ischemic Stroke with Artificial Intelligence : Progress in Binary Classification Models and clinical implementation », *IEEE Journal of Biomedical and Health Informatics*, 27(4), pp 1945-1956, 2023
[11] Nguyen T.T. & Zhang J., « Handling Class Imbalance in Stroke Datasets using cost-sensitive support Vector Machines and Evolutionary Algorithm », *Applied Soft Computing*, vol. 132, 109864, 2023
[12] Rodriguez J. & Silva M., « Cost-sensitive SVM Optimization using Multi-Objective Evolutionary Algorithms for Asymetric classification », *Expert Systems with Application*, 2023
[13] Gupta R., Kumar S. & Taylor P., « Explainable AI in Stroke Diagnosis : Multi-Objective Hyperparameters Tuning for Reliable SVM Models », *Artificial Intelligence in Medicine* , 2024
[14] Catalin Stocan et Ruxandra Stocan, « Support vector Machines and Evolutionary Algorithms for classification : Single or Together ? », *Springer* ,2020
[15] Kalyanmoy Deb, « Multi-Objective Optimization Using Evolutionary Algorithms (3rd Edition) », *Wiley* , 2023
[16] Ma H., Zhang Y., Sun S., Liu T. & Shan Y., « A Comprehensive Survey on NSGA-II for Multi-Objective Optimization and applications », *Artificial Intelligence Review*, 56, 15217-15270, 2023
[17] Yang Y., & Men J., « Multi-Objective Optimization of Construction Project Management Based on NSGA-II Algorithm Improvement », *International Journal of Advanced Computer Science and Applications (IJACSA)*, 16(1), 2025
[18] Deb K. & Kalyan A., « Evolutionary Multi-Objective Optimization : Beyond NSGA-II for complex Medical Diagnostics », *Evolutionary Computation Journal, MIT Press*, 2025
[19] Patel M. et al. (2020), « A Systematic Review of Machine Learning Technics for Stroke prediction : Performance and Scalability », *Healthcare Informatics Research*, vol 26, n°3, pp 160-175, 2020
[20] Arjun Panesar, « Machine Learning and AI in Healthcare : Data, Models and Applications », *Apress*, 2021
[21] Fernandes et al, Rochmawati & Maryani, « A comprehensive Machine Learning Framework for Stroke risk prediction », *Global Academic Journal of Medical Sciences*, 2026
[22] Acharya U. R. & Oh S. L., « Deep learning and SVM hybrids for stroke risk stratification : A Multi-Objective Approach », *Journal of clinical Medicine*, Vol 13, n°4, pp 112-128, 2024
[23] Divers contributeurs (sous la direction de springer nature), « Interpretable Machine Learning for Healthcare », *Springer*, 2025
[24] Karthik N. & Srimathi C., « An intelligent Framework for Stroke Prediction using Hybrid Feature Selection and Machine Learning Classifiers », *Computers in Biology and Medicine*, 146, 105612, 2022
[25] Liu Y. & Singh R., « Real-Time Stroke Risk Assessment using Optimized SVM on Edge Devices », *Computers in Biology and Medicine*, 2026
[26] Zhang L., Chen Y. & Wang H., « A Hybrid NSGA-II and SVM Model for Stroke Risk prediction with Imbalanced Medical data », *Journal of Biomedical Informatics*, 2021
[27] Zhao Q. & Li S. (2024), « Interpretable Multi-Objective Optimization for Medical SVMs : Balancing Accuracy and Transparency », *Artificial Intelligence in Medicine*, Vol 142, 102582,2024