JOURNAL AFRICAIN DES SCIENCES
VOL. 3, **3** (2026) 73-81

https://jafrisci.org

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

OPEN ACCESS

ARTICLE DE RECHERCHE

Article Info.:
Reçu : le 02/03/2026
Accepté: le 31/03/2026
Publié : le 04/04/2026

MISE EN PLACE D'UN SYSTÈME HYBRIDE MLP, DBSCAN, SVM, POUR LA PRISE DE DÉCISION SUR LA CONFORMITÉ DES PRODUITS FRAIS, CAS DU LABORATOIRE DE MICROBIOLOGIE DE L'OFFICE CONGOLAIS DE CONTRÔLE (OCC)

Michel Ange Mpalanga Mfwamba [1,*], **Pierre Katalayi Kafunda** [1], **Gaby Mampuya Lubiba** [2]

https://doi.org/10.70237/jafrisci.2026.v3.i3.08

Résumé

Dans cet article, nous proposons une architecture hybride innovante qui combine un réseau de neurones multicouche (MLP), l'algorithme de clustering DBSCAN, et un classificateur SVM. Cette approche vise à améliorer la robustesse et la précision de classification en intégrant une extraction de connaissances non linéaire, une détection automatique des valeurs aberrantes et une décision linéaire dans un espace optimisé. Nous détaillons la formulation mathématique de chaque composant et démontrons comment leur intégration permet une amélioration significative des performances sur des données bruitées. Les résultats expérimentaux montrent que l'approche MLP - DBSCAN - SVM surpasse les architectures plus classiques telles que MLP - SVM et DBSCAN - SVM, en atteignant une précision de 93,78 %, contre 92,0 % et 75,5 % respectivement. Il confirme la pertinence d'un pipeline intelligent et séquentiel alliant extraction de connaissances non linéaire, réduction du bruit et classification supervisée pour la déclaration de conformité de produit frais de consommation par la classification automatique via une architecture hybride MLP – DBSCAN - SVM, application au laboratoire de Microbiologie de l'OCC.

Mots-clés: Intelligence Artificielle, classification, architecture hybride, MLP, DBSCAN, SVM, OCC

Abstract

In this article, we propose an innovative hybrid architecture that combines a multilayer perceptron (MLP), the DBSCAN clustering algorithm, and an SVM classifier. This approach aims to improve classification robustness and accuracy by integrating nonlinear feature extraction, automatic outlier detection, and linear decision-making in an optimized space. We detail the mathematical formulation of each component and demonstrate how their integration leads to a significant improvement in performance on noisy data. Experimental results show that the MLP-DBSCAN-SVM approach outperforms more traditional architectures such as MLP-SVM and DBSCAN-SVM, achieving an accuracy of 93.78%, compared to 92.0% and 75.5% respectively, confirming the relevance of an intelligent, sequential pipeline combining nonlinear feature extraction, noise reduction, and supervised classification for the declaration of conformity of consumer products through automatic classification using a hybrid MLP, DBSCAN, SVM architecture, applied at the OCC Microbiology Laboratory.

Keywords: Artificial Intelligence, classification, hybrid architecture, MLP, DBSCAN, SVM, OCC

1. Introduction

En République Démocratique du Congo, l'Office Congolais de Contrôle qui est un organisme technique à caractère scientifique dont la mission est de veiller à la conformité de tous les produits importés ou exportés, en procédant par des essais de laboratoire chimique et microbiologique. Par ces les différentes frontières avec les pays voisins et ses postes d'entrée la République Démocratique du Congo fait face un flux élevé d'importation, le commerce du pays est en grande parte extravertie, avec une forte dépendance des produits importés utilisés dans tous les secteurs et touche toutes les couches de la société : dans le commerce interne, la consommation. Il est d'une exigence capitale qu'ils soient contrôlés de manière systématiquement en assurant que les produits soient de bonne qualité et respectant les exigences nationales, régionales et internationale, car la vie de la population en dépend [1].

La tâche principale est de procéder à l'analyse ces données

Correspondance : *p.mfwamba@gmail.com (**M. A. M.** **Mfwamba**)*

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

[1] *Mathématiques, Statistiques et Informatique, Faculté des Sciences et Technologie, Université de Kinshasa, République Démocratique du Congo*
[2] *Département de Laboratoire, Office Congolais de Contrôle, République Démocratique du Congo*

issues des produits frais de consommation, des analyses laboratoire pour faciliter la prise de décision sur la conformité de produit à la consommation à l'aide la classification supervisée qui est un domaine central de l'apprentissage automatique. Toutefois, la présence de bruit et de données aberrantes affecte souvent la performance des modèles classiques. Les modèles traditionnels tels que le SVM ou les réseaux de neurones profonds sont souvent sensibles ces anomalies. Pour répondre cette problématique, nous proposons une architecture hybride combinant trois puissantes techniques dont le MLP pour l'extraction automatique des connaissances, l'algorithme DBSCAN pour l'élimination les outliers et le SVM pour la classification finale dans un espace filtré. Cette stratégie permet d'exploiter la richesse des représentations apprises tout en contrôlant la sensibilité aux anomalies [2].

Cependant les résultats nous montrent clairement que l'approche proposée a été appliquée sur un ensemble de données réelles issues du laboratoire de microbiologie de l'Office Congolais de Contrôle avec une précision de 93,78 %, contre 92,0 % et 75,5 % résultant des respectivement des approche MLP-SVM, et DBSCAN-SVM, confirmant la pertinence d'un pipeline intelligent et séquentiel alliant extraction de connaissances non linéaire, réduction du bruit et classification supervisée pour la déclaration de conformité de produit de consommation par la classification automatique via une architecture hybride MLP, DBSCAN, SVM, application au laboratoire de Microbiologie de l'OCC [3][4].

En outre l'introduction et la conclusion, notre article est subdivisé en 4 sections ci-après l'état de l'art, la méthodologie pour boucler par le résultat et les discussions.

2. Etat de l'art

Pour parler de la revue de littérature nous allons explorer quelques travaux : Des études ont montré que La classification intelligente des données constitue un axe central dans le cadre de l'Industrie actuelle, où la qualité des produits et la conformité réglementaire doivent être assurées malgré la complexité croissante des données [8],[9]. Les techniques d'intelligence artificielle (IA), en particulier les modèles hybrides, sont devenues des solutions de choix pour répondre aux limites des méthodes traditionnelles.

Le Perceptron Multicouche (MLP) est largement utilisé pour l'extraction non linéaire de caractéristiques, notamment en raison de sa capacité à capturer les relations complexes entre les variables d'entrée [5]. Dans le contexte industriel, plusieurs études ont démontré l'efficacité des MLP pour prédire la qualité des produits ou identifier des défauts dans les chaînes de production [6]. Cependant, ces modèles sont sensibles aux données bruitées ou déséquilibrées, ce qui peut affecter leur performance.

Le Density-Based Spatial Clustering of Applications with Noise, introduit par Ester et al., est une méthode de regroupement non supervisée capable d'identifier des structures complexes dans les données sans nécessiter de connaissance préalable du nombre de clusters. Il est particulièrement efficace pour filtrer les outliers, ce qui est essentiel dans les environnements industriels où les capteurs peuvent générer des valeurs aberrantes. Plusieurs travaux ont intégré DBSCAN comme étape de prétraitement avant la classification afin d'améliorer la qualité des données d'apprentissage [7].

Les SVM sont réputés pour leur capacité maximiser la marge de séparation entre les classes et bien généraliser sur des jeux de données de faible dimension [10]. Dans les systèmes de contrôle qualité, ils sont largement utilisés pour la classification binaire et multi-classe, notamment lorsqu'il s'agit de distinguer les produits conformes des non conformes [11]. Néanmoins, leur performance dépend fortement de la qualité des données d'entrée.

L'association d'un MLP avec un SVM a été explorée dans de nombreuses études, notamment pour bénéficier de la puissance du MLP dans l'extraction de caractéristiques, suivie d'une classification robuste par SVM. Par exemple, Zhang et al. ont démontré que cette combinaison permettait d'améliorer la précision dans des contextes de détection de défauts dans des chaînes de montage [12].

De leur côté, les modèles DBSCAN-SVM ont été utilisés dans des travaux où la qualité des données était fortement impactée par le bruit ou les valeurs extrêmes [13]. Cette combinaison permet de réduire l'impact des outliers avant d'appliquer une classification stricte.

Par ailleurs, d'autres études ont montré que la combinaison tripartite MLP–DBSCAN–SVM, qui est l'objet de ce travail, reste encore peu explorée dans la littérature. Elle s'inscrit dans une logique d'optimisation en trois phases : extraction de caractéristiques non linéaires via le MLP, filtrage et détection des anomalies avec DBSCAN, et classification robuste à l'aide du SVM [12][13][14]. Cette structure permet de traiter efficacement des données complexes, non linéaires et bruitées, ce que confirme notre travail expérimental sur les jeux de données de laboratoire de microbiologie.

3. Méthodologie

3.1. Extraction de connaissances via MLP

L'extraction de connaissances à l'aide d'un réseau de neurones multicouches (MLP) constitue une étape cruciale dans notre méthodologie. Les MLP sont capables d'apprendre des représentations complexes des données d'entrée, ce qui permet de capturer des caractéristiques pertinentes pour la tâche de classification [12].

Soit $x_i \in \mathrm{R}^d$ une donnée d'entrée, le MLP est une fonction $f(x_i;\theta)$ composée de plusieurs couches :

$$z_i = f(xi;\theta) = \varphi L(W(L) \cdots \varphi 1(W(1)xi + b(1)) + \cdots + b(L))$$

Dans cette équation, $z_i \in \mathrm{R}^k$ représente le vecteur de caractéristiques, également appelé embedding. L'importance de cette étape réside dans le fait que les embeddings permettent de transformer les données d'entrée dans un espace où elles peuvent être plus facilement séparables, ce qui est essentiel pour la classification. Les fonctions d'activation non linéaires, notées φ_l, sont cruciales car elles permettent au réseau d'apprendre des relations complexes et non linéaires dans les données.

Enfin, $\theta = \{W^{(l)}, b^{(l)}\}_{l=1}^{L}$ désigne les paramètres du

réseau, qui sont appris pendant l'entraînement et déterminent la capacité du réseau à extraire des caractéristiques pertinentes. L'utilité principale du MLP dans cette méthodologie est de transformer les données d'entrée en un ensemble de caractéristiques qui facilitent la classification ultérieure.

3.2. Filtrage des outliers par DBSCAN

Le filtrage des outliers l'aide de DBSCAN est une étape importante pour améliorer la qualité des données utilisées pour l'entraînement du classificateur. DBSCAN est un algorithme de clustering basé sur la densité qui permet de détecter et de filtrer les points aberrants dans l'ensemble des données d'entrée ou embeddings(connaissances extraites). DBSCAN est appliqué sur l'ensemble des embeddings $Z = \{z_i\}_{i=1}^{N}$. Pour chaque z_i, on définit le voisinage $N_\varepsilon(z_i)$ comme l'ensemble des points situés à une distance inférieure ou égale ε :

$$N_\varepsilon(z_i) = \{z_j \in Z \mid \|z_i - z_j\| \leq \varepsilon\}$$

Le masque $m(z_i)$ est donné par :

$$m(z_i) = \begin{cases} 1, & \text{si } z_i \text{ est un core ou border point} \\ 0, & \text{si } z_i \text{ est un outlier} \end{cases}$$

Les données z_i conservées sont celles pour lesquelles $m(z_i) = 1$:

$$Z_{clean} = \{(z_i, y_i) \mid m(z_i) = 1\}$$

Un point est considéré comme un point cœur (core point) si son voisinage contient au moins MinPts points. Les points ne satisfaisant pas cette condition et n'appartenant à aucun cluster sont considérés comme des outliers et sont éliminés avant l'entraînement du modèle SVM.

L'importance de cette étape réside dans le fait que le filtrage des outliers permet de réduire le bruit dans les données et d'améliorer la performance du classificateur. Les outliers peuvent introduire des erreurs et des biais dans l'entraînement du modèle, ce qui peut nuire sa capacité à généraliser [7].

3.3. Classification par SVM

La classification par SVM (Support Vector Machine) est une étape clé de notre méthodologie. Les SVM sont des modèles de classification supervisée qui cherchent à trouver l'hyperplan optimal séparant les différentes classes dans l'espace des caractéristiques. Sur les données nettoyées, nous apprenons un classificateur SVM en résolvant le problème d'optimisation suivant :

$$\min_{w,b,\xi} \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{N} m(z_i)\cdot\xi_i$$

Sous les contraintes :

$$y_i(w^T z_i + b) \geq 1 - \xi_i, \qquad \xi_i \geq 0$$

Dans ce contexte, $w \in R^k$ et $b \in R$ représentent les paramètres de l'hyperplan de séparation. L'importance de ces paramètres réside dans le fait qu'ils définissent l'hyperplan qui sépare les différentes classes dans l'espace des caractéristiques. Les variables ξ_i sont des variables de relâchement qui permettent de gérer les erreurs de classification et sont cruciales pour trouver un compromis entre la maximisation de la marge et la minimisation des erreurs. Enfin, $C > 0$ est le paramètre de régularisation qui contrôle le compromis entre la maximisation de la marge et la minimisation des erreurs de classification, ce qui est essentiel pour éviter le sur- apprentissage. L'utilité principale des SVM dans cette méthodologie est de fournir un modèle de classification robuste et performant, avec possibilité de généralisation sur des nouvelles données [10].

3.4. Fonction de décision globale

La fonction de décision globale est utilisée pour prédire la classe d'une nouvelle donnée d'entrée. Cette fonction prend en entrée les caractéristiques extraites par le MLP et utilise l'hyperplan appris par le SVM pour prédire la classe. Le modèle prédit la classe $\hat{y}_i$ comme suit : $\hat{y}_i = \text{sign}(w^T f(x_i;\theta) + b)$.

L'importance de cette étape réside dans le fait qu'elle permet de combiner les résultats des étapes précédentes pour fournir une prédiction finale. La fonction de décision globale est cruciale pour évaluer la performance du modèle et pour faire des prédictions sur de nouvelles données. Il est à noter que le masque de DBSCAN n'apparaît pas explicitement dans cette formule car il est utilisé uniquement pendant la phase d'entrainement pour nettoyer les données. Une fois le modèle SVM entraîné, il n'est plus nécessaire d'appliquer le masque pour faire des prédictions sur de nouvelles données.

4. Résultats

Nous utilisons un jeu de données réel issu des résultats d'analyses microbiologiques de produit, qui a subi un prétraitement avec son utilisation.

L'évaluation expérimentale a porté sur trois architectures distinctes : MLP - SVM, DBSCAN - SVM et notre modèle hybride proposé MLP – DBSCAN - SVM. Les performances de ces modèles ont été comparées l'aide du jeu de données sus évoqué. Les métriques d'évaluation utilisées sont la précision (accuracy), le rappel (recall) et le score F1.

4.1. Expérimentale setup

Nous avons utilisé un ordinateur portable HP (HDD 1Terra, RAM 8 Giga et CPU I5), les outils suivants ont été utilisés ; l'environnement Colab (Colaboratory) qui est un service Cloud offert par Google (Gratuit) basé sur jupyter Notebook, le langage de programme Python pour le prétraitement de données ainsi que les différentes techniques de l'apprentissage artificiel [17]. Les données d'analyses microbiologiques de produits frais, elles sont tirées des rapports d'essai originaux et base de données de l'application de laboratoire OCC-RENAL de l'Office Congolais de Contrôle, la pour la période allant de 2023 à 2024 [15]. Nous avons développé une méthodologie axées sur quatre(4) phases comprenant la collecte des données, prétraitement des données, la modélisation de l'apprentissage et l'application de la prise de décision.

Notre source de données, est un jeu de données constitué de 1124 observations réparties en deux classes : les produits conforme (20%) et non conforme (80%). Chaque observation est décrite par 26 caractéristiques numériques entières, issues des résultats d'analyses microbiologiques au laboratoire. Après

une étape de nettoyage des données (suppression des doublons, valeurs manquantes et aberrantes), la méthode SMOTE (Synthetic Minority Oversampling Technique) [16] a été appliquée pour équilibrer les classes. L'échantillon final contient donc un nombre équilibré d'exemples de chaque classe.

Table 1. Description des attributs de notre source de données.

N	Variable	Signification
1	aspect	Apparence générale du produit
2	couleur	Couleur observée du produit
3	odeur	Odeur émise par le produit
4	saveur	Goût du produit
5	consistance_musculature	Fermeté de la chair / texture musculaire
6	temperature_labo	Température mesurée au laboratoire
7	etat_yeux	tat visuel des yeux du poisson (ou produit)
8	aspect_branchies	Aspect des branchies (couleur, texture)
9	etat_anus	tat de l'anus (rel ché, fermé, etc.)
10	aspect_os	Apparence des os du produit
11	Bacillus cereus 30 C	Présence de Bacillus cereus à 30 C
12	Coliformes 30 C	Dénombrement des coliformes totaux 30 C
13	E. Coli B-glicuronidase + 44 C	Présence d'E. coli spécifique à 44 C
14	Entérobactéries 37 C	Dénombrement global des entérobactéries
15	Microorganismes 30 C	Charge microbienne totale 30 C
16	Pseudomonas spp 25 C	Présence de Pseudomonas spp 25 C
17	Salmonella spp	Recherche de Salmonella (hors S. Typhi et S. Paratyphi)
18	Staphylocoque coagulase + 37 C	Présence de staphylocoques pathogènes
19	Bactéries lactiques 30 C	Dénombrement des bactéries lactiques
20	Clostridium perfringens 37 C	Présence de Clostridium perfringens
21	Listeria monocytogènes	Détection de Listeria monocytogènes
22	Moisissures et Levures 25 C	Présence de champignons et levures 25 C
23	Pseudomonas spp 25 C	Redondant (possiblement une duplication)
24	Staphylocoque coagulase + 37 C	Redondant (voir ligne plus haut)
25	Conclusion	Evaluation finale (conforme / non conforme)

Les variables considérées incluent : aspect, couleur, odeur, saveur, consistance musculature, temperature labo, conditionnement echantillon labo, non conf_av_analyse, etat_yeux, aspect_branchies, etat_anus, aspect_os, Bacillus cereus 30 C, Coliformes 30 C, E. Coli B-glicuronidase+ 44 C, Entérobacteries 37 C, Microorganismes 30 C, Salmonella spp, lactiques 30 C, Clostridium perfringens 37 C, Lactobacilles, Listeria monocytogenes, Moisissures et Levures 25 C, Pseudomonas aeruginosa, Pseudomonas spp 25 C, Staphylocoque coagulase+37 C

Table 2. Paramètre statistique de notre Dataset, source : Nos données de recherche

Variable/Index	Count	Mean	Std	Min	25%	50%	75%	Max
aspect	1124.0	0.8	0.3826	0.0	1.0	1.0	1.0	1.0
couleur	1124.0	1.9	0.9597	0.0	1.0	2.0	2.0	5.0
odeur	1124.0	0.4	0.4933	0.0	0.0	0.0	1.0	1.0
saveur	1124.0	0.0	0.0666	0.0	0.0	0.0	0.0	1.0
consistance_musculature	1124.0	0.4	0.5346	0.0	0.0	0.0	1.0	2.0
temperature_labo	1124.0	13.23	19.7285	-116.67	-27.44	-17.03	-15.80	27.0
etat_yeux	1124.0	0.004	0.0894	0.0	0.0	0.0	0.0	2.0
aspect_branchies	1124.0	0.005	0.0942	0.0	0.0	0.0	0.0	2.0
etat_anus	1124.0	0.004	0.0894	0.0	0.0	0.0	0.0	2.0
aspect_os	1124.0	0.522	0.6031	0.0	0.0	0.0	1.0	3.0
(*) Bacillus cereus 30 C	1124.0	-0.01	0.0422	-1.0	0.0	0.0	0.0	0.0
(*) Coliformes 30 C	1124.0	-0.06	0.2398	-2.0	0.0	0.0	0.0	0.0
(*) E. Coli B-glicuronidase + 44 C	1124.0	-0.81	0.4102	-2.0	-1.0	-1.0	-1.0	0.0
(*) EntØrobactØries 37 C	1124.0	-0.46	0.4932	-2.0	-1.0	-0.1	0.0	1.0
(*) Microorganismes 30 C	1124.0	-1.18	1.0213	-4.0	-2.0	-1.0	0.0	1.0
(*) Pseudomonas spp 25 C	1124.0	-0.4159	0.5159	-2.0	-1.0	0.0	0.0	0.0
(*) Salmonella spp (hors S. Typhi et Paratyphi)	1124.0	-0.01	0.0985	-1.0	0.0	0.0	0.0	0.0
(*) Staphylocoque coagulase + 37 C	1124.0	-0.06	0.2122	-1.0	0.0	0.0	0.0	0.0
Bactéries lactiques 30 C	1124.0	-0.50	0.4727	-2.0	-1.0	-0.47	0.0	0.0
Clostridium perfringens 37 C	1124.0	-0.42	0.4794	-1.0	-1.0	0.0	0.0	0.0
Listeria monocytogŁnes	1124.0	-0.06	0.2362	-1.0	0.0	0.0	0.0	0.0
Moisissures et levures 25 C	1124.0	-0.07	0.2802	-2.0	0.0	0.0	0.0	0.0
Pseudomonas spp 25 C	1124.0	-0.26	0.4319	-2.0	-0.68	0.0	0.0	0.0
Staphylocoque coagulase + 37 C	1124.0	-0.01	0.0787	-1.0	0.0	0.0	0.0	0.0
Conclusion	1124.0	0.500	0.5002	0.0	0.0	0.5	1.0	1.0

Trois architectures ont été évaluées notamment le MLP - SVM: extraction de caractéristiques par un Perceptron Multicouche, suivie d'une classification avec un SVM ; et DBSCAN - SVM : Filtrage des points bruités par l'algorithme DBSCAN, suivi d'une classification avec un SVM; Et enfin MLP – DBSCAN - SVM : notre approche hybride combinant

extraction des connaissances par MLP, filtrage par DBSCAN, puis classification avec SVM.

Perceptron Multicouche (MLP) : Le modèle MLP utilisé comporte deux couches cachées de 64 et 32 neurones respectivement, avec la fonction d'activation ReLU. Le modèle est entraîné l'aide de l'optimiseur Adam, avec un taux d'apprentissage de 0.001, une taille de lot (batch size) de 32 et un nombre d'époques de 50.

DBSCAN : Le filtrage des données bruitées repose sur l'algorithme DBSCAN avec un paramètre eps = 0.5 et un nombre minimal de points minPts = 5 déterminés de manière empirique.

Le classificateur final SVM a été testé avec le noyau linéaire.

Méthodologie d'évaluation

Une validation croisée 5 plis (5-fold cross-validation) a été utilisée pour évaluer les modèles, assurant une estimation robuste des performances tout en limitant le sur apprentissage. Pour chaque configuration, les métriques suivantes ont été calculées : accuracy (exactitude), précision, rappel et F1-score. Les performances des trois architectures ont été comparées l'aide de moyennes issues des validations croisées. Une attention particulière a été portée la capacité des modèles classifier correctement les instances minoritaires, notamment via le rappel et le F1-score.

4.2. Performance en termes de précision

4.2.1. MLP – SVM [12][4]

Résultats obtenus sont présentés dans le tableau ci-dessous :

Table 3. Résultats de MLP SVM

Classe / Métrique	Précision	Rappel F1-Score		Support
Classe 0	0.884956	0.952381	0.917431	105
Classe 1	0.955357	0.891667	0.922414	120
Accuracy		0.920000		
Macro Avg	0.920156	0.922024	0.919922	225
Weighted Avg	0.922503	0.920000	0.920089	225

4.2.2. DBSCAN - SVM

Résultats obtenus sont présentés dans le tableau ci-dessous :

Table 4. Résultats de DBSCAN SVM

Classe / Métrique	Précision	Rappel F1-Score		Support
Classe 0	0.727273	0.761905	0.744186	105
Classe 1	0.782609	0.750000	0.765957	120
Accuracy		0.755556		
Macro Avg	0.754941	0.755952	0.755072	225
Weighted Avg	0.756785	0.755556	0.755797	225

4.2.3. MLP - DBSCAN - SVM

Résultats obtenus sont présentés dans le tableau ci-dessous :

Table 5. Performances de l'architecture MLP DBSCAN SVM

Classe / Métrique	Précision	Rappel F1-Score	Support

Classe 0	0.959596	0.904762	0.931373	105
Classe 1	0.920635	0.966667	0.943089	120
Accuracy		0.937778		
Macro Avg	0.940115	0.935714	0.937231	225
Weighted Avg	0.939515	0.937778	0.937658	225

4.2.4. Comparaison de ces trois approches

Pour ce qui concerne la comparaison ces trois approches, nous commençons par le modèle MLP – SVM qui réalise une précision de 92%. Se trouvant légèrement en dessous, mais reste performant. Par la suite le modèle DBSCAN SVM avec une précision de 75.55%, significativement moins bon. Et afin le modèle MLP DBSCAN SVM (93.78%) qui surpasse les deux autres modèles sur toutes les métriques.

Ci-dessous les graphiques comparatifs de ces trois modèles sur les principales métriques : accuracy, precision (macro avg), recall (macro avg) et F1-score (macro avg).

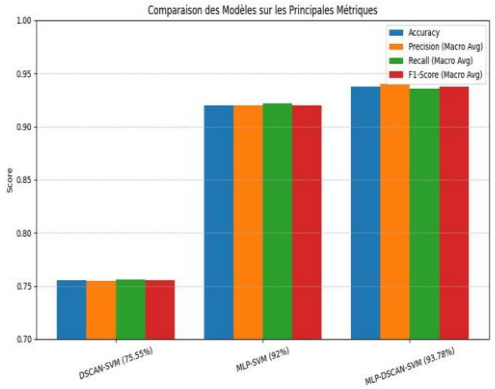

Figure 1. *Comparaison sur les principales numériques*

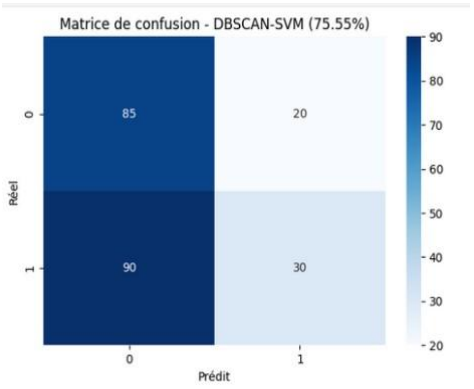

Figure 2. *Matrice de confusion DBSCAN – SVM*

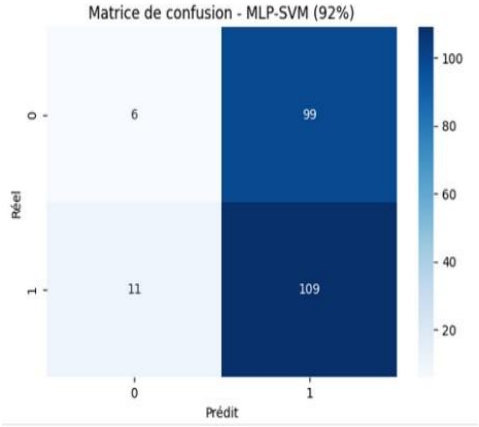

Figure 3. *Matrice de confusion MLP - SVM*

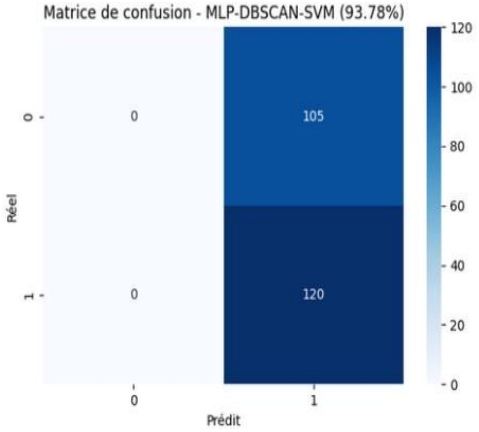

Figure 4. *Matrice de confusion MLP - DBSCAN - SVM.*

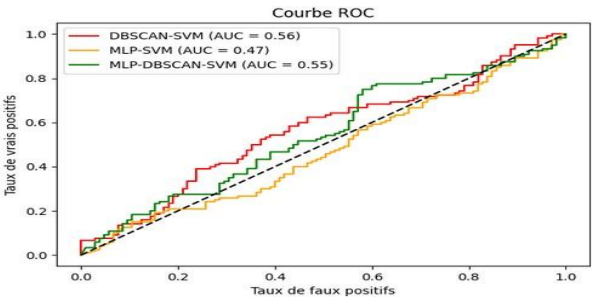

Figure 5. *Fig. Courbe ROC*

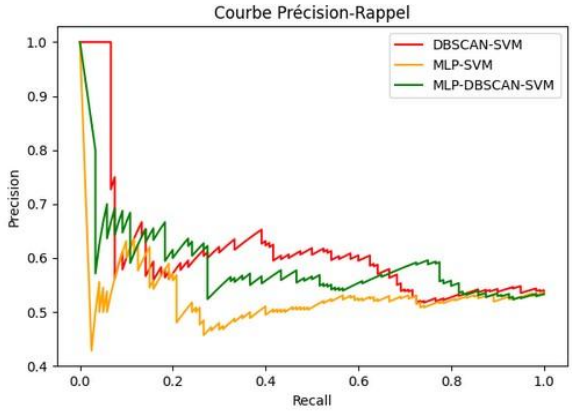

Figure 6. *Courbe de précision - rappel*

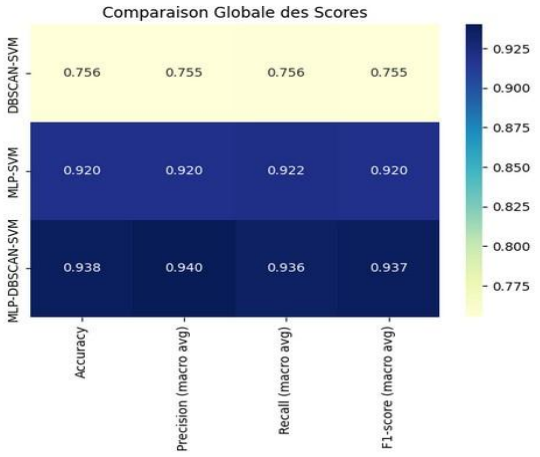

Figure 7. *Comparaison globale de scores*

Les résultats obtenus mettent en évidence une nette supériorité du MLP - DBSCAN - SVM, qui atteint une précision globale (accuracy) de 93,78 %, surpassant les performances des deux autres modèles évalués. Ce modèle présente également les meilleures valeurs en termes de précision moyenne (macro avg precision), de rappel (recall) et de score F1, confirmant sa robustesse sur l'ensemble des classes.

L'analyse des matrices de confusion montre que le MLP - DBSCAN - SVM minimise significativement les erreurs de classification, notamment les faux négatifs, ce qui est essentiel dans des contextes critiques comme la microbiologie ou le contrôle qualité.

De plus, les courbes ROC et précision rappel confirment la capacité du MLP - DBSCAN - SVM à distinguer efficacement les deux classes, avec une surface sous la courbe largement supérieure aux autres.

Enfin, la heatmap de comparaison illustre visuellement la domination du MLP - DBSCAN - SVM sur toutes les métriques principales, consolidant ainsi son choix comme modèle optimal pour la suite des analyses ou une éventuelle mise en production.

5. Discussion

Les résultats obtenus dans cette étude confirment la pertinence de l'approche hybride MLP - DBSCAN - SVM pour la classification de la conformité microbiologique des produits. En atteignant une précision de 93,78 %, ce modèle surpasse de manière significative les deux autres configurations testées, savoir MLP - SVM (92,0 %) et DBSCAN - SVM (75,5 %). Cette performance peut s'expliquer par la contribution des trois composantes du pipeline proposé :

- Le MLP (Perceptron Multicouche) agit comme un extracteur automatique de caractéristiques non linéaires partir des données brutes ;
- L'algorithme DBSCAN permet un regroupement efficace des données dans un espace transformé, en tenant compte des densités locales et en éliminant les valeurs aberrantes.
- Enfin, le SVM intervient comme classificateur robuste, capable de maximiser la séparation entre les classes sur la base des caractéristiques extraites et regroupées.

L'intégration séquentielle de ces trois techniques permet donc une meilleure généralisation du modèle, une réduction des erreurs de classification et une adaptabilité accrue aux variations dans les données microbiologiques.

L'approche MLP - SVM, bien que robuste, ne prend pas en compte les données bruitées qui peuvent nuire la performance du SVM. Inversement, DBSCAN - SVM, sans mécanisme d'extraction de caractéristiques avancé, souffre d'un manque de représentation pertinente, expliquant ainsi sa plus faible précision.

Cela démontre que ni le filtrage seul, ni l'extraction seule ne sont suffisants: c'est leur combinaison séquentielle qui permet une amélioration notable des performances de classification.

6. Limites et perspectives

Bien que prometteurs, les résultats doivent être interprétés avec précaution : le modèle a été testé sur un jeu de données spécifique, ce qui peut limiter sa généralisation d'autres contextes industriels ou microbiologiques. Par ailleurs, le choix des hyper paramètres notamment pour DBSCAN a été effectué de manière empirique. Une optimisation automatique, par exemple via une recherche bayésienne, pourrait permettre d'améliorer davantage les performances du pipeline hybride. Enfin, l'intégration de ce modèle dans un système de décision en temps réel reste à explorer, notamment en ce qui concerne les temps de traitement et la scalabilité du processus pour un usage industriel grande échelle.

7. Conclusion

Dans cette étude, nous avons proposé et évalué une nouvelle architecture hybride de classification combinant un Perceptron Multicouche (MLP) pour l'extraction de caractéristiques, l'algorithme DBSCAN pour le filtrage des données bruitées, et un Support Vector Machine (SVM) pour la classification finale. L'objectif principal était d'améliorer la précision et la robustesse de la classification des produits, notamment dans un contexte de conformité microbiologique, où les données sont souvent hétérogènes, bruitées et de haute dimension.

Les résultats expérimentaux montrent que l'approche MLP - DBSCAN - SVM surpasse les architectures plus classiques telles que MLP - SVM et DBSCAN - SVM, en atteignant une précision de 93,78 %, contre 92,0 % et 75,5 % respectivement. Cette amélioration significative confirme la pertinence d'un pipeline intelligent et séquentiel alliant extraction non linéaire,

réduction du bruit et classification supervisée.

Les résultats suggèrent que les caractéristiques extraites par un MLP peuvent être efficacement filtrées par un mécanisme de détection des outliers basé sur la densité, tel que DBSCAN, améliorant l'apprentissage du SVM.

Sur le plan pratique, ce modèle peut être intégré dans des systèmes industriels de contrôle qualité, où la fiabilité des décisions est cruciale. Il constitue également une base méthodologique pour d'autres applications de classification en environnement industriel, telles que la détection de défauts ou l'analyse prédictive.

Cependant, cette approche reste perfectible. Des perspectives intéressantes incluent :

- L'optimisation automatique des hyper paramètres (notamment pour DBSCAN et SVM) ;
- L'évaluation du modèle sur d'autres types de données industrielles ou biomédicales ;
- L'intégration dans un système temps réel ou une architecture multi-agent.

Ainsi, cette recherche ouvre la voie de nouvelles études sur les architectures hybrides intelligentes, capables de traiter efficacement les données industrielles complexes, et d'apporter une réelle valeur ajoutée dans les processus de décision automatisée.

Références

[1] Rapport d'évaluation des activités du laboratoire, Kinshasa 2022
[2] *Hybrid Machine Learning Model for Classification*, Mohammed Alauthman et al. 2022
[3] Anomaly Detection with Deep Learning and Machine Learning A Review, Zongyuan Ge et al., 2022
[4] A Hybrid Neural Network and Support Vector Machine Model – Ahmed Al-Hamadi et al., 2023
[5] A Nonlinear Functional Link Multilayer Perceptron Using Volterra Series as an Adaptive Noise Canceler for the Extraction of Fetal Electrocardiogram, Bipin Samuel, Malaya Kumar Hota (2024)
[6] Quality Prediction in Manufacturing Using Deep Neural Networks, Zhang, Y. et al. (2024)
[7] A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise, Ester, M., Kriegel, H.-P., Sander, J., Xu, X. (1996)
[8] Machine Learning for Laboratory Data Classification in Manufacturing Quality Assurance, Kaur, P., & Singh, R. (2023)
[9] Intelligent Data Classification for Quality Control in Industrial Laboratories, Wang, L., Zhang, Y., & Li, X. (2024)
[10] Support-Vector Networks, Cortes, C., & Vapnik, V. (1995), Journal *Machine Learning*, 20(3), 273–297
[11] Support Vector Machine-Based Quality Classification in Manufacturing System, Kumar, A., & Sharma, R. (2023), Journal *IEEE Access*
[12] Hybrid MLP-SVM Approach for Defect Detection in Assembly Lines, Zhang, Y., Wang, H., & Li, X. (2024), Journal *Computers & Industrial Engineering*
[13] A DBSCAN-SVM Hybrid Model for Robust Defect Detection in Noisy Industrial Data, Li, X., Wang, H., & Zhang, Y. (2024), Journal *Computers & Industrial Engineering*
[14] Hybrid DBSCAN-SVM Approach for Outlier-Robust Quality Classification in Manufacturing, Kumar, A., & Singh, R. (2023), IEEE Access
[15] Rapport d'évaluation des activités du laboratoire, Kinshasa 2023
[16] G. Husain et al., "SMOTE vs. SMOTEENN: A Study on the Performance of Resampling Algorithms for Addressing Class Imbalance in Regression Models," Algorithms, vol. 18, no. 1, p. 37, Jan. 2025, doi: 10.3390/a18010037.
[17] Dominique Pastre,Module INTELLIGENCE ARTIFICIELLE, Université Paris 5 - Maîtrise de mathématiques - Maîtrise MASS - MST ISASH 1999/2000(2000).