INDABAX
Abidjan 2023

3rd
EDITION

INDABAX 2023 Côte d'Ivoire Proceedings

3rd Conference of Intelligence Artificial on Artificial Intelligence and Smart City

October 10-11 2023

Université Félix Houphouët-Boigny

Cocody Abidjan, Côte d'Ivoire

Comité d'édition
Beman Hamidja KAMAGATE
Mamadou DIARRA

SOMMAIRE

1. Avant propos

Le Conférence INDABAX, a débuté en 2018 en tant programme visant à renforcer les échanges d'expériences et d'expertises entre les chercheurs autour des thématiques de l'intelligence artificielle, de l'apprentissage machine et l'apprentissage profond. INDABAX vise à permettre à davantage de personnes de contribuer à l'avancée de l'intelligence artificielle sur notre continent. A ce jour, 25 pays organisent des évènements INDABAX chaque année.

La présente édition constitue la troisième édition que le comité INDABAX Côte organise en collaboration avec l'Unité de Formation de Recherche de Mathématiques Informatique (UFR MI) de l'Université Félix Houphouët Boigny (UFHB) et du Centre National de Calcul de Côte d'Ivoire (CNCCI).

Le programme de cette édition, se décline en 10 communications scientifiques, sélectionnées parmi 14 soumises et 01 conférences présentées par des spécialistes de renommée internationale. Cette édition a aussi vu la réalisation d'une séance de formation sur le *Deep Learning* à l'endroit des étudiants et les enseignants-chercheurs. Il faut aussi noter la présence effective des entreprise CIE et SAH Analytics qui en plus des communications sur des cas d'utilisation en entreprise ont contribué à l'animation d'un panel d'expert sur le thème : « **Développement des Smart Cites en Côte d'Ivoire : Enjeux et Défis »**

2. Participants et partenaires

INDABAX Abidjan 2023 — 3rd EDITION	LAMI — Laboratoire Mécanique et informatiq
LARIT — LABORATOIRE DE RECHERCHE EN INFORMATIQUE ET TELECOMMUNI	Lastic — Laboratoire des Sciences et Technologies de l'Information et de la Communication
INPHB — Institut National Polytechniqu Félix HOUPHOUËT-BOIGNY	CIE — Compagnie Ivoirienne d'Electricité
SaH Analytics	Institut Pasteur de Côte d'Ivoire
cncci — Centre National de Calcul de Côte d'Ivoire	Université Félix HOUPHOUËT-BOIGNY

3. Comité d'organisation

Dr. Amadou DIABAGATE, Université Félix Houphouët-Boigny (Côte d'Ivoire)
Dr. Ismael KONE, Université Virtuelle de Côte d'ivoire (Côte d'ivoire)
Dr. Louis Paul SEKA, Université Félix Houphouët-Boigny (Côte d'Ivoire)
Dr. Mamadou DIARRA, Université Félix Houphouët-Boigny (Côte d'Ivoire)
Dr. Abdoul MAIGA, Université Félix Houphouët-Boigny (Côte d'Ivoire)
Dr. Ida Brou ASSIE, Université Félix Houphouët-Boigny (Côte d'Ivoire)
Dr. Mohamed DIALLO, Université Félix Houphouët-Boigny (Côte d'Ivoire)
Dr. Issa Traoré, Institut de Recherche en Mathématiques (Côte d'Ivoire)
Dr. Beman Hamidja KAMAGATE, Ecole Supérieure Africaine des TIC (Côte d'Ivoire)
M. Medard KOUASSI, Université Félix Houphouët-Boigny (Côte d'Ivoire)
Dr. Ngolo KONATE, Université Félix Houphouët-Boigny (Côte d'Ivoire)
M. Stanislas ASSOHOUN, Université Jean Lorougnon Guédé (Côte d'Ivoire)
M. Patrice BROU, Université Félix Houphouët-Boigny (Côte d'Ivoire)
M. Ibrahim BAKAYOKO, Université Félix Houphouët-Boigny (Côte d'Ivoire)
Dr. Bertin KAYE BI, Université Félix Houphouët-Boigny (Côte d'Ivoire)

4. Comité scientifique

Prof. Assohoun ADJE, Université Félix Houphouët-Boigny (Côte d'Ivoire)

Prof. ADOU KABLAN, Université Félix Houphouët-Boigny (Côte d'Ivoire)

Prof. Adama COULIBALY, Université Félix Houphouët-Boigny (Côte d'Ivoire)

Prof. Michel BABRI, Institut National Polytechnique H.B (Côte d'ivoire)

Prof. Albert BIFET, Telecom Paris (France)

Prof. Vincent MONSAN, Université Félix Houphouët-Boigny (Côte d'Ivoire)

Prof. Selam WAKTOLA, Netherlands Cancer Institute (Pays-Bas)

Prof. Bernard DOUSSET, Université Paul Sabatier de Toulouse (France)

Prof. Souleymane OUMTANAGA, Institut National Polytechnique H.B (Côte d'ivoire)

Prof. Abdelghani CHIBANI, Université Paris-Est Créteil (France)

5. Listes des communications

1. Architecture des Villes Intelligentes
2. Adaptive Lightning : optimisation de l'éclairage publique
3. Identification des personnes susceptibles d'avoir la covid-19 par les techniques de l'apprentissage automatique
4. Chat with your Data
5. Modèle d'optimisation multi-objectif prenant en compte le rayon de courbure des virages et l'inclinaison des pentes des routes
6. An Agent Based Model to Improve Traffic Flow and Assess the Impact of Smart Traffic Ligh
7. Sécurité et Smart city : optimisation de la traçabilité des véhicules volés
8. Construction d'un modèle de gestion intelligente de la consommation d'électricité
9. Nouvelle approche de sélection des valeurs optimales d'hyperparamètre
10. Using Text to Mesure Bias in Research

6. Liste des conférences invitées

1. Endoscopic images and clinical data to assess the response of radiation therapy

7. Programme de la conférence

Mardi 10 octobre 2023		
HORAIRE	**ACTIVITES**	**AUTEURS**
8h00–08h30	**Réception des invités**	Comité d'Organisation
8h30–08h50	**Maître de cérémonie :** *Mamadou DIARRA* *Presente les personnalités presentes et le deroulés de la céremonie* **Cérémonie d'Ouverture** • **Discours du Directeur du Centre National de Calcul de Côte d'ivoire** *Dr Traoré Issa :* *Remerciement du Prof Arouna, représentant du Ministre* *Remerciement du Adou UFR Math* *Remerciement e Coulibaly Adama, IRMA* *Remerciement de Brou , INPHB* *Remerciement du ministère de la construction , de la CIE* *Le CNCCI est heureux pour l'accueil de cette conférence sur le thème de villes intelligentes et IA.* • **Discours du Président du Comité d'Organisation** **Dr Diabagaté Amadou :** Presentation des civilités aux officiels **-Conférence qui fait la promotion de l'intelligence artificielle** **- possible grâce aux puissance de calcul** **- creation de l'initiatif INDABAX pour ne pas que l'afrique rate le train des 4IR** **- LAMI , IDTS, CNCCI : ont le mis en place la conference** **- theme de la conférence : Smart city et intelligence artificielle** • **Discours du Directeur de l'UFR Mathématiques et Informatique** **Prof. Adou Kablan** - Civilité	

	- Présentation du contexte des 4IR pour la transformation digitale, un processus irréversible. Notre pays doit s'y inscrire résolument. - C'est pourquoi le LAMI, l'IDTS organise cette conférence, qui est à sa 3ème edition - Heureux que la conférence passe à deux jours ; - Objectif : faire de l'UFR MI un pole d'excellence en IA, promouvoir la formation et la recherche en IA - Remerciement de Prof Arouna pour le rôle qui joue dans la promotion du CNCCI et son activisme pour reveiller le pôle de simulation - Remerciements des partenaires **Représentant du ministre** **Prof. Arouna** **-civilité** -heureux d'être present, et le thème est tres pertinent, (les atouts de smart city et de l'intelligence artificielle) -problème des villes(urbanisation , forte croissance , - préoccupation legitime lié à la vie privé et la protection de la vie • **-** l'intelligence artificiel ne doit pas créer l'inégalité • - la société civil, les gouvernements, les chercheurs pour mettre à disposition l'IA • - le ministère doit travailler pour avoir une stratégie en IA. • **- il faut arriver à préserver notre culture à travers l'IA et les smart cities** • **- remerciements des membres du comité d'organisation.**	
08h50–08h55	**Présentation du thème**	Président du Comité Scientifique
9h00–09h25	**Construction d'un modèle de gestion intelligente de la consommation d'électricité**	Dr Yazid Hambally Yacouba, Université Félix Houphouët-Boigny (UFHB)
09h25–10h10	**Cocktail**	Comité d'Organisation
10h10–10h35	**Communication** **Adaptive Lightning : optimisation de l'éclairage publique**	M. Didier Michael KRE, Compagnie Ivoirienne d'Electricité (CIE)

10h35–12h25	**Série de Communications scientifiques**	
	Identification des personnes susceptibles d'avoir la covid-19 par les techniques de l'apprentissage automatique	Dr Diako Doffou Jérôme, Ecole Supérieure Africaine des TIC (ESATIC)
	Chat with your Data	M. Michel TOFFE, Data 365
	Modèle d'optimisation multi-objectif prenant en compte le rayon de courbure des virages et l'inclinaison des pentes des routes	Dr Coulibaly Kpinna Tiekoura, Ecole Supérieure Africaine des TIC (ESATIC)
	An Agent Based Model to Improve Traffic Flow and Assess the Impact of Smart Traffic Ligh	Dr N'Golo Konaté, Université Félix Houphouët-Boigny (UFHB)
	Sécurité et Smart city : optimisation de la traçabilité des véhicules volés	M. Tiorna COULIBALY, Compagnie Ivoirienne d'Electricité (CIE)
12h30-13h30	**Déjeuner**	Comité d'Organisation
14h00–15h20	**Série de Communications scientifiques**	
	Communication **Architecture des Villes Intelligentes**	Dr Mamadou Diarra, Université Félix Houphouët-Boigny (UFHB)
	Nouvelle approche de sélection des valeurs optimales d'hyperparamètre	Dr Koppoin Charlemagne, Ecole Supérieure Africaine des TIC (ESATIC)
	Using Text to Mesure Bias in Research	Dr Marlène KOFFI, Université de Toronto (CANADA)
	Conférence **Endoscopic images and clinical data to assess the response of radiation therapy**	Dr Selam WAKTOLA, Université Johns-Hopkins (Etats-Unis)

15h25–16h00	**Pause-Café**	Comité d'Organisation
16h00–17h00	**Panel** **Développement des Smart Cites en Côte d'Ivoire : Enjeux et Défis**	• Dr Jules KALA (Compagnie Ivoirienne d'Electricité) • Dr Beman KAMAGATE (ESATIC) • Mme Sali OUATTARA (SAH ANALYTICS) • Dr Philippe PANGO, Conseiller Spécial, Ministère de la Construction, du Logement et de l'Urbanisme
17h00	**Clôture**	Comité d'Organisation
Mercredi 11 Octobre 2023		
8h30-09h00	**Ouverture**	Comité d'Organisation
09h00-12h00	**Formation (Deep Learning) - Partie 1**	Dr Ismael KONE, Université Virtuelle de Côte d'Ivoire
12h00-13h00	**Déjeuner**	Comité d'Organisation
13h30–16h30	**Formation (Deep Learning) - Partie 2**	Dr Ismael KONE, Université Virtuelle de Côte d'Ivoire
16h30-17h15	**Clôture**	Comité d'Organisation

8. Communications

8.1. Construction d'un modèle de gestion intelligente de la consommation d'électricité

Yazid Hambally Yacouba[1], Adama Coulibaly[2]

[1]Université de Bondoukou, UBKOU, Côte d'Ivoire
[2]Félix Houphouet Boigny University UPHB, Côte d'Ivoire
Email de l'auteur correspondant : yazid.hambally@gmail.com

Résumé. Le volume important des données de consommation d'électricité milite à l'agrégation de ces données. La mise en œuvre de méthodes d'agrégation constitue à cet effet une préoccupation majeure à laquelle une réponse est apportée en présentant un cas d'agrégation des données de consommation d'électricité à l'aide du processus de saut. Un jeu de données a permis de faire des simulations et de présenter les résultats obtenus pour les agrégations journalières, mensuelles et annuelles. Il est mis en exergue le principe d'utilisation du processus de saut pour l'agréation de ces données. Ce travail est une présentation concrète d'une simulation pour l'agrégation des données de consommation d'électricité dans un réseau de capteurs sans fil que peut constituer un réseau de compteurs intelligents. L'approche de ce travail consiste à utiliser des méthodes d'agrégation pour diminuer le flux d'échanges des données dans des réseaux de capteurs sans fil.

Mots clés: simulation, agrégation des données, comptage intelligent, consommation d'électricité, processus de saut

1. Introduction

L'agrégation des données des compteurs intelligents est une problématique importante dans la gestion des données volumineuses de la consommation d'électricité. L'agrégation des données de consommation d'électricité dans les réseaux intelligents est essentielle pour les échanges des données et pour le traitement de ces données dans le cadre de la réduction du coût du stockage et de la charge d'énergie nécessaire au fonctionnement d'un réseau de compteurs intelligents alimenté par des batteries. Rohit Gupta and Krishna Teerth Chaturvedi [1] expliquent le rôle important des réseaux intelligent pour la gestion de l'électricité et des données entre les fournisseurs et les consommateurs. Cette publication procède aussi à une comparaison de méthodes de traitement des données.
Les réseaux de compteurs intelligents peuvent dans certaines zones isolées être assimilés à des réseaux de capteurs sans fil qui sont très souvent de petites tailles et qui ont une capacité de traitement limitée avec une alimentation basée sur des batteries généralement de faibles autonomies. Cette contrainte d'alimentation expose ces réseaux de capteurs à de nombreuses pannes. Aussi, l'agrégation des données peut contribuer à l'amélioration de la résilience aux pannes de ces réseaux intelligents. De même, les besoins d'atténuation des contraintes de bande passante, d'énergie et de débit souffrent de la capacité d'adaptation des réseaux sans fils aux topologies des réseaux informatiques dont le trafic des données est à la fois dynamique et

imprévisible. Zhiyi Chen et al. [2] présentent une revue des défis, des opportunités et applications relatifs à l'utilisation des compteurs intelligents ainsi que l'apport de ces compteurs dans les réseaux intelligents. S. R. Selva Jeevitha [3] évoque le besoin de stockage des signaux de puissance dans les réseaux électriques intelligents et propose une méthode de compression du signal haute fréquence et de reconstruction du signal original à l'aide de la transformée de Hilbert.
L'objectif de ce travail consiste à présenter des méthodes d'agrégation qui facilitent le traitement des données d'électricité à différentes échelles d'analyse tout en améliorant la durée d'autonomie des batteries. Soham Dutta et al. [4] exposent la nécessité de faire face de façon permanente à l'apparition de pannes imprévisibles et dangereuses dans les réseaux de distribution utilisant des compteurs intelligents. Shampa Banik et al. [5] donne un apercu de l'apport des réseaux intelligents dans la détection des anomalies. Un tel processus de gestion des données des compteurs intelligents nécessite le stockage et l'agrégation des données au niveau de certains compteurs intelligents dont les critères de sélection dans un bâtiment ou un quartier ne sont pas abordés dans ce travail. L'apport de ce système réside dans la facilitation de l'interprétation des données et la mise en œuvre d'une bonne utilisation de l'énergie.

2. Etat de l'art

Il existe une pluralité de techniques d'agrégation des données qui visent à répondre aux défis de traitement des données et de réduction de la consommation d'énergie dans les systèmes et les réseaux informatiques. Cette section présente un état de l'art de méthodes d'agrégation ainsi que des solutions apportées.
Les techniques d'agrégation des données sont mises en œuvre pour plusieurs raisons parmi lesquelles :
La compression des données :
Lulu Wen et al [6] présentent une étude exhaustive des techniques et des méthodes potentielles de compression des entrepôts de données issues des compteurs intelligents. Pour améliorer les capacités de stockage et de communication. Lee J et al. [7] proposent un modèle de compression basé sur le choix de modèles existants pour déterminer le modèle qui améliore la qualité de la reconstruction tout en maintenant le taux de compression grâce à l'analyse de la variation dans le temps des propriétés spectrales des données.
La performance de la collecte des données
Dans un souci de transférer les données des compteurs intelligents vers le centre de contrôle ; Sung Tien-Wen et al. [8] proposent un schéma de placement de points d'agrégation de données (PAD) et présentent les algorithmes correspondants pour réduire le nombre de PAD et limiter l'impact sur la qualité de la communication. Ramesh Rajagopalan and Pramod K. Varshney [9] s'intéressent aux problèmes d'agrégation de données dans les réseaux de capteurs à énergie limitée et proposent différents algorithmes dans le but de collecter et d'agréger les données d'une manière économe sur la base des mesures de performance telles que la durée de vie, la latence et la précision des données. Geetika Dhand et S.S. Tyagi [10] expliquent l'importance de la collecte de données et comparent différentes approches de clustering hiérarchique.

Mohamed Saleem Haja Nazmudeen et al. [11] proposent une architecture de gestion des communications en vue de faciliter la collecte, le stockage et le traitement des données des compteurs intelligents. Jagdish Chandra Pandey and Mala Kalra [12] montrent l'intérêt de la compression et du cryptage des données dans les réseaux intelligents pour communiquer de manière sécurisée des grandes quantités de données tout en réduisant la consommation de mémoire et du temps d'exécution.

L'amélioration de la performance du réseau

Cette publication [13] explore les algorithmes d'agrégation de données sur la base de la topologie du réseau ainsi que les compromis possibles au sein ces algorithmes. Les travaux de Hassan Harb et al. [14] consistent à une agrégation locale des données au niveau d'un nœud capteur pour la classification périodique basée sur le clustering tout en permettant au cluster de tête d'éliminer les jeux de données redondants générés par les nœuds voisins grâce à l'utilisation de trois méthodes d'agrégation de données méthodes. Ces méthodes sont basées respectivement sur les fonctions de similarité des ensembles, le modèle Anova unidirectionnel avec tests statistiques et les fonctions de distance. Karthikeyan Vaidyanathan et al. [15] présentent une simulation mettant en exergue les réductions d'énergie et de délai de transmission de bout en bout des données. Ils proposent également une agrégation hybride tout en permettent aux nœuds de capteurs de passer d'une technique d'agrégation à l'autre en fonction de la charge du réseau. Mohammad Ghiasi et al. [16] proposent des modèles mathématiques et une simulation dans le but de contribuer à un développement énergétique durable à faible émission de carbone en utilisant les technologies de l'Internet des objets (IoT), de l'Internet de l'énergie (IoE) et les systèmes intelligents. L'étude réalisée par Adrian Lang et al. [17] met en exergue l'importance de l'emplacement des points d'agrégation de données dans l'amélioration de la rentabilité et de la qualité de service dans les réseaux de compteurs intelligents.

La sécurisation des échanges de données

Dans l'étude [18], Abbasian Dehkordi et al. font une revue des techniques et des protocoles d'agrégation des données dans le but de présenter de nouvelles approches tout en décrivant leurs avantages et leurs inconvénients. L'article [19] met en exergue le déploiement de plusieurs nœuds IoT dans le domaine de la sécurité et le besoin de faire un compromis entre la consommation d'énergie et la fiabilité en éliminant la redondance des données jusqu'à un certain seuil. Guguloth Ravi et al. [19] propose un schéma d'agrégation fiable de données (CRDA : cluster based reliable data aggregation) pour l'économie d'énergie et qui est appliqué à un cluster de capteurs IoT en charge de la collecte et de l'agrégation des donnés. Les auteurs mettent ensuite en œuvre un algorithme d'agrégation des données qui calcule le degré de confiance de chaque capteur IoT pour déterminer le cluster de tête (CH : cluster head). Un réseau neuronal (ROL-DNN : reformative optimal–learning-based deep neural network) est aussi utilisé pour le calcul des itinéraires entre les capteurs IoT pour garantir la fiabilité du transport des données. Ashutosh Kumar Singh and Jatinder Kumar [20] proposent un modèle

sécurisé et préservant la confidentialité de l'agrégation et de la classification des données dans une architecture brouillard (fog) et nuage (cloud). Ils présentent également un système d'agrégation de données multidimensionnelles qui préserve la confidentialité grâce à un traitement des requêtes basé sur la gestion de l'identité et du temps [21].

2. Problématique

L'énergie est l'un des plus grands défis et aussi l'un des principaux piliers de la croissance économique. Les évolutions technologiques dans le domaine de la communication et des échanges de données constituent une véritable opportunité pour le développement du secteur de l'électricité. De même, plusieurs facteurs rappellent la nécessité de mise en œuvre de nouvelles stratégies parmi lesquels la disparité dans l'avancement de l'électrification entre les zones rurales et les zones urbaines, l'insuffisance des moyens financiers, l'augmentation de la demande, l'isolement et l'éloignement de certaines zones d'habitation. Dramé et Cheikh [22] montrent que la résolution des problèmes d'accès à l'électricité en Afrique de l'Ouest passe par le déploiement des réseaux intelligents avec la Côte d'Ivoire comme étude de cas. En effet, cette étude recommande la réalisation des infrastructures électriques opportunes au développement des réseaux intelligents telles que les technologies d'énergie renouvelable, de communication par ligne électrique, de tarification basée sur les compteurs intelligents et de prépaiement à l'aide des téléphones mobiles et d'autres équipements informatiques. Sebastian Sterl et al. [23] expliquent que la fiabilité de l'approvisionnement en électricité dépend de l'utilisation complémentaire de l'hydroélectricité, des énergies solaire et éolienne contrairement à la tendance basée sur les énergies fossiles. S. M. Kadri et al. [24] mettent en exergue la pauvreté et la faiblesse technologique comme obstacles de l'électrification en Afrique de l'Ouest et présentent un état des approches en cours pour la production distribuée d'électricité. Osama Majeed et al. [25] donnent un aperçu des réseaux intelligents et de leurs fonctionnalités tout en expliquant l'apport de ces technologies dans l'évolution et le renforcement du système de distribution d'électricité. Young et Jacob R. [26] présentent les réseaux intelligents et font un rapport de l'état de mise en œuvre dans les pays en voie de développement. Aussi, Fernando Antonanzas-Torres et al. [27] font l'état des lieux des installations de recherche et d'exploitation d'électricité en Afrique de l'Ouest et énumèrent les défis de développement des mini-réseaux particulièrement les systèmes solaires domestiques.

3. Principe fondamental des méthodes d'agrégation des données de notre système d'électricité à l'aide du processus de saut

Dans notre publication [28], nous avons déjà posé les bases de notre méthode d'agrégation des données à l'aide du processus de saut.

Les agrégations de données sont modélisées en utilisant le processus de saut comme suit :

$\boldsymbol{n}$: le nombre d'enregistrement de données

$\boldsymbol{X_n}$: la consommation d'électricité entre T_{n-1} et T_n à l'instant t

$\mathbb{1}_{\boldsymbol{A}}(\boldsymbol{t}) = \begin{cases} \mathbf{1}\ \boldsymbol{si}\ \boldsymbol{t} \in \boldsymbol{A} = [\boldsymbol{T_{n-1}}, \boldsymbol{T_n}[\\ \mathbf{0}\ \boldsymbol{sinon} \end{cases}$, $\mathbb{1}$ est une fonction binaire qui permet de constater un

changement d'état dans le système

$\boldsymbol{M}$: le nombre fini d'évènements au bout duquel le processus de comptage se termine

La consommation d'électricité d'un compteur intelligent à un instant t est défini par :

$\boldsymbol{Z_t = \sum_{n=1}^{M} X_n \mathbb{1}_{[T_{n-1},T_n[}(t)}$

Le cumul de la consommation d'électricité d'un compteur intelligent jusqu'à l'instant t est :

$$\boldsymbol{Y_t = \sum_{n=1}^{M} X_n \mathbb{1}_{T_n<t}}$$

$\boldsymbol{Y_t}$ est un processus de saut, les valeurs $\boldsymbol{X_n}$ du système de comptage intelligent étant différentes.

Ci-dessous le tableau récapitulatif des agrégations des données de consommation d'électricité :

Tableau 1. Tableau récapitulatif des agrégations des données de consommation d'électricité

Consommation totale d'électricité à l'instant t		**Consommation totale d'électricité jusqu'à l'instant t**	
$\boldsymbol{S_N = \sum_{k=1}^{N} Z_t^k}$		$\boldsymbol{R_N = \sum_{k=1}^{N} Y_t^k}$	
Temps t	$\boldsymbol{Z_t^k = \sum_{n=1}^{M} X_n^k \mathbb{1}_{[T_{n-1}^k,T_n^k[}(t)}$	Temps t	$\boldsymbol{Y_t^k = \sum_{n=1}^{M} X_n^k \mathbb{1}_{T_n^k<t}}$
Zone i	$\boldsymbol{Z_t^k(i) = \sum_{n=1}^{M} X_n^k(i) \mathbb{1}_{[T_{n-1}^k,T_n^k[}(t)}$	Zone i	$\boldsymbol{Y_t^k(i) = \sum_{n=1}^{M} X_n^k(i) \mathbb{1}_{T_n^k<t}}$
Usage j	$\boldsymbol{Z_t^k(j) = \sum_{n=1}^{M} X_n^k(j) \mathbb{1}_{[T_{n-1}^k,T_n^k[}(t)}$	Usage j	$\boldsymbol{Y_t^k(j) = \sum_{n=1}^{M} X_n^k(j) \mathbb{1}_{T_n^k<t}}$
Zone i et Usage j	$\boldsymbol{Z_t^k(i,j) = \sum_{n=1}^{M} X_n^k(i,j) \mathbb{1}_{[T_{n-1}^k,T_n^k[}(t)}$	Zone i et Usage j	$\boldsymbol{Y_t^k(i,j) = \sum_{n=1}^{M} X_n^k(i,j) \mathbb{1}_{T_n^k<t}}$

Simulation sur les agrégations journalières, mensuelles et annuelles

a. Contexte de réalisation des tests de simulation

Les données de simulation utilisées sont des données d'électricité de 370 clients au Portugal sur la période 2011 à 2014 [29]. L'ensemble de données ne contient aucune valeur manquante. Les valeurs sont en kW au pas de 15 min. Pour convertir les valeurs en kWh, les valeurs doivent être divisées par 4.

Chaque colonne représente un client. Certains clients ont été créés après 2011. Dans ces cas, la consommation était considérée comme nulle.
Toutes les périodes de temps se rapportent à l'heure portugaise. Cependant tous les jours présentent 96 mesures (24*15). Chaque année, au mois de mars, jour de changement d'heure (qui ne compte que 23 heures), les valeurs entre 1h00 et 2h00 sont nulles pour tous les points. Chaque année, en octobre, jour de changement d'heure (qui compte 25 heures), les valeurs entre 1h00 et 2h00 cumulent la consommation de deux heures.
Dans le cadre de la mise en œuvre des tests de simulation, une application a été développée dans la technologie Java. La solution proposée récupère les données de simulation [18] classées par ordre croissant des dates d'arrivé et à des pas de temps de 15 minutes. Aussi, chaque ligne du fichier des données de simulation présente individuellement l'ensemble des données de consommation des 370 clients. L'application met en œuvre des programmes de calcul et d'affichage du résultat des agrégations journalières, mensuelles et annuelles des consommations des clients.

b. Simulation sur les agrégations des données au pas de temps journalier

Ci-dessous les résultats de l'agrégation journalière des données pour les clients MT_001, MT_210, MT_369, MT_370 :

Tableau 2. Extraction des résultats des agrégations journalières

Date	MT_001	MT_210	MT_369	MT_370
11/02/2012	258.883	78294.573	73447.214	0.0
12/02/2012	239.847	73067.183	67326.246	0.0
13/02/2012	314.720	73312.661	77876.832	0.0
14/02/2012	953.045	74255.813	81609.970	0.0
15/02/2012	1507.614	72359.173	82350.439	0.0
16/02/2012	1515.228	71214.470	81455.278	0.0
17/02/2012	1512.690	75111.111	81635.630	0.0
18/02/2012	1365.482	75307.493	74744.868	0.0
19/02/2012	1441.624	71772.609	70006.598	0.0
20/02/2012	1378.172	76356.589	78653.225	0.0
21/02/2012	1506.345	72720.930	77706.011	0.0
22/02/2012	1425.126	73832.041	82166.422	0.0
23/02/2012	1380.710	73170.542	82224.340	0.0
24/02/2012	1439.086	74583.979	82342.375	0.0
25/02/2012	1661.167	76865.633	76005.131	0.0

Simulation sur les agrégations des données au pas de temps mensuel

Ci-dessous les résultats de l'agrégation mensuelle des données pour les clients MT_001, MT_012, MT_120, MT_370 :

Tableau 3. Extraction des résultats des agrégations mensuelles

Date	MT_001	MT_012	MT_120	MT_370
11/2013	46880.710	430848.936	0.0	4.5739286486E7
12/2013	8558.375	510874.468	0.0	4.4630864864E7
01/2014	7220.812	542934.042	61001.047	3.9909351351E7
02/2014	6106.598	501002.127	99151.911	3.7458540540E7
03/2014	7111.675	435482.978	108907.805	4.5485783783E7
04/2014	8361.675	409714.893	107146.673	5.1986270270E7
05/2014	14086.294	386185.106	112424.305	5.6521513513E7
06/2014	6431.472	375808.510	115553.693	5.8213675675E7
07/2014	6771.573	396870.212	142814.562	6.2059459459E7

Simulation sur les agrégations des données au pas de temps annuel

Ci-dessous les résultats de l'agrégation annuelle des données pour les clients MT_001, MT_021, MT_221, MT_370 :

Tableau 4. Extraction des résultats des agrégations annuelles

Date	MT_001	MT_021	MT_221	MT_370
2011	0.0	0.0	5972502.342	0.0
2012	193131.979	5786534.031	5539887.558	0.0
2013	221593.908	5731217.277	5364469.547	6.02657670269E8
2014	142197.969	5559989.528	5380239.458	6.20697837837E8
2015	2.538	185.863	71.837	7135.135

Critères de mesure de la performance des méthodes d'agrégation proposées

Nous avons déjà montré [28] que le comportement de la consommation d'électricité est également dépendant des caractéristiques des appareils électriques branchés aux compteurs intelligents.

Soit T_l^n la durée de consommation de l'appareil l dans l'intervalle $[T_{n-1}, T_n[$ et soit C_l la consommation par unité de temps de cet appareil.

La consommation de l'appareil l dans l'intervalle $[T_{n-1}, T_n[$ peut alors être représentée par

$C_l^n = C_l \times T_l^n$ (1)

La consommation totale d'électricité du compteur intelligent k pour M appareils dans l'intervalle $[T_{n-1}, T_n[$ est :

$X_n^k = \sum_{l=1}^{M} C_l^{n,k} = \sum_{l=1}^{M} C_l^k \times T_l^n$ (2)

avec $C_l^{n,k}$ la consommation lue par le compteur k pour l'appareil l dans l'intervalle $[T_{n-1}, T_n[$

et C_l^k la consommation par unité de temps de l'appareil l relativement aux caractéristiques du compteur intelligent k .
Nous déduisons alors que les caractéristiques des appareils électriques influent sur la consommation d'électricité et donc sur le volume de données à agréger.
Nous avons par ailleurs présenté les formules de calcul des agrégations des données de notre système de gestion de la consommation d'électricité et nous pouvons déduire que ces agrégations dépendent des critères suivants :

- $\boldsymbol{n}$: le nombre d'enregistrement de données ;
- $\boldsymbol{X_n}$: la consommation d'électricité entre T_{n-1} et T_n à l'instant t ;
- $\boldsymbol{M}$: le nombre fini d'évènements au bout duquel le processus de comptage se termine.

Nous disons donc que la performance des méthodes d'agrégation des données de notre système d'électricité prend en compte plusieurs facteurs et il convient donc de faire usage de nos méthodes d'agrégation en fonction de l'importance des critères de performance susmentionnés.

4. Conclusion

Plusieurs initiatives sont en cours dans le monde pour une gestion de l'énergie basée sur le comptage intelligent. Une réalité à laquelle le continent africain sera confrontée après le développement des infrastructures énergétiques et l'intégration des sources alternatives d'énergie du nombre desquelles le solaire et l'éolien.
Le potentiel solaire de l'Afrique est une source de motivation pour l'utilisation des compteurs intelligents et le développement d'un système d'information qui permettra la gestion à distance de toutes les fonctions du compteur évolué. Le défi majeur repose cependant sur la mise en œuvre d'un système d'information capable de répondre aux objectifs attendus surtout dans un contexte où la réduction des coûts constitue une problématique parfois vitale. En effet, des investissements importants sont nécessaires pour la réalisation des centres décentralisés de production de l'électricité. L'implantation de ces centres doit également prendre en compte le rapprochement avec les zones de consommation pour réduire les pertes techniques. Ce travail a permis de définir un cadre d'agrégation des données des compteurs intelligents. La spécificité des données dans un tel système a conduit à la mise en place de méthodes d'agrégation de données pour faciliter le traitement et l'analyse de ces données.
L'état de l'art a permis de présenter d'autres méthodes d'agrégation des données et plus précisément celles en lien avec les données énergétiques.
L'originalité de ce travail réside dans la présentation d'une simulation relative aux traitements des données des compteurs intelligents. Il s'agit de réaliser une simulation sur des données issues de compteurs intelligents par le biais de l'agrégation de ces données à l'aide du processus de saut. Le manque d'informations sur les systèmes de comptage intelligent rend toute étude comparative difficile. En effet, la mise en œuvre de ces systèmes reste fermée même si certaines fonctionnalités sont bien connues des utilisateurs.

La complexité de la gestion des données des systèmes de comptage intelligent est mise en évidence par la proposition d'une approche globale qui inclut un cas concret de mise en œuvre.
Les limites de ce travail résident principalement dans l'absence d'un cadre expérimental de mise en œuvre des méthodes d'agrégation proposées pour vérifier leurs impacts dans un réseau de capteurs intelligents. Le jeu de données n'a pas permis de présenter des simulations sur les agrégations par zone et par usage également possibles à l'aide des méthodes d'agrégation mises en œuvre.
Nous allons dans la suite de nos travaux analyser les performances de nos méthodes d'agrégation comparativement à d'autres méthodes relativement à la diminution de la consommation d'énergie dans les réseaux de capteurs sans fil.
Cependant, les nouvelles méthodes d'agrégation proposées constituent des nouvelles techniques éligibles à la réduction du trafic et à l'amélioration de l'efficacité énergétique dans les réseaux de capteurs sans fil.

Références

1. Gupta, R.; Chaturvedi, K.T. Adaptive Energy Management of Big Data Analytics in Smart Grids. Energies 2023, 16, 6016. https://doi.org/10.3390/en16166016.
2. Chen, Z.; Amani, A.M.; Yu, X.; Jalili, M. Control and Optimisation of Power Grids Using Smart Meter Data: A Review. Sensors 2023, 23, 2118. https://doi.org/10.3390/s23042118.
3. S. R. Selva Jeevitha. Making ease of smart grid communication through compression of power system disturbance signals, 21 July 2023, PREPRINT (Version 1) available at Research Square [https://doi.org/10.21203/rs.3.rs-3172718/v1].
4. Soham Dutta, Sourav Kumar Sahu, Millend Roy, Swarnali Dutta, A data driven fault detection approach with an ensemble classifier based smart meter in modern distribution system, Sustainable Energy, Grids and Networks, Volume 34, 2023, 101012, ISSN 2352-4677, https://doi.org/10.1016/j.segan.2023.101012.
5. S. Banik, S. K. Saha, T. Banik and S. M. M. Hossain, "Anomaly Detection Techniques in Smart Grid Systems: A Review," *2023 IEEE World AI IoT Congress (AIIoT)*, Seattle, WA, USA, 2023, pp. 0331-0337, doi: 10.1109/AIIoT58121.2023.10174485.
6. Lulu Wen, Kaile Zhou, Shanlin Yang, Lanlan Li, Compression of smart meter big data: A survey, Renewable and Sustainable Energy Reviews, Volume 91, 2018, Pages 59-69, ISSN 1364-0321, https://doi.org/10.1016/j.rser.2018.03.088.
7. Lee J, Yoon S, Hwang E. Frequency Selective Auto-Encoder for Smart Meter Data Compression. Sensors. 2021; 21(4):1521. https://doi.org/10.3390/s21041521
8. Sung, Tien-Wen et al. 'Optimizing Data Aggregation Point Location with Grid-based Model for Smart Grids'. 1 Jan. 2022 : 3189 – 3201.
9. Rajagopalan, Ramesh and Varshney, Pramod K., "Data aggregation techniques in sensor networks: A survey" (2006). Electrical Engineering and Computer Science - All Scholarship. 22. https://surface.syr.edu/eecs/22
10. Geetika Dhand, S.S. Tyagi, Data Aggregation Techniques in WSN:Survey, Procedia Computer Science, Volume 92, 2016, Pages 378-384, ISSN 1877-0509, https://doi.org/10.1016/j.procs.2016.07.393.
11. M. S. H. Nazmudeen, A. T. Wan and S. M. Buhari, "Improved throughput for Power Line Communication (PLC) for smart meters using fog computing based data aggregation approach," 2016 IEEE International Smart Cities Conference (ISC2), Trento, Italy, 2016, pp. 1-4, doi: 10.1109/ISC2.2016.7580841.
12. Jagdish Chandra Pandey and Mala Kalra. An approach for secure data transmission in smart grids. Published Online:February 2, 2023pp 348-365https://doi.org/10.1504/IJICS.2023.128830.
13. Pandey, Vaibhav, Amarjeet Kaur, and Narottam Chand. "A review on data aggregation techniques in wireless sensor network." Journal of Electronic and Electrical Engineering 1.2 (2010): 01-08.
14. H. Harb, A. Makhoul, S. Tawbi and R. Couturier, "Comparison of Different Data Aggregation Techniques in Distributed Sensor Networks," in IEEE Access, vol. 5, pp. 4250-4263, 2017, doi: 10.1109/ACCESS.2017.2681207.

15. Vaidyanathan, K., Sur, S., Narravula, S., & Sinha, P. (2004). Data aggregation techniques in sensor networks. Osu-cisrc-11/04-tr60, The Ohio State University.
16. Ghiasi, M., et al.: Evolution of smart grids towards the Internet of energy: concept and essential components for deep decarbonisation. IET Smart Grid. 6(1), 86–102 (2023). https://doi.org/10.1049/stg2.12095.
17. A. Lang, Y. Wang, C. Feng, E. Stai and G. Hug, "Data Aggregation Point Placement for Smart Meters in the Smart Grid," in IEEE Transactions on Smart Grid, vol. 13, no. 1, pp. 541-554, Jan. 2022, doi: 10.1109/TSG.2021.3119904.
18. ABBASIAN DEHKORDI, Soroush, FARAJZADEH, Kamran, REZAZADEH, Javad, et al. A survey on data aggregation techniques in IoT sensor networks. Wireless Networks, 2020, vol. 26, p. 1243-1263.
19. Guguloth Ravi, M. Swamy Das, Karthik Karmakonda, Reliable cluster based data aggregation scheme for IoT network using hybrid deep learning techniques, Measurement: Sensors, Volume 27, 2023, 100744, ISSN 2665-9174, https://doi.org/10.1016/j.measen.2023.100744.
20. Singh, A.K., Kumar, J. A secure and privacy-preserving data aggregation and classification model for smart grid. Multimed Tools Appl 82, 22997–23015 (2023). https://doi.org/10.1007/s11042-023-14599-4.
21. Singh, A.K., Kumar, J. A privacy-preserving multidimensional data aggregation scheme with secure query processing for smart grid. J Supercomput 79, 3750–3770 (2023). https://doi.org/10.1007/s11227-022-04794-9
22. Dramé, Cheikh (2014) : Resolving West Africa's electricity dilemma through the pursuit of smart grid opportunities, 20th Biennial Conference of the International Telecommunications Society (ITS): "The Net and the Internet - Emerging Markets and Policies" , Rio de Janeiro, Brazil, 30th-03rd December, 2014, International Telecommunications Society (ITS), Calgary.
23. Sebastian Sterl, Inne Vanderkelen, Celray James Chawanda, Daniel Russo, Robert J. Brecha, Ann van Griensven, Nicole P. M. van Lipzig and Wim Thiery. Smart renewable electricity portfolios in West Africa. Nat Sustain 3, 710–719 (2020). https://doi.org/10.1038/s41893-020-0539-0.
24. S. M. Kadri, A. O. Bagré, M. B. Camara, B. Dakyo and Y. Coulibaly, "Electrical Power distribution status in West Africa: Assessment and Perspective Overview," 2019 8th International Conference on Renewable Energy Research and Applications (ICRERA), 2019, pp. 511-515, doi: 10.1109/ICRERA47325.2019.8997112.
25. Osama Majeed Butt, Muhammad Zulqarnain, Tallal Majeed Butt. Recent advancement in smart grid technology: Future prospects in the electrical power network. Ain Shams Engineering Journal 12 (2021) 687–695.
26. Young, Jacob R., "Smart grid technology in the developing world" (2017). Honors Projects. 68. https://digitalcommons.spu.edu/honorsprojects/68.

5.

27. Fernando Antonanzas-Torres, Javier Antonanzas and Julio Blanco-Fernandez. State-of-the-Art of Mini Grids for Rural Electrification in West Africa. Energies 2021, 14(4), 990; https://doi.org/10.3390/en14040990.
28. Yazid Hambally Yacouba, Amadou Diabagaté, Michel Babri and Adama Coulibaly, “Design, Aggregation and Analysis of Power Consumption Data using the Jump Process” International Journal of Advanced Computer Science and Applications(IJACSA),12(5), 2021. http://dx.doi.org/10.14569/IJACSA.2021.0120567.
29. Artur Trindade, artur.trindade '@' elergone.pt, Elergone, NORTE-07-0202-FEDER-038564 Data type: TS, Smart Meter Data Portal, 370 Client Electricity Loads 2011 - 2014, https://smda.github.io/smart-meter-data-portal.

8.2. Identification des personnes susceptibles d’avoir la covid-19 par les techniques de l’apprentissage automatique

Diako Doffou Jérome[12], JOHNSON Grâce Yénin Edwig [12] Sandjé Marcellin [23]

[1] Ecole supérieur Africaine de technologie, Abidjan
[2] Inphb , Yamoussoukro
[3] LARIT , LASTIC
correspondent author mail : jerome_diako@esatic.edu.ci

Résumé.

Un nouveau coronavirus est apparu dans la ville de Wuhan (province du Hubei, Chine) en décembre 2019. La maladie, nommée COVID-19, est causée par le coronavirus 2 du syndrome respiratoire aigu sévère (SRAS-CoV-2) étant très contagieuse. Le virus s'est rapidement propagé en dehors de la Chine et l'Organisation mondiale de la santé (OMS) a reconnu l'épidémie comme une pandémie en mars 2020. L'infection par le SRAS-CoV-2 est caractérisée par de la fièvre, une faiblesse généralisée, une toux sèche, des maux de tête, une dyspnée, une myalgie, ainsi qu'une leucopénie, une lymphopénie, une neutrophilie, des taux élevés de protéine C-réactive, de D-dimères et de cytokines et la perte de l'odorat et du goût au stade précoce de l'infection. Vu la propagation rapide du COVID-19, il est nécessaire que l'on s'intéresse à l'utilisation des techniques de l'apprentissage automatique sur différents fronts tels que le diagnostic rapide et précis et l'identification des personnes les plus sensibles à la maladie. L'objectif de cet article est de proposer une méthode d'identification précoce des patients à risque accru de développer des symptômes graves de la COVID-19 en utilisant une base de données disponible de l'État Ivoirien. En utilisant des techniques de l'apprentissage automatique, un problème de classification peut être résolu dans le but de prédire l'état (négatif, positif) des personnes par rapport à la COVID-19. Les 75 % de l'ensemble de données d'apprentissage ont été utilisés pour l'apprentissage des modèles, tandis que les 25 % restants ont été utilisés pour tester les modèles. Le résultat de l'évaluation des performances des modèles a montré que le modèle d'arbre de décision a la plus grande exactitude (97 %).

Mots-clés. COVID-19, SARS-CoV-2, Apprentissage Automatique

1 Introduction

Un nouveau coronavirus est apparu dans la ville de Wuhan (province du Hubei, Chine) en décembre 2019. La maladie, nommée COVID-19, est causée par le coronavirus 2 du syndrome respiratoire aigu sévère (SRAS-CoV-2) étant très contagieuse. Le virus s'est rapidement propagé en dehors de la Chine et l'Organisation mondiale de la santé (OMS) a reconnu l'épidémie comme une pandémie en mars 2020. L'infection par le SRAS-CoV-2 est caractérisée par de la fièvre, une faiblesse généralisée, une toux sèche, des maux de tête, une dyspnée, une myalgie, ainsi qu'une leucopénie, une lymphopénie, une neutrophilie, des taux élevés de protéine C-réactive, de D-dimères et de cytokines , de raideur musculaire et la perte de l'odorat et du goût au stade précoce de l'infection. Cependant, l'état peut rapidement évoluer vers un syndrome de détresse respiratoire aiguë (SDRA), une tempête de cytokines, un dysfonctionnement de la coagulation, une lésion cardiaque aiguë, une insuffisance rénale aiguë et un dysfonctionnement de plusieurs organes si la maladie n'est pas résolue, entraînant le décès du patient.
Vu la propagation rapide du COVID-19, il est nécessaire que l'on s'intéresse à l'utilisation des techniques de l'apprentissage automatique sur différents fronts tels que le diagnostic rapide et précis et l'identification des personnes les plus sensibles à la maladie.
L'objectif de cet article est de faire un pronostic ou une identification précoce des patients à risque accru de développer des symptômes graves de la COVID-19 en utilisant une base de données disponible de l'État Ivoirien. En utilisant des techniques de l'apprentissage automatique, un problème de classification peut être résolu

dans le but de prédire l'état (négatif, positif) des personnes par rapport à la COVID-19, sur la base d'informations individuelles, en plus des comorbidités et des symptômes cliniques variés. Ce processus peut être d'une grande importance pour aider à la prise de décision au sein des hôpitaux, car les ressources deviennent chaque jour limitées. Les patients classés comme ayant une maladie plus grave peuvent être prioritaires dans ce cas.
L'apprentissage automatique est capable de détecter les maladies et les infections virales avec plus de précision afin que la maladie des patients puisse être diagnostiquée à un stade précoce, que les stades dangereux des maladies puissent être évités et qu'il y ait moins de patients. Dans cet article, nous nous interrogerons comment l'apprentissage automatique (Machine Learning) permet de développer des solutions intelligentes pour diagnostiquer la COVID-19 chez les populations.
Afin de traiter le sujet et de répondre aux questionnements émis, un plan de recherche a été établi. Tout d'abord, nous avons mené des entretiens avec des professionnels. La recherche empirique a enfin été complétée par de nombreuses lectures sur le sujet.
Dans ce travail, nous nous sommes appuyés sur un ensemble de données épidémiologiques étiquetées pour les cas positifs et négatifs de COVID-19 de la Côte d'Ivoire et sur des algorithmes d'apprentissage supervisés tels que l'arbre de décision, la foret aléatoire, le gradient boosting,

2 Etat de l'art sur la prédiction de la maladie COVID19 à partir des algorithmes de machine Learning

Swapnarekha et al, (2020) ont présenté une revue de littérature sur l'utilisation des algorithmes intelligents dans la prédiction de la maladie COVID19. Il s'agit des algorithmes d'apprentissage automatique, de deep learning et ceux basés sur les modèles mathématiques et statistiques. Selon l'étude réalisée, l'analyse de la littérature révèle que 39% des recherches sont axés sur l'utilisation des algorithmes de deep learning, 37% concernent l'utilisation des algorithmes de machine learning et les autres 24% sont attribués à l'application des modèles mathématiques et statistiques. Il ressort que ces approches informatiques intelligentes ont été utilisées avec profit dans la prédiction et le dépistage de la pandémie de COVID19 [1].

Kwekha-Rashid et al, (2021) a montré que parmi les travaux utilisant les algorithmes de machine learning, les algorithmes supervisés sont beaucoup utilisés comparativement aux algorithmes non supervisés. Dans cette étude, on estime leur utilisation à 92,9% et l'utilisation des algorithmes non supervisés à 7,1%. De plus, les modèles construits sont à 86% basés sur la classification, 7% pour les modèles de régression et 7% pour les modèles à base de clustering. Cette étude porte essentiellement sur la regression logistique, les réseaux de neurones artificiels, les réseaux de neurones convolutionels, la régression linéaire, le K-means, les K-NN et le Naïve Bayes. Bien que cette étude présente les avantages de l'utilisation des algorithmes supervisés, nous notons que cette liste est non exhaustive. En effet, dans le cadre de l'utilisation des algorithmes de machine learning pour la prédiction de la maladie COVID19, plusieurs autres algorithmes ont été utilisés, ces dernières années pour des prédictions [2].

Prakash et al, (2020) ont présenté un cadre d'analyse de la maladie COVID19, en vue de la comprendre comment elle affecte les personnes. L'étude révèle que les personnes âgées de 20 à 50 ans sont les plus touchés par la maladie. Les algorithmes d'apprentissage automatique tels que SVM, KNN+NCA, Decision Tree, Gaussian Naïve Bayes, Multilinear Regression, Logistic Regression et XGBoost ont été utilisé pour construire des modèles de prédiction. Les données utilisées sont caractérisées par l'âge, le nombre de décès, de guéris, d'étrangers confirmés et de nationaux confirmés. Le RF a réalisé la meilleure performance avec un accuracy de 96%.[3]

Rustam et al, (2020) ont démontré la capacité des modèles de machine learning à prévoir le nombre de patients affectés par le COVID19. Quatre modèles de prédiction ont été utilisés à savoir, la régression linéaire (LR), l'opérateur de sélection et de rétrécissement le moins absolu (LASSO), la machine à vecteur de support (SVM) et le lissage exponentiel (ES). Cette étude avait pour but de réaliser trois types de prédiction telles que le nombre le nombre de nouveaux cas d'infection, le nombre

de décès et le nombre de guérisons dans les 10 prochains jours. Les résultats ont indiqué que l'ES était le mieux adapté à l'étude des 3 cas suscités, ensuite viennent le LR et le LASSO. Cependant le SVM, dans tous les scénarii de prédiction a été peu performant [4].

Une autre étude a été réalisée par **N. N. Sun et al(2020)** dont le but était d'extraire les facteurs de risque des données cliniques des patients infectés par le COVID19 en utilisant quatre algorithmes tels que le LR, le SVM, le DT, le RF et le DNN. Les résultats indiquent que le modèle prédictif LR a obtenu de meilleures performances en termes de précision, de sensibilité et d'accuracy[5]. Cet article décrit la capacité des méthodes d'apprentissage automatique à améliorer la précision et la rapidité du diagnostic précoce de cette maladie.

Les travaux concernant l'utilisation des algorithmes de l'apprentissage automatique pour la prédiction de la maladie COVID19 sont nombreux et s'étendent principalement sur ces deux dernières années. Nous les résumons dans le tableau 1 ci-dessous. Ce tableau présente une vue générale des algorithmes de l'apprentissage automatique utilisés au cours des années 2020 et 2021.

*Table 2:**Algorithmes utilisés dans la prédiction de la maladie à COVID19***

Année	Auteurs	Algorithmes utilisés	Meilleurs Algorithmes
2020	Prakash et al [3]	RF, SVM, DT, GNB, MLR, LR, XGBoost, KNN	RF
	Rustam et al [4]	LR, LASSO, SVM, ES	ES
	Batista et al [6]	RN, RF, GBT, LR, SVM	SVM
	Swapna Rekha et al [1]	SVM, RF, FFT-Gabor Sheme	SVM
	Schwab et al [7]	LR, RN, SVM, RF, GB	SVM, RF et XGB selon les cas
	Sun et al [5]	LR, RN, SVM, RF, DT, DNN	LR
2021	Zoabi et al [8], Kukar et al [9]	XGBoost	XGBoost
	Rahman et al [10]	CSDC-SVM, LR, LR/CNN	CSDC-SVM
	Muhammad et al [11]	DT, SVM, Naïve Bayes, LR, ANN	DT
	Charlyn et al [12]	J48 DT, RF, SVM, KNN, Naïve Bayes	SVM et RF
	Sharma et al [13]	SVM, FS-SVM, HPO-FS-SVM	HPO-FS-SVM
	Dutta et al [14]	KNN, RF, Bagging	RF

3 Approche proposée

3.1 Dataset

Cette étude repose sur un ensemble de données épidémiologiques concernant les cas positifs et négatifs de COVID-19 en Côte d'Ivoire. Ces données ont été recueillies par L'Institut National d'Hygiène Publique (INHP). L'ensemble de données provient du Système de Surveillance Épidémiologique. L'ensemble de données comprend 10000 enregistrements, comportant 13 caractéristiques, notamment des symptômes.

Table 3: Profile information of the dataset

N°	Fièvre	Sensation fièvre	Fatigue	Toux	Courbatures	Perte de gout et /ou odorat	Maux de tête	Diarrhée	Contact ayant les signes	Contact cas confirmé	co-morbidités	traitement quotidien	Decision
001	1	1	1	1	1	1	1	1	1	1	0	0	1
002	1	1	1	1	1	1	1	0	0	1	0	0	1
003	1	0	1	1	1	1	1	1	0	1	0	0	1
004	1	0	1	1	1	1	1	1	1	1	0	0	1
....	.....	.....	.....	.....	.....	.....	.	.....	.....	.....	.....	.....	.....

Les noms des caractéristiques sont : Fièvre Sensation fièvre, Fatigue, Toux, Courbatures Perte de gout et /ou odorat, Maux de tête, Diarrhée, Contact ayant les signes, Contact cas confirmé , comorbidités, traitement quotidien et Decision. Dans le cadre de ce travail, ces caractéristiques symptomatiques, ont été pris en compte dans l'ensemble de données. L'ensemble de données original code 1 pour les résultats positifs et 0 pour les résultats négatifs.

3.2 Algorithme IDPATIANT

Pour mettre en place notre modèle, nous avons suivi les étapes suivantes :
Etape 1 : Charger les données prétraitées
Etape 2 : Diviser les données en ensembles d'entraînement et de test
Etape 3 : Une mise à l'échelle des caractéristiques
Etape 4 : des modèles et des grilles de recherche d'hyperparamètres
Etape 5 : Dictionnaires pour stocker les résultats
Etape 6 : Entraîner et évaluer les modèles
Etape 7 : Afficher les scores
Etape 8 : Tracer les graphiques
Etape 9 : Bar plot des scores d'exactitude, de précision, de rappel et de fscore
Etape 10 : Afficher les graphiques
Etape 11 : Générez les règles de l'arbre de décision
Etape 12 : Affichez l'image de l'arbre de décision

3.3 Experimentation

Les algorithmes d'apprentissage automatique supervisés suivants, l'arbre de décision, la forêt aléatoire et le gradient boosting ont été exécutés à l'aide du langage de programmation Python dans un environnement de système d'exploitation Windows déployé sur un système informatique de marque Thinkpad (ordinateur portable), Corei7 avec 16 Go de RAM et une vitesse de processeur de 2,9 GHz. Toutes les bibliothèques nécessaires ont été installées sur un notebook Python et utilisées pour l'analyse des données, y compris l'analyse de corrélation et le développement des modèles. Les modèles ont été évalués à l'aide d'une mesure d'évaluation des performances d'exactitude, de sensibilité et de spécificité afin de déterminer leur efficacité et leur qualité.

3.4 Resultat et discussion

La prédiction précoce du patient de la COVID-19 est très utile car elle réduit le fardeau qui pèse sur les systèmes de santé en aidant à diagnostiquer les patients atteints de la COVID-19. Dans ce travail, des modèles d'arbre de décision, de forêt aléatoire et de gradient boosting pour la prédiction de l'infection au COVID-19 à l'aide d'un ensemble de données épidémiologiques pour les cas positifs et négatifs de COVID-19 en Côte d'Ivoire ont été développés. Les performances de tous les modèles ont été évaluées sur la base de paramètres de précision. Le résultat des performances des modèles est présenté respectivement dans la table 3.

Table 4: Evaluation de performance

Modèle	Exactitude	Précision	Rappel
Random Forest	0.87	0.85	0.875
Gradient Boosting	0.8	0.85	0.875
Decision Tree	0.97	0.98	0.925

Le meilleur modèle pour avec une haute précision et un rappel élevé est le Decision Tree.

Les règles générées

```
Règles de l'arbre de décision :
|--- Mauxdetete <= -0.75
|  |--- Diarrhee <= -0.58
|  |  |--- class: 0
|  |--- Diarrhee >  -0.58
|  |  |--- class: 0
|--- Mauxdetete >  -0.75
|  |--- Contactcasconfirme <= -0.91
|  |  |--- comorbidites <= 0.29
|  |  |  |--- class: 0
|  |  |--- comorbidites >  0.29
|  |  |  |--- class: 1
|  |--- Contactcasconfirme >  -0.91
|  |  |--- Courbatures <= -0.62
|  |  |  |--- Contactayantlessignes <= -0.15
|  |  |  |  |--- traitementquotidien <= 0.26
|  |  |  |  |  |--- class: 0
|  |  |  |  |--- traitementquotidien >  0.26
|  |  |  |  |  |--- class: 1
|  |  |  |  |--- Contactayantlessignes >  -0.15
|  |  |  |  |  |--- Sensationfievre <= 0.13
|  |  |  |  |  |  |--- class: 0
|  |  |  |  |  |--- Sensationfievre >  0.13
|  |  |  |  |  |  |--- class: 0
|  |  |  |--- Courbatures >  -0.62
|  |  |  |  |--- Fatigue <= -0.18
|  |  |  |  |  |--- Pertedegoutetouodorat  <= -0.34
|  |  |  |  |  |  |--- class: 0
|  |  |  |  |  |--- Pertedegoutetouodorat  >  -0.34
|  |  |  |  |  |  |--- Toux <= -0.66
|  |  |  |  |  |  |  |--- class: 0
|  |  |  |  |  |  |--- Toux >  -0.66
|  |  |  |  |  |  |  |--- Fievre <= -0.05
|  |  |  |  |  |  |  |  |--- class: 0
|  |  |  |  |  |  |  |--- Fievre >  -0.05
|  |  |  |  |  |  |  |  |--- class: 1
|  |  |  |  |--- Fatigue >  -0.18
|  |  |  |  |  |--- Pertedegoutetouodorat  <= -0.34
|  |  |  |  |  |  |--- comorbidites <= 0.29
|  |  |  |  |  |  |  |--- Fievre <= -0.05
|  |  |  |  |  |  |  |  |--- class: 1
|  |  |  |  |  |  |  |--- Fievre >  -0.05
|  |  |  |  |  |  |  |  |--- class: 1
|  |  |  |  |  |  |--- comorbidites >  0.29
|  |  |  |  |  |  |  |--- class: 1
|  |  |  |  |  |--- Pertedegoutetouodorat  >  -0.34
|  |  |  |  |  |--- class: 1
```

L'arbre de décision ci-dessus est utilisé pour prédire si une personne est infectée par la COVID-19. Il est basé sur un ensemble de données de patients avec des symptômes et des facteurs de risque de la COVID-19. L'arbre de décision est divisé en trois branches principales :

- Branche 1 : Si la personne a un mauvais état général (Mauxdetete), l'arbre de décision vérifie si elle a de la diarrhée. Si c'est le cas, la personne est probablement infectée par la COVID-19. Sinon, la personne est probablement saine.
- Branche 2 : Si la personne n'a pas un mauvais état général, l'arbre de décision vérifie si elle a été en contact avec un cas confirmé de COVID-19. Si c'est le cas, l'arbre de décision vérifie si la personne a des comorbidités. Si la personne a des comorbidités, elle est probablement infectée par la COVID-19. Sinon, la personne est probablement saine.
- Branche 3 : Si la personne n'a pas été en contact avec un cas confirmé de COVID-19, l'arbre de décision vérifie si elle a des symptômes spécifiques de la COVID-19, tels que des courbatures, de la fatigue, une perte de goût ou d'odorat, une toux ou de la fièvre. Si la personne a un ou plusieurs de ces symptômes, elle est probablement infectée par la COVID-19. Sinon, la personne est probablement saine.

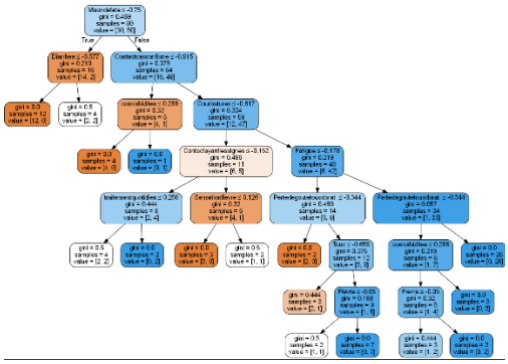


Fig.1 ***Modèle d'arbre de décision pour la prédiction d'infection au COVID-19***

Références

6. [1] H. Swapnarekha, H. S. Behera, J. Nayak, and B. Naik, "Role of intelligent computing in COVID-19 prognosis: A state-of-the-art review," *Chaos Solitons Fractals*, vol. 138, p. 109947, Sep. 2020, doi: 10.1016/j.chaos.2020.109947.
7. [2] A. S. Kwekha-Rashid, H. N. Abduljabbar, and B. Alhayani, "Coronavirus disease (COVID-19) cases analysis using machine-learning applications," *Appl Nanosci*, no. 0123456789, May 2021, doi: 10.1007/s13204-021-01868-7.
8. [3] K. B. Prakash, "Analysis, Prediction and Evaluation of COVID-19 Datasets using Machine Learning Algorithms," *International Journal of Emerging Trends in Engineering Research*, vol. 8, no. 5, pp. 2199–2204, May 2020, doi: 10.30534/ijeter/2020/117852020.
9. [4] F. Rustam *et al.*, "COVID-19 Future Forecasting Using Supervised Machine Learning Models," *IEEE Access*, vol. 8, pp. 101489–101499, 2020, doi: 10.1109/ACCESS.2020.2997311.
10. [5] N. N. Sun *et al.*, "A prediction model based on machine learning for diagnosing the early COVID-19 patients," *medRxiv*, pp. 1–12, 2020, doi: https://doi.org/10.1101/2020.06.03.20120881.
11. [6] B. Afm *et al.*, "COVID-19 diagnosis prediction in emergency care patients: a machine learning approach," *medRxiv*, vol. 102, 2020, doi: https://doi.org/10.1101/2020.04.04.20052092 ;
12. [7] P. Schwab, A. DuMont Schütte, B. Dietz, and S. Bauer, "Clinical Predictive Models for COVID-19: Systematic Study," *J Med Internet Res*, vol. 22, no. 10, p. e21439, Oct. 2020, doi: 10.2196/21439.
13. [8] Y. Zoabi, S. Deri-Rozov, and N. Shomron, "Machine learning-based prediction of COVID-19 diagnosis based on symptoms," *NPJ Digit Med*, vol. 4, no. 1, p. 3, Dec. 2021, doi: 10.1038/s41746-020-00372-6.
14. [9] M. Kukar *et al.*, "COVID-19 diagnosis by routine blood tests using machine learning," *Sci Rep*, vol. 11, no. 1, p. 10738, Dec. 2021, doi: 10.1038/s41598-021-90265-9.
15. [10] Atta-ur-Rahman *et al.*, "Supervised Machine Learning-Based Prediction of COVID-19," *Computers, Materials & Continua*, vol. 69, no. 1, pp. 21–34, 2021, doi: 10.32604/cmc.2021.013453.
16. [11] L. J. Muhammad, E. A. Algehyne, S. S. Usman, A. Ahmad, C. Chakraborty, and I. A. Mohammed, "Supervised Machine Learning Models for Prediction of COVID-19 Infection using Epidemiology Dataset," *SN Comput Sci*, vol. 2, no. 1, p. 11, Feb. 2021, doi: 10.1007/s42979-020-00394-7.
17. [12] C. N. Villavicencio, J. J. E. Macrohon, X. A. Inbaraj, J. H. Jeng, and J. G. Hsieh, "Covid-19 prediction applying supervised machine learning algorithms with comparative analysis using weka," *Algorithms*, vol. 14, no. 7, 2021, doi: 10.3390/a14070201.
18. [13] D. K. Sharma, M. Subramanian, Pacha. Malyadri, B. S. Reddy, M. Sharma, and M. Tahreem, "Classification of COVID-19 by using supervised optimized machine learning technique," *Mater Today Proc*, vol. 56, pp. 2058–2062, 2022, doi: 10.1016/j.matpr.2021.11.388.
19. [14] P. Dutta, S. Paul, and A. Kumar, "Comparative analysis of various supervised machine learning techniques for diagnosis of COVID-19," in *Electronic Devices, Circuits, and Systems for Biomedical Applications*, Elsevier, 2021, pp. 521–540. doi: 10.1016/B978-0-323-85172-5.00020-4.

8.3. Modèle d'optimisation multi-objectif prenant en compte le rayon de courbure des virages et l'inclinaison des pentes des routes

Coulibaly Kpinna Tiekoura[1] , Diaby Moustapha[1]

[1] LaSTIC, Ecole Supérieure Africaine des TIC, Abidjan, Côte d'Ivoire, tiekoura77@yahoo.fr

Abstract. La croissance économique en Afrique a entrainé une augmentation significative du trafic routier et singulièrement du transport public. Cette situation entraîne malheureusement des effets néfastes sur le plan environnemental et économique dans les zones urbaines. Pour y remédier, de nombreuses études ont été menées afin d'optimiser ce mode de transport. Dans cet article, nous proposons un système de gestion optimal pour le transport public, visant à minimiser divers critères tels que la distance parcourue, le temps de service des véhicules et les contraintes de sécurité (rayon de courbure des virages et inclinaison des pentes). Pour la résolution de notre modèle, nous avons utilisé une méthode exacte qui donne de bons résultats pour les instances de petites tailles.

Keywords: Smart-city, optimisation, graphe, intelligence artificielle, transport urbain, transport en commun

1. Introduction

Une gestion efficace des transports contribue à assurer un environnement sain en réduisant les émissions de gaz à effet de serre et en maîtrisant l'empreinte carbone liée au transport de biens et de services. Sur le plan économique, optimiser les déplacements des véhicules permet de diminuer les coûts énergétiques associés à l'utilisation de combustibles fossiles, dont les réserves s'amenuisent et dont les prix augmentent de manière insoutenable pour les consommateurs. Par exemple, dans un rapport, la Banque Mondiale [1] indique que les habitants d'Abidjan, la capitale économique de la Côte d'Ivoire, effectuent 10 millions de trajets par jour et qu'en 2017, la perte de temps et de productivité, ainsi que le désordre et les coûts liés aux déplacements quotidiens dans cette ville, représentaient près de 5 % du PIB. L'un des types de transport les plus pratiqués dans les villes africaines demeure le transport urbain. Ainsi l'amélioration de ce transport représenterait une solution efficace pour assurer un développement durable tout en répondant aux attentes des usagers. Cette optimisation permet d'obtenir un meilleur rapport qualité-prix en établissant un équilibre entre le coût d'exploitation et la qualité du service proposé aux utilisateurs.

Dans ce travail, nous nous intéressons au transport à la demande (TAD) qui est considéré comme un mode de transport collectif, individualisé et activé selon les besoins. Notre approche d'optimisation prend en compte, en plus de la minimisation du coût de transport et du temps de

parcours, la minimisation du rayon de courbure des virages et d'inclinaison des pentes au niveau des routes, afin de garantir les meilleurs itinéraires aux usagers.

Notre étude se structurera en quatre sections principales. La première section présentera une description de l'état de l'art. Dans la deuxième section, nous définirons la problématique. La troisième section sera dédiée à l'analyse conceptuelle et à l'implémentation de notre système. Enfin, la quatrième section portera sur la discussion des résultats obtenus. Nous conclurons par une synthèse et des perspectives de recherche.

2.État de l'art

La plupart des algorithmes d'optimisation des transports en commun se focalisent principalement sur la réduction de la distance parcourue par les véhicules. Dans les études existantes, on fait la distinction entre les algorithmes à critère unique et ceux à critères multiples. Nous allons ici exposer quelques recherches académiques importantes sur ce thème.
L'un des algorithmes fondateurs à la base des différentes recherches de plus court chemin dans un graphe est celui de Dijkstra [2].
L'algorithme de Dijkstra est utilisé pour créer un arbre des plus courts chemins à partir d'une source x dans un graphe, permettant d'atteindre tous les autres nœuds accessibles. Il convient de souligner que la complexité des algorithmes d'optimisation pour les transports en commun, selon la littérature, dépend du nombre de critères ou d'objectifs considérés. Ainsi, on distingue dans les travaux existants des approches d'optimisation à un critère, à deux critères et multicritères.
Tomhave et al. [3] ont également développé un modèle d'optimisation qui prend en compte les profils des utilisateurs plutôt que de se concentrer uniquement sur le chemin le plus court. Ils partent du principe que de nombreux usagers préfèrent éviter le chemin le plus court en distance pour se déplacer d'un point A à un point B, en raison des longs temps d'attente qu'il peut engendrer. Dans un premier temps, les auteurs identifient le chemin le plus court parmi plusieurs itinéraires pertinents générés à l'aide de l'algorithme de Dijkstra. Ce chemin est ensuite utilisé pour calculer d'autres trajets en supprimant certains nœuds et arcs qu'il emprunte. Après plusieurs étapes de suppression et d'exploration de nœuds et d'arcs, le nombre souhaité de résultats est atteint. Enfin, les résultats sont filtrés afin d'éliminer les doublons et les trajets jugés trop longs par rapport au chemin le plus court.
Par ailleurs, Zidi et ses collègues [4] présentent une approche dynamique multicritères en recourant à l'algorithme de recuit simulé multi-objectif. Cette méthode intègre également la gestion des profils utilisateurs afin de personnaliser l'offre de services grâce à un Système Multi-Agents.

3. Objectif et Problématique

La majorité de ces études ne tiennent pas compte des contraintes de circulation dans leur approche d'optimisation des trajets. En effet, il est logique que les usagers préfèrent généralement des itinéraires qui soient à la fois sûrs et qui minimisent le nombre de détours. Ainsi, un trajet peut être court en distance, mais comporter des risques en raison de certains obstacles, tels que le rayon de courbure des virages ou l'inclinaison des pentes. Cela peut amener un usager à opter pour un chemin plus long, mais avec moins d'obstacles [5].

Ce choix peut être justifié par le gain de temps qu'il permet, car plus un virage est serré, plus le conducteur doit ralentir pour le négocier, comme l'indique une étude menée par Khoo et al [6].

Aussi, contrairement à certaines recherches présentées dans cet état de l'art, une optimisation efficace du transport en commun nécessite de prendre en compte plusieurs critères. L'objectif de notre étude est donc de développer un système de gestion optimal pour le transport public, visant à minimiser divers critères tels que le coût du trajet, le temps de service des véhicules et les risques sécuritaires (angle des virages et degré des pentes).

4 Méthodologie

4.1 Modélisation du problème

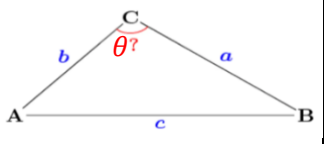
Dans cette section nous présentons la modélisation mathématique de notre proposition.

- *Variables du problème :*
  - P = pente de l'arc emprunté de i vers j
  - θ = angle de virage ou de courbure entre deux arcs du trajet de i vers j
  - Δ = matrice indiquant la distance euclidienne entre deux nœuds de $\mathcal{N}$
  - C = nœud courant visité, (C $\in \mathcal{N}$)
  - A = avant-dernier nœud du chemin liant le point de départ I à C (A $\in \mathcal{N}$)
  - B = nœud voisin de C n'ayant pas été visité (B $\in \mathcal{N}$)
  - $\mathcal{A}$: Ensemble des arrêtes ou arcs du graphe $\mathcal{G}$
  - n : Nombre de demandes de transport.
  - $\mathcal{D} = \{1, \dots, n\}$: ensemble de points d'arrêt des usagers à transporter
  - $\mathcal{F} = \{n+1, \dots, 2n\}$: Ensemble de points d'arrivée des usagers.
  - $\mathcal{S}$: Ensemble des points de stationnement (ou dépôts) des véhicules.
  - $\mathcal{K}$ = Ensemble des points limitant les arrêtes du graphe $\mathcal{G}$.
  - $\mathcal{N}$ = Ensemble de tous les nœuds du graphe $\mathcal{G}(\mathcal{N}, \mathcal{A})$
  - $\mathcal{V}$: Ensemble des véhicules destinés au transport des usagers.
  - Q_v : Capacité du véhicule v.
  - n_{vi} : Nombre d'usagers pris par le véhicule v à la station i tel que i∈D.
  - $[t_i\ k_i]$: Intervalle de temps associé au départ i∈D.

- $[t_{i+n}\ \ k_{i+n}]$: Intervalle de temps associé à l'arrivée i+n ∈D.
- C_v: Coût d'utilisation du véhicule v au kilomètre.
- $C_{ijv} = C_{ij} \times C_v$: Coût de transport de i vers j avec le véhicule v tel que
- T_{ijv} : Durée de transport de i vers j avec le véhicule v.
- t_{siv}: Heure de début de service pour la demande i avec le véhicule v.
- t_{aiv}: Heure d'arrivée de la demande i à destination avec le véhicule v.
- N_{iv}: Nombre de passagers dans le véhicule *v* après avoir visité le point i tel que i ∈ $\mathcal{N}$.
- N_{sv}: Nombre de passagers dans le véhicule *v* à la sortie d'un dépôt
- ξ_{ijv}: Variable de décision du problème, ξ_{ijv}= 1 si le véhicule emprunte un chemin direct de i vers j, sinon $\xi_{ijv} = 0$.
- ψ_v = coordonnées *(x,y,z)* des nœuds de l'ensemble $\mathcal{N}$
- $\mathcal{L}_{\mathcal{A}}$ = vecteur indiquant la longueur des arcs de l'ensemble $\mathcal{A}$

- *La Fonction Objectif* :
  Notre fonction objectif globale F est composé de la somme du coût de transport (f_{CT}), du temps de service (f_{TS}) et de la sécurité de la route (f_{SR}).

$F = f_{CT} + f_{TS} + f_{SR}$	(E.1)
$f_{CT} = \sum_{i\in D}\sum_{j\in D}\sum_{v\in \mathcal{V}} \xi_{ijv} C_{ijv}$	(E.2)
$f_{TS} = \sum_{i\in D}\sum_{v\in \mathcal{V}} (t_{aiv} - t_{siv})$	(E.3)
$f_{SR} = Pente\ (P) + Courbure\ virage\ (CV)^{-1}$	(E4)
$Pente\ (P) = 100 \times \frac{\psi_v(B,3) - \psi_v(C,3)}{\mathcal{L}_{\mathcal{A}}(CB)}$	(E5)

La courbure de virage (CV) représentée par l'angle de courbure θ (fig. 1) se calcule selon la formule d'Al Kashi, comme suit :	 *Figure 1. Angle de courbure θ*
$\theta = \arccos\left(\frac{a^2 + b^2 - c^2}{2ab}\right)$ C'est-à-dire :	(E6)
$\theta = \arccos\left(\frac{\mathcal{L}_{\mathcal{A}}(AC)^2 + \mathcal{L}_{\mathcal{A}}(CB)^2 - \Delta(A,B)^2}{2\mathcal{L}_{\mathcal{A}}(AC) \times \mathcal{L}_{\mathcal{A}}(CB)}\right)$ Ainsi :	(E7)

$f_{SR} = \sum_{B\in\mathcal{N}}\sum_{C\in\mathcal{N}}\left[100 \times \frac{\psi_v(B,3) - \psi_v(C,3)}{\mathcal{L}_{\mathcal{A}}(CB)} + \arccos^{-1}\left(\frac{\mathcal{L}_{\mathcal{A}}(AC)^2 + \mathcal{L}_{\mathcal{A}}(CB)^2 - \Delta(\mathrm{A},\mathrm{B})^2}{2\mathcal{L}_{\mathcal{A}}(AC) \times \mathcal{L}_{\mathcal{A}}(CB)}\right)\right]$ avec	(E8)
$\Delta(\mathrm{A},\mathrm{B}) = \sqrt{\left(\psi_{\mathrm{v}}(\mathrm{B},1) - \psi_{\mathrm{v}}(\mathrm{A},1)\right)^2 + \left(\psi_{\mathrm{v}}(\mathrm{B},2) - \psi_{\mathrm{v}}(\mathrm{A},2)\right)^2}$	(E9)
Par conséquent, la fonction objectif global est :	
$F = \sum_{i\in D}\sum_{j\in D}\sum_{v\in\mathcal{V}} \xi_{ijv} C_{ijv} + \sum_{i\in D}\sum_{v\in\mathcal{V}}(t_{aiv} - t_{siv}) + \sum_{B\in\mathcal{N}}\sum_{C\in\mathcal{N}}\left[100 \times \frac{\psi_v(B,3) - \psi_v(C,3)}{\mathcal{L}_{\mathcal{A}}(CB)} + \arccos^{-1}\left(\frac{\mathcal{L}_{\mathcal{A}}(AC)^2 + \mathcal{L}_{\mathcal{A}}(CB)^2 - \Delta(\mathrm{A},\mathrm{B})^2}{2\mathcal{L}_{\mathcal{A}}(AC) \times \mathcal{L}_{\mathcal{A}}(CB)}\right)\right]$	(E10)

- *Notre modèle mathématique :*

La modélisation du problème consiste à minimiser la fonction objectif F, sous différentes contraintes que nous préciserons ci-dessous.

$\begin{cases} Min\ F(\xi_{ijv}) \\ s.c\ X \in C \end{cases}$	(E11)
▪ Un véhicule *v* ne commence le service en *j* qu'après avoir fini le service en *i* et emprunté l'arc (*i, j*) :	
$\xi_{ijv}(T_{siv} + T_{ijv} - T_{sjv}) \leq 0,\ \forall v \in \mathcal{V}, (i,j) \in \mathcal{A}$	(E12)
▪ Afin de réaliser le service d'embarquement à l'heure, chaque véhicule *v* se doit de respecter l'intervalle de temps de demande à un nœud de départ *i* :	
$t_i \leq T_{siv} \leq k_i,\ \forall i \in \mathcal{N},\ v \in \mathcal{V}$	(E13)
▪ Afin de réaliser le service à l'heure, chaque véhicule *v* se doit de respecter l'intervalle de temps d'arrivée à un nœud de d'arrivée *i+n* :	
$t_{i+n} \leq T_{aiv} \leq k_{i+n},\ \forall i \in \mathcal{N},\ v \in \mathcal{V}$	(E14)
▪ Le nombre d'usager dans un véhicule *v* après avoir quitté un nœud de départ *i* est supérieur à celui collecté en i et inferieur à la capacité maximale du véhicule :	
▪ $n_{vi} \leq N_{vi} \leq Q_v\ , \forall i \in \mathcal{D},\ v \in \mathcal{V}$	(E15)

5 Expérimentation et évaluation

5.1 Cadre expérimental

Dans le cadre de notre expérimentation, nous avons eu recours aux données provenant du benchmark établi par Cordeau et Laporte en 2003 [7]. Ce benchmark comprend 20 instances présentant des problèmes de tailles variées, allant de 24 à 144 demandes de transport (voir tableau 1). De plus, entre 3 et 13 véhicules homogènes ont servis ces demandes de transport. Nous avons également enrichi ces données avec des informations simulées concernant les arcs et les nœuds des routes, afin de simplifier le calcul des pentes et des angles de courbure des virages.

Tableau 1 : Caractéristiques des instances de test

Instances	Nombre de requêtes ou demandes	Nombre de véhicules k1 (a)	Nombre de véhicules k2 (b)
I_1	24	1	2
I_2	48	2	3
I_3	72	3	4
I_4	96	3	6
I_5	120	5	7
I_6	144	5	8
I_7	36	2	2
I_8	72	3	3
I_9	108	3	5
I_{10}	144	5	6
I_{11}	24	3	-
I_{12}	48	5	-
I_{13}	72	7	-
I_{14}	96	9	-
I_{15}	120	11	-
I_{16}	144	13	-
I_{17}	36	4	-
I_{18}	72	6	-
I_{19}	108	8	-
I_{20}	144	10	-

Dans la présente étude, nous avons évalué notre méthode sur cinq instances du benchmark proposé par Cordeau et Laporte (2003) à savoir : **I_1b** (24 demandes et 2 véhicules), **I_2b** (48 demandes et 3 véhicules), **I_4b** (96 demandes et 6 véhicules), **I_{19}a** (108 demandes et 8 véhicules) et **I_{20}a** (144 demandes et 10 véhicules)

La résolution du modèle a été faite avec une méthode exacte. Nous avons utilisé le solveur CPLEX 12.10. La figure 2 présente le schéma du système de résolution par cette méthode.

Notre approche a été exécutée sur un ordinateur portable Surface pro 7, de processeur 3,1 GHz x64 Intel, Core i7 quatre cœurs avec une mémoire vive de 16 Go 2133 MHz avec Système d'exploitation 64 bits.

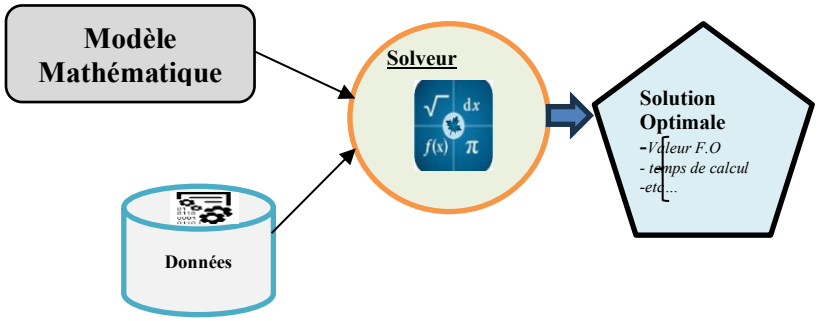


Figure 2 : Schéma du système de résolution par CPLEX.

5.2 Résultats et discussion

Le tableau ci-dessous présente les résultats obtenus à l'issue des tests

Tableau 2. Résultats obtenus avec la méthode exacte.

Instances	Distance parcourue	Durée du trajet (min)	Temps d'exécution (min)
I_1b	245,15	608,44	0,11
I_2b	812,56	2710,95	0,35
I_4b	1028,07	2929,42	2,28
I_{19}a	1776,32	3876,30	3,84
I_{20}a	2011,14	4118,77	6,16
TOTAL	**5 873,24**	**14243,88**	**12,74**

L'analyse du tableau de résultats montre que la durée des trajets ainsi que le temps d'exécution de l'algorithme augmentent exponentiellement en fonction de la taille de l'instance. Plus l'instance comporte un grand nombre de demandes et de véhicules, plus le temps d'exécution augmente. Ainsi, pour des tailles d'instances plus grandes, le problème deviendra impossible à résoudre de façon exacte en un temps raisonnable. Cependant, avec de petites tailles d'instances (telles que les instances I_1b et I_2b), on obtient une solution optimale en des temps assez raisonnables. Cela montre que pour la résolution des problèmes NP-difficile comme le présent problème d'optimisation des trajets, il est judicieux d'utiliser une heuristique. Ainsi, la prochaine étape de nos travaux sera l'implémentation de notre modèle en utilisant une heuristique telle que la recherche taboue.

Par ailleurs, il est important de souligner qu'avec notre approche, l'algorithme ne privilégie pas la distance du trajet en termes de longueur, mais tiens compte des contraintes de circulation (pente et angle de virage) qui en pratique augmentent le temps de parcourt. Avec notre approche, un véhicule empruntera donc un itinéraire long et comportant moins de virage et pente par rapport à un itinéraire court avec plus de virages et pentes.

6 Conclusion et futures travaux

Ce travail de recherche vise à améliorer la qualité des transports en commun, dans le but de fournir aux prestataires service de transport et aux usagers des services de qualité, alliant sécurité, économie et respect de l'environnement. Nous avons proposé un modèle d'optimisation du transport permettant de déterminer le trajet le plus court, tout en prenant en compte des contraintes qui influencent la sécurité et la durée du trajet, telles que les angles de virage et les pentes des routes.

Après la présentation de quelques recherches de la littérature concernant le problème des transports en commun, nous avons élaboré notre propre approche. Notre modèle mathématique intègre la minimisation de divers critères, tels que le coût du trajet, le temps de service des véhicules, ainsi que les contraintes de sécurité, notamment les angles de virage et les pentes.

Nous avons ensuite utilisé une méthode exacte pour la résolution de notre problème. Cela nous a permis d'obtenir des résultats avec un temps d'exécution assez réduit pour les instances de petites tailles, contrairement aux grandes instances.

En perspectives à ce travail nous comptons utiliser une heuristique telle que la recherche tabou afin de permettre une bonne exploration de l'espace de recherche ainsi que la réduction du temps de calcul. Concernant les données, il serait intéressant d'utiliser de réelles données sur le réseau routier avec des mesures sur les pentes et les angles de virage.

References

20. OxFord Business Group, The Report: Cote d'Ivoire 2019, https://oxfordbusinessgroup.com/news/pourquoi-la-c%C3%B4te-d%E2%80%99ivoire- besoin-de-moderniser-ses-infrastructures-de-transport [consulté le 7/09/2023]
21. Dijkstra, E. W. A note on two problems in connexion with graphs. Nume- rische mathematik, vol. 1, no 1, p. 269-271, 1959.
22. Tomhave B., Khani A. (2022). Refined choice set generation and the investigation of multi-criteria transit route choice behavior. *Transportation Research Part A: Policy and Practice*, *Vol. 155*. pp 484-500, https://doi.org/10.1016/j.tra.2021.11.005.
23. Zidi I. Zidi K. , Ghedira K, . Mesghouni K. (2010), A Multi-Objective Simulated Annealing for the Multi-Criteria Dial a Ride Problem11th IFAC/IFIP/IFORS/IEA Symposium on Analysis, Design, and Evaluation of Human-Machine Systems, Valenciennes France, 2010.
24. Akgol K., Guany B., Eldemir F., Samasti M. (2020). A new method to measure the rationalities of transit route layouts. *Case Studies on Transport Policy, Vol. 8*, pp. 1518-1530. https://doi.org/10.1016/j.cstp.2020.11.002
25. Khoo H. L., Ahmed M. (2018). Modeling of passengers' safety perception for buses on mountainous roads. Accident Analysis & Prevention,Vol. 113. Pp 106-116. https://doi.org/10.1016/j.aap.2018.01.025
26. Cordeau J-F, Laporte G (2003) A tabu search heuristic for the static multi-vehicle dial-a- ride problem, Transportation Research B 37:579–594

8.4. An Agent Based Model to Improve Traffic Flow and Assess the Impact of Smart Traffic Ligh

N'golo Konaté[1], Batiebo Mory Richard[2] and Mamadou Diarra[1]

[1]Université Felix Houphouet Boigny and [2]Université Virtuelle de Cote d'Ivoire

Abstract: Traffic in developing countries undergoes severe perturbations due to either a misunderstanding of the physical phenomenon or the bad calibration of equipment such as traffic lights. In such circumstances, using mathematical models to predict various traffic scenarios becomes either computationally or mathematically challenging. Contrarily, agent-based modeling is a relatively recent method for simulating complex systems made up of agents whose behavior is governed by straightforward rules. In terms of traffic modeling, agent-based models are closer to microscopic models where the driver is considered as one entity.

In this paper, an agent-based model to estimate the impact of the environment on the flow is defined. The suggested approach is intended to model the spatiotemporal transmission process. The considered agent in the model decision depends on the rules programmed. The different rules are built around the space gap, the traffic light, and the presence of pedestrians. The driver's decision to go either to the left or right lane is a stochastic process. Several hypothetical scenarios have been considered to show the performance of the proposed model. Experimental results show the ideal lane and traffic light time on a principal and arterial lane

Keywords: Agent-Based Model; Dynamic traffic light ; Urban Mobility

1. Introduction to this Template

Traffic congestion has become a pervasive problem in urban areas across the globe, giving rise to substantial economic, environmental, and social consequences (huang2023urban; litman2021well). As urban populations continue to expand, and cities experience unprecedented growth, the need for efficient and sustainable transportation systems has never been more crucial. Researchers and policymakers have been exploring various approaches to address these challenges, including the development of more accurate traffic flow simulation models, the optimization of traffic light control systems, and the integration of intelligent transportation systems (ITS) (barmpounakis2015intelligent; zhu2019parallel).

One of the primary factors contributing to traffic congestion is vehicle density, defined as the number of vehicles per unit of road space (berhanu2023examining). Higher vehicle densities often result in increased travel times, fuel consumption, and greenhouse gas emissions, while simultaneously decreasing road safety and overall quality of life for residents (kawajiri2020lightweight). Consequently, a comprehensive understanding of the relationship between vehicle density and traffic flow is essential for the development of effective traffic management strategies that can mitigate the negative effects of congestion.

Traffic light control systems serve as a critical component in regulating traffic flow and reducing congestion, particularly at inter- sections (nedim2023advanced). The duration of green and red light

phases significantly impacts the efficiency of traffic movement, and optimizing these timings can help alleviate traffic issues in urban areas (huang2014affectroute). Nonetheless, determining the optimal traffic light timings is a complex problem contingent on numer- ous factors, such as traffic demand, road network characteristics,
and traffic flow dynamics (nigam2023review). Over the years, re- searchers have employed a variety of computational methods and simulation models to study traffic flow and optimize traffic light control systems (barmpounakis2015intelligent) (wang2020optimizing) (li2022human). (li2022human) proposed an optimal scheme that reduces the queue delay but the number of passing vehicles remains unchanged
These methods range from relatively simple cellular automata models (nagel1992cellular) to more sophisticated agent- based simulations (chen2018agent) and machine learning approaches (miletic2022review). These studies have provided valuable insights into traffic flow dynamics and the effects of various traffic light control strategies on congestion levels.
This paper seeks to contribute to the existing body of knowledge by presenting a comprehensive multi-lane traffic flow simulation model that investigates the impact of lanes number and traffic light timings on urban traffic congestion. Our model employs a simple grid-based representation of a road network with multiple lanes and integrates traffic light control mechanisms that can be adjusted to study their effects on traffic flow. We further examine the potential of incorporating intelligent transportation systems (ITS) and machine learning techniques to optimize traffic light control strategies based on real-time data (barmpounakis2016unmanned).
By analyzing the results of the simulation under various scenarios and traffic conditions, we aim to provide a deeper understanding of the factors that contribute to traffic congestion and explore potential strategies for mitigating its adverse effects. Moreover, our study seeks to inform policymakers and urban planners about the potential benefits and challenges associated with implementing advanced traffic management systems and inspire further research in this crucial field. The aim of this paper is to give the traffic engineer a simulation perspective of traffic evaluation in different scenarios.

2. Methodology

The model definition is the first part of traffic simulation. It involves giving a mathematical model of the traffic, describing the global environment (road structures, drivers and their characteristics, traffic light, and rules)

2.1 Agent-based modeling

A new way of simulating systems with interacting autonomous parts is called agent-based modeling(cuevas2020agent). Agents are arti- ficial individuals programmed to perform pre-defined operations (bemthuis2020using).
Agent represents the different entities that make up the environ- ment(lattice or multidimensional). Hence they can either cooperate to form a multi-agent model or work alone. Therefore the agent's behavior can be described as a physical system, such as simulations of evacuations, traffic etc.

The main advantage of agents is that they do not need strong mathematical theory or a high level of computing power. Despite their simple behaviors, they are capable of producing several globally complicated models (comportments) as a result of the modeling characteristics generated by the interactions of a group of simple agents. The term "global behavioral models" refers to coherent microscale patterns, such as recognized patterns of temporal, spatial, and behavioral structure, patterns of distribution, or patterns that are temporally, spatially, and behaviorally coherent.

2.2 Simulation framework

The model defined later corresponds to a traffic situation where vehi- cle follows each other near a traffic light. Therefore the environment is composed by:

• Scenarios: The study focuses on a traffic flow near a junction with a traffic light. The traffic timing is static and we want to check the impact of traffic light timing, vehicle density, and lane number on the flow. The simulation is conducted on a road of 20 meters

Figure 1: Carrefour 9 Kilo

• Agent:To achieve those objectives, we define our agents.

– Car agent: In a simulation context drivers and vehicles could be seen as one entity. However, they are not in real life. Because, the vehicle simulation is straightforward with speed, and acceleration, the driver is a decision-taker through an analysis of the traffic situation. Therefore we assumed that each agent moved toward a certain direction, could change their behaviors (decisions), and had a set of path solutions. they can be defined as adaptative, reactive, autonomous, and mobile agents. Each agent is considered as he is going at his own speed and the decision of changing lanes is taken by the observation of the next following cells. Although the driver has preferred cells, he will go to the empty one

– Pedestrian agent : Pedestrian agents are individuals person who walk through the transportation environment(road). He has to go from one place to another with a velocity from standing to a maximum speed corresponding to running. The pedestrians can either be walking, waiting, or halted due to an obstacle. In the context of this paper, the only interactions considered are the presence of a pedestrian on the road which can yield to an accident and disrupt the traffic.

– Traffic Light : A traffic Light is a road signal device that con- trols vehicle traffic at an intersection by alternating displaying traffic state. The traffic light interacts with the car agent in the flowing way red(mandatory stop), and green (vehicle move on). The transition between the two

states is guided by the traffic density. If the traffic density decreases to a given threshold the traffic becomes red otherwise it stays green. In this paper, the traffic light is supposed to have a random density from an interval

• Simulation: Simulating a single agent on a road section is quite simple since it just consists of acceleration and deceleration. However, when other agents get involved, it becomes necessary to take into consideration how the driver evolved and the decision he takes considering other agents. The simulation is done with discrete values, whereby each iteration is a time step. The initialization value is taken for the Ministry of Transportation. Then the traffic evolution is investigated through the analysis of travel time, and traffic throughput. Each agent is simulated by a Class function. The validation is done by real values taken from conductors on busy days. The implementation of an efficient simulation lies in the definition of a set of rules.

• Rules: The scenario is guided by different rules chosen among the traffic parameters. The traffic light timing is static therefore the traffic cannot absorb the flow in different situations.

2.3 Mathematical model

Since the traffic flow is an interaction between drivers and their environment. In this paper, we combined a set of mathematical equations describing the actions of considered components. Considering the aforementioned characteristics of the agents, we will define some guideline rules.

2.3.1 car following rules

The car following model used is an extension of Nagel-Schreckenberg nagel1992cellular in Titamare2020Analysis to express the velocity in different situations. The car following rules depends on the other drivers' behavior. Therefore there is cooperation between a car agent and others in the following situations. In each situation, the car adopts a new velocity Vnew

• Acceleration: The car accelerate if the space headway is suffi- ciently large and the preceding vehicle maintains a velocity higher than the considered vehicle, the car velocity is then between the current one and the maximal velocity given by the preceding. The lane changing and acceleration will be considered in the next section :

$$V_{new} = min[(V_{curr} + 1), V_{max}] \quad (1)$$

- Braking: The car brakes if the preceding is lower than the one of the considered vehicle. However, the driver in this situation maintains a safe space headway sufficient to react to an obstacle. The new velocity is therefore between the current vehicle and minimum velocity Vmin

$$V_{new} = min[(V_{min}, V_{curr} - 1)] \quad (2)$$

- Driving: When there is no other vehicle, the car chooses its veloc- ity between an interval from a minimum velocity to a maximal velocity given by the authority

$$V_{new} = [V_{min}, V_{max}] \quad (3)$$

- Random velocity: A random velocity can occur when there is an unexpected event on the road. The only unexpected events that can be considered here are an accident or when a pedestrian comes too

close to the first lane. Since those situations are discreet events. It's introduced with a binomial negation distribution

$f(k; n, p) = (k + n - 1 \quad k)^r p^n q^k$ (4) Where p is the probability that an accident occur and $q = 1 - p$,
k is the number of iteration and n is the number of succes

2.3.2. Lane Changing rules

The lane-changing rules are taken from the Nagel-Schrenberg model and state guidelines for vehicular agents when deciding to move from their current lane to an adjacent lane. The Lane changing occurs in two situations

- Discretionary Lane Changing: It occurs when the driver sees a queue far away and change lane to go faster
- Mandatory Lane Changing: This is when the driver has to change lanes because of an obstacle or to follow a given direction. Con- sidering the speed loss caused by the mandatory lane-changing behavior,

Before initiating a lane change, a car agent should consider several options:

- Necessity: A lane change might be triggered by a turn ahead, a slow-moving vehicle in the present lane, or a planned maneuver to reduce the agent's travel time.
- A lane change should be performed only when it is safe to do so. Safety checks should verify that there is enough space in the target lane and that the speeds of cars in both the current and target lanes are considered.

$$V_{new} = max(V_{new} - 1, 0) \quad (5)$$

2.4 Experimental Design

As shown in figure 1, the experiment is conducted on a road section not far from a traffic light. The road has 3 lanes and the priority is to keep right unless overtake. The structure of this intersection is such that in addition to the 4 lanes, there is a mini roundabout. Therefore the traffic static traffic timing is not optimal during the whole day. Throughout the simulation, the road network acts as a stage for the interactions between vehicles, traffic lights, and the surrounding environment. It provides the spatial context within which the multi-lane traffic flow and traffic light control strategies are evaluated

The cells grid is defined by a basic numpy matrix. The number of cars is generated using the initial density and the number of lanes. the initial car positions are given by the random function of numpy and a loop over the enumeration of car position fill the grid.

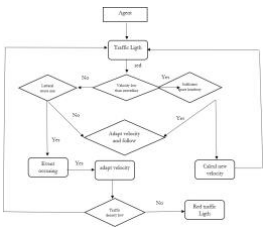
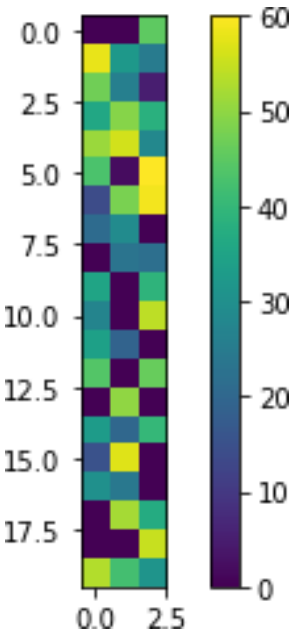


Figure 2: Agent Based Evacuation Model

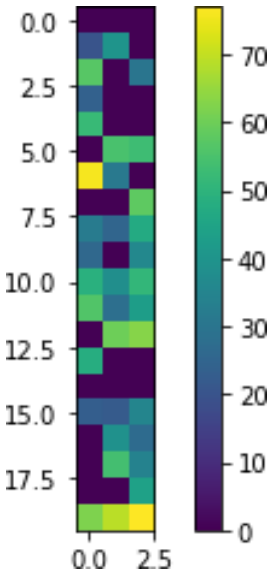
Basically, the most important part of the evacuation is in the search module. When the traffic light goes from red to green, the drivers initiate a movement from one cell to the other. However before he could move, he has to check if the desired next cell is empty.

3. Model validation

The model validation is conducted by checking an obvious assump- tion, the correlation between traffic density and road congestion.



(a) Simulation at time t0



(b) Simulation at the time the end

Figure 3: Simulation at the time the end

The simulations show that the traffic flow is smooth when vehicle density is low. In Figure 3a, there are fewer empty cells than in 3b. So the model could be used to improve by more realistic assumptions (lane changing, pedestrians, and buses) to check the impact of density, light timing, and lane number on the flow.

4. Results and discussion

4.1 Impact of Lane Number on Traffic Congestion

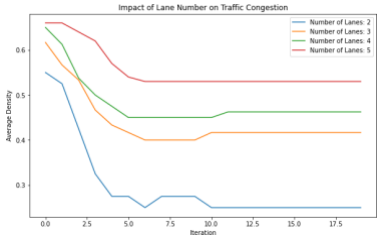
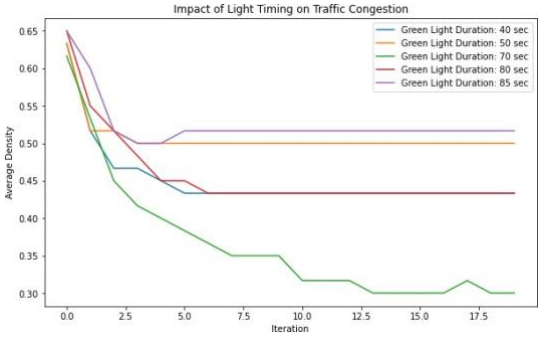


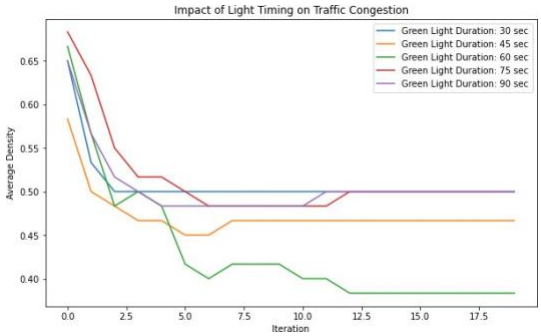
Figure 4: Impact of lane numbers on the traffic

Figure4 result on the variation of lane number with a constant density equal to 0.8. The plot show that the average density function has an inverse variation to the lane number. This is because more lane allow a huge number of vehicle to cross the junction. Since the result can help traffic regulator to increase the mobility experience, the model should take consideration of others perspectives like speed limit, number of interection and light timing. Moreover the result is consistent with Texas Transportation Institute who found that increasing the number of lanes from two to four can reduce traffic congestion by up to 20.

4.2 Impact of Traffic Light Timing on Traffic Congestion



(a) Light timing varies from 30 to 91 s



(b) Light timing varies from 40 to 70 s

Figure 5: Impact of Light timing vairiation in the flow

Figure 5 show the relationship between a green light duration and traffic congestion. It can be seen that increasing the ligth timing reduces the traffic congestion. However the rate improvement decrease over time as traffic reaches a more balanced state.

4.3 Impact of Pedestrian On the Traffic flow

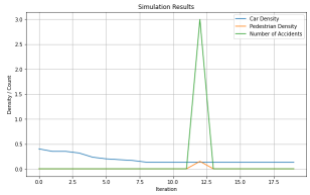


Figure 6: Impact of Pedestrian on The Flow

In Figure 6 are considered as sidewalker randomly placed. However this requires dirivers to slow down when approaching them. This results to a increasing probability of accident at each iteration. Acci- dents disrupt traffic flow and demonstrate the impact of unexpected events on congestion. By incorporating these assumptions, we are more comprehensive about the reaction of traffic on unexpected events.

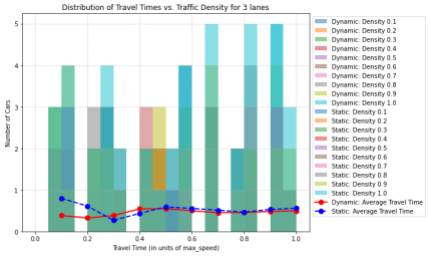


Figure 7: Distribution of Travel Times vs. Traffic Density for 3 lanes

4. Conclusion

This paper is a multi-agent-based model where agent rules are based on the cellular automata equation. The simulation shows that this pattern is a good representation of traffic congestion. However, the model could be improved with streaming data provided by the sensor whereby the density of the different agents would be more accurate. Furthermore the slightly difference between dynamic and static travel time can be due to the use of empirical data during simulation. However, this result may prompt road planners to consider implementing other technologies such as like fly-over or fully connected traffic light to reroute the traffic in case of emergency

Conflict of interest statement

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

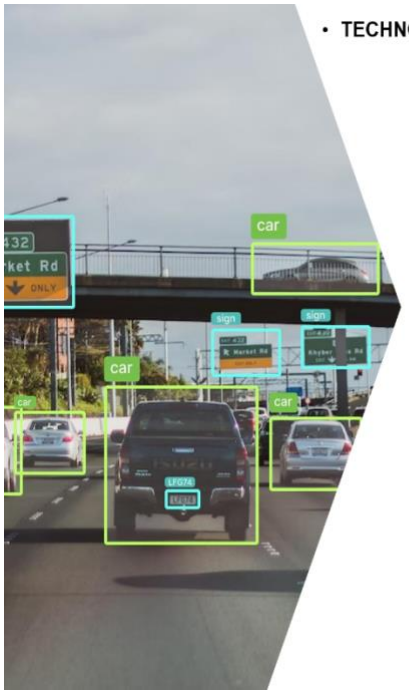
8.5. Sécurité et Smart city : optimisation de la traçabilité des véhicules volés

M. Tiorna COULIBALY, Compagnie Ivoirienne d'Electricité (CIE)

1. INTRODUCTION

(Source: https://www.planetoscope.com/Criminalite/1203-nombre-de-vols- de-voitures-dans-le-monde.html)

2. L'IA COMME SOLUTION



- TECHNOLOGIES DE RECONNAISSANCE AUTOMATIQUE DES PLAQUES D'IMMATRICULATION (ALPR)

ALPR : Technologie de traitement de l'image utilisée pour identifier les véhicules à partir de leurs plaques d'immatriculation.

Segmentation: Utilisée dans la reconnaissance des plaques pour isoler la plaque du reste de l'image

OCR: Technique qui permet d'extraire les caractères des images.

- **PRÉREQUIS POUR DES ALPR DANS LA TRAÇABILITÉ DES VÉHICULES VOLÉS**

01

Infrastructures nécessaires

- Caméra : Pour prendre les photos avant et arrière
- Illumination : Une lumière pour éclairer les plaques de nuit comme de jour et qui est invisible pour le chauffeur. Ex: Infrarouge

02

Le module d'ALPR

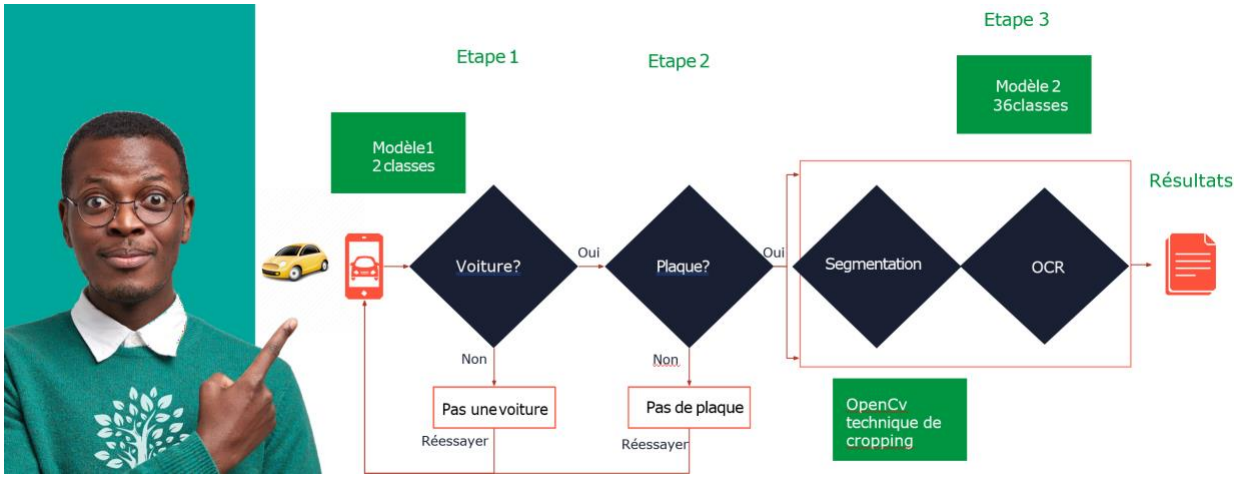
- Segmentation
- OCR

03

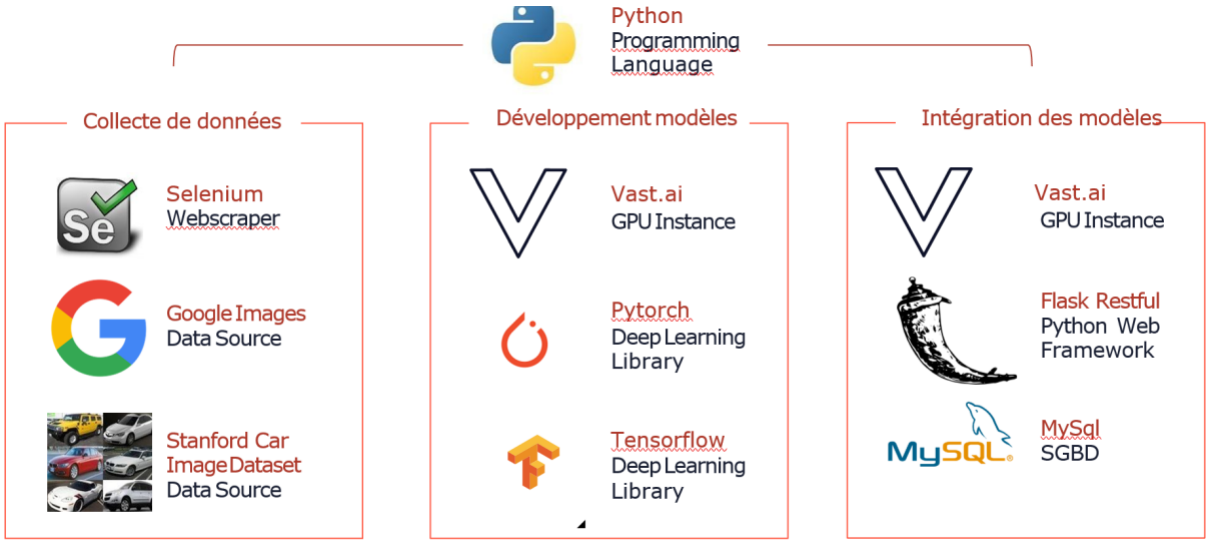
Des bases de données

- Base de données centralisée des déclarations et des informations des plaques
- Base de données de stockage des résultats du traitement du ALPR

- **PIPELINE DES MODÈLES**



- **OUTILS ET FRAMEWORK**



3. LIMITATIONS ACTUELLES ET PERSPECTIVES D'AMELIORATION

- **DÉFIS ET CONTRAINTES**

01

Aspects juridiques et éthiques

- Protection de la vie privée
- Stockage sécurisé des données
- Utilisation responsable des informations

02

Coût de déploiement

- Coût des caméras et des installations
- Gestion du stockage (disque dur et SGBD)
- Coût des GPU

03

Modèles de plaques variées

- Tenir compte des différents modèles de plaque

04

Conditions météorologiques

- Impact sur la qualité des images
- Techniques de traitement supplémentaires nécessaires
- Risque de mauvaise précision des modèles

- **PERSPECTIVES D'AMÉLIORATION**

Repenser le système ALPR

- Possible d'utiliser des modèles existants via API
- Amélioration constante du pipeline mis en place
- S'assurer de la vitesse de traitement

Base de données centralisée

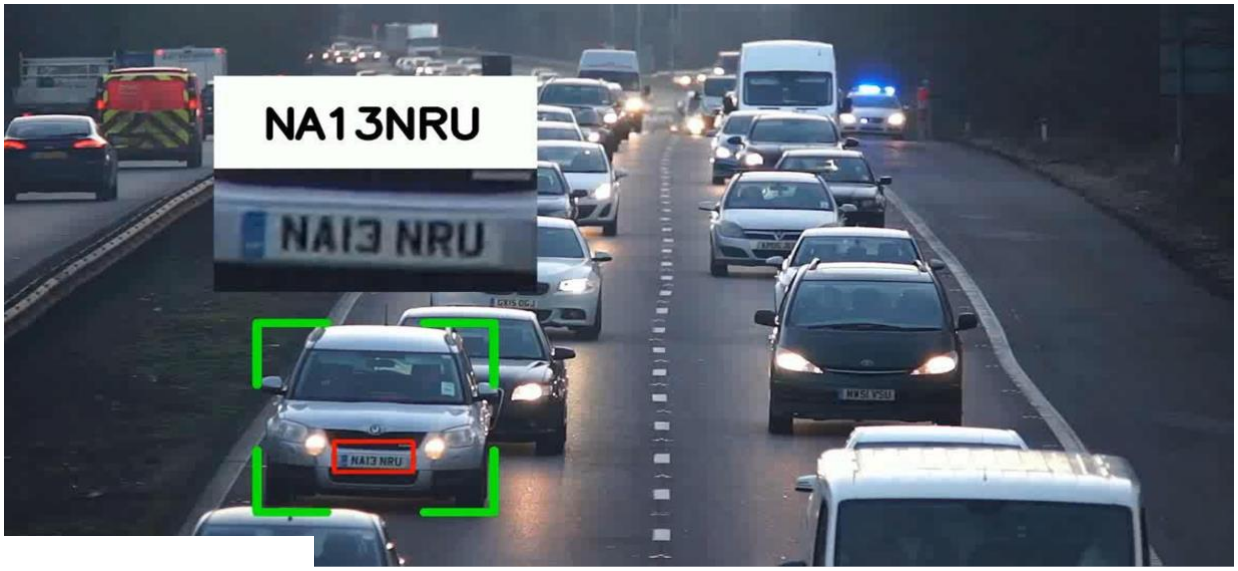
- Trouver un moyen pour centraliser les déclarations de vol
- Fiabiliser les données

Amélioration de la capacité de traitement des systèmes ALPR

- Améliorer la vitesse de traitement
- Améliorer le nombre de détections simultanées

Meilleurs équipements

- Achat de camera de dernière technologie



4. CONCLUSION

8.6. Architecture des Villes Intelligentes

Mamadou Diarra, Ballo Abou Bakary, Ayikpa Kacoutchy Jean, Adou Kablan Jérôme
Université Félix Houphouët-Boigny (UFHB), Unité de Formation et de Recherche Mathématiques et Informatique (UFR-MI)

1. Introduction

Nombreuses définitions des villes intelligentes: smart, intelligent, numérique.

Tableau 1. Définitions de villes intelligentes [1]

Termes	Définitions
Intelligent City Ville intelligente	Les villes intelligentes s'efforcent délibérément d'utiliser les technologies de l'information pour transformer la vie et le travail. Le terme "intelligente" implique la capacité de soutenir l'apprentissage, le développement technologique et l'innovation dans les villes. En ce sens, toute ville numérique n'est pas nécessairement intelligente, mais toute ville intelligente a une composante numérique, même si la composante "population" n'est toujours pas incluse dans une ville intelligente, comme c'est le cas dans une ville intelligente
Ubiquitous City Ville omniprésente	Une extension du concept de ville numérique dans les tems de large accessibilité. Elle met l'informatique omniprésente à la disposition des éléments urbains partout. Elle se caractérise par la création d'un environnement dans lequel tout citoyen peut obtenir n'importe quel service, n'importe où et n'importe quand, à l'aide de n'importe quel appareil. La ville ubiquitaire est différente de la ville virtuelle car, alors que la ville virtuelle reproduit des éléments urbains en les visualisant dans un espace virtuel, la ville ubiquitaire est créée par l'inclusion de puces informatiques ou de capteurs dans les éléments urbains.
Digital city Ville numérique	Une communauté connectée qui combine une infrastructure de communication à large bande pour répondre aux besoins des gouvernements, des citoyens et des entreprises. L'objectif final d'une ville numérique est de créer un environnement propice au partage d'informations, à la collaboration, à l'interopérabilité et à des expériences transparentes partout dans la ville.
Virtual city Ville virtuelle	La ville devient un concept hybride composé d'une réalité, avec ses entités physiques et ses habitants réels, et d'une ville virtuelle parallèle de contreparties, un cyberespace.

2. Villes intelligentes

- Explosion du nombre d'habitants au sein des villes — en 2050, 70 % de la population sera urbaine [7] — a conduit les administrations à réfléchir sur leur optimisation et leur viabilité à long terme
- Caractéristiques :
- Hétérogène : les systèmes basés sur l'IoT, l'hétérogénéité pour permettre à des systèmes indépendants, distribués, stockés d'être utilisés par différents utilisateurs ;
- Contrainte de ressources: mémoire, vitesse de traitement , stockage, etc.;
- Mobilité :la mobilité dans les villes intelligentes est personnalisée grâce à une infrastructure de communication bien développée;
- Connectivité: La connectivité permet à tout appareil de se connecter au monde

intelligent:

- Implication de l'utilisateur: Les facteurs humains (apprentissage, créativité et éducation) sont également essentiels au développement des villes intelligentes [6].

3. Problématique

Objectif :

Le concept de ville intelligente découle de la nécessité de gérer plusieurs problèmes causés par la croissance démographique effrénée dans les centres urbains.

L'objectif est d'améliorer la gestion des villes par l'optimisation de toutes les ressources urbaines en vue proposer une mielleure qualité de vie des citoyens.

Contexte :

Améliorer la gestion des villes en termes de ressources humaines, naturelles et infrastructures au moyen d'outil d'aide à la décision utilisant des interfaces ergonomiques,

Utiliser des TIC dans les composantes de la vie culturelle, socio-professionnelle: la gouvernance, domotique, bâti, transport, éducation, santé, information, éducation, agriculture, environnement, la sécurité, etc,

Outils :

- □ Infrastructures TIC: Télécoms, IoT, Internet, Logiciels
- □ Intelligence artificielle :outils de gestion et de décision
- □ Jumeaux numériques: la réalité virtuelle et la réalité augmentée pour faciliter la planification des villes intelligentes
- □ Robotique
- □ Normes : IUT, ISO

4. Paradigme de l'Intelligence Artificielle

4.1.Intelligence artificielle

L'intelligence artificielle (IA) consiste à mettre en œuvre un certain nombre de techniques visant à permettre aux machines d'imiter une forme d'intelligence réelle.

L'objectif de l'IA est de permettre à des ordinateurs de penser et d'agir comme des êtres humains.

4.2.Apprentissage Automatique ou Machine Learning:

Le Machine Learning (ML) est une technologie d'intelligence artificielle permettant aux ordinateurs d'apprendre sans avoir été programmés explicitement à cet effet.

ML s'appuie sur des algorithmes (principalement statistiques) pour permettre à une machine « d'apprendre » à partir d'un certain nombre de réponses correctes disponibles connues au départ (échantillon ou base d'apprentissage).

4.3.Apprentissage profond ou Deep Learning

Le Deep Learning ou apprentissage profond est un type d'intelligence artificielle dérivé du machine Learning (apprentissage automatique) où la machine est capable d'apprendre par elle-même, contrairement à la programmation où elle se contente d'exécuter à la lettre des règles prédéterminées.

Le Deep Learning s'appuie sur un réseau de neurones artificiels s'inspirant du cerveau humain

5. Apprentissage profond et villes intelligentes

L'apprentissage profond est utilisé pour construire des systèmes intelligents dans :

- la modélisation urbaine des villes intelligentes
- L'élaboration d'infrastructures intelligentes des villes intelligentes: urbanisation, routes, bâtiments, parking, véhicule, communication
- Apprentissage profond pour la mobilité et le transport intelligents
- la gouvernance urbaine intelligente: administration, énergie, déchets, environnement, qualité de l'air
- la résilience et la durabilité des villes intelligentes : anticipation des risques
- l'éducation intelligente
- les solutions de santé intelligente
- la sécurité et la confidentialité des villes intelligentes
- TIC: IoT, data center, Cloud, Blockchain

- Les extensions de l'environnement: économie, citoyens, agriculture, espaces verts, sure, élevage
- l'accès aux données ouvertes ou open data

6. Proposition d'architecture

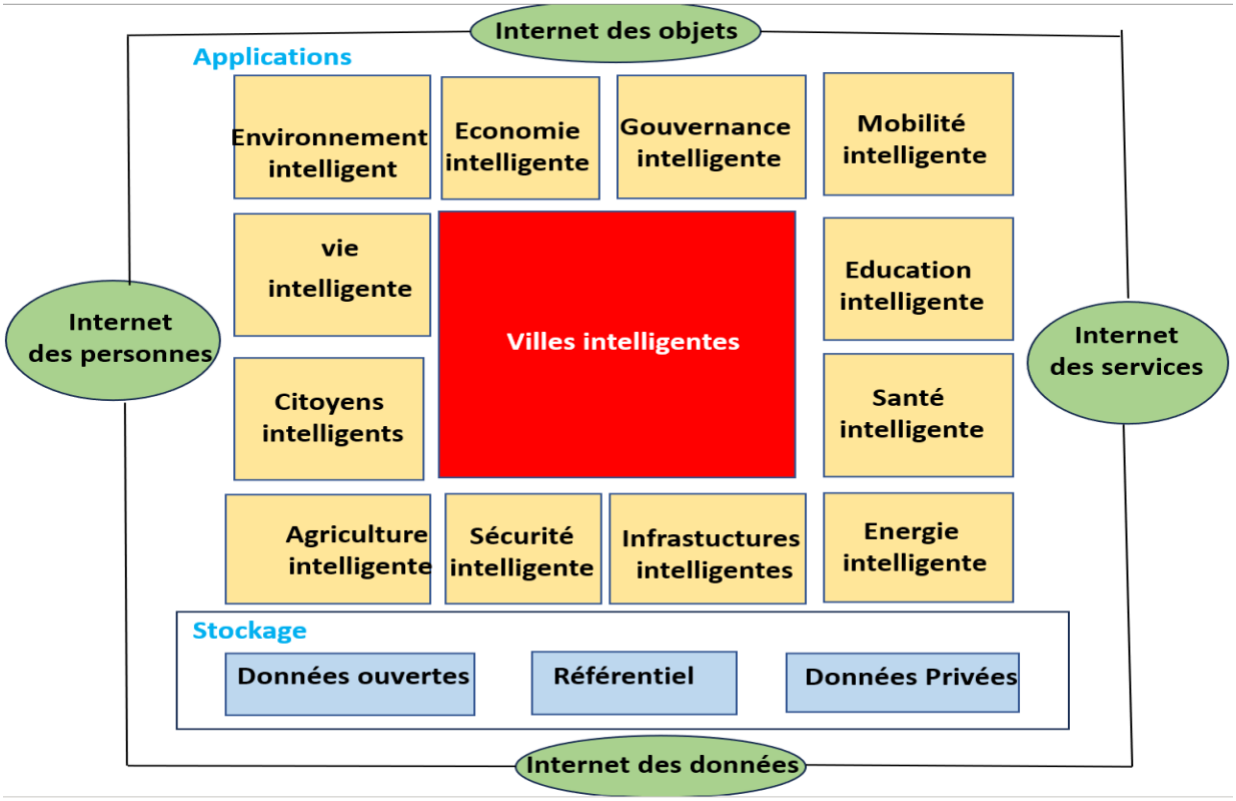


Figure 1. Architecture proposée pour villes intelligentes

7. Application aux véhicules: Internet of Vehicles (IoV)

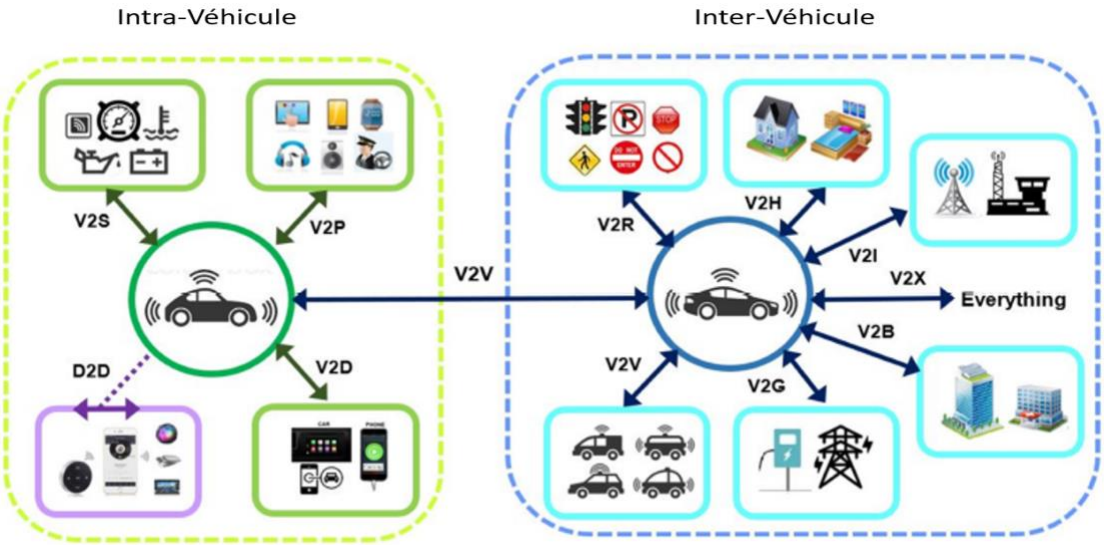


Figure 2. Modèles d'interaction de l'Internet des Véhicules [2]

8. Application à l'éducation

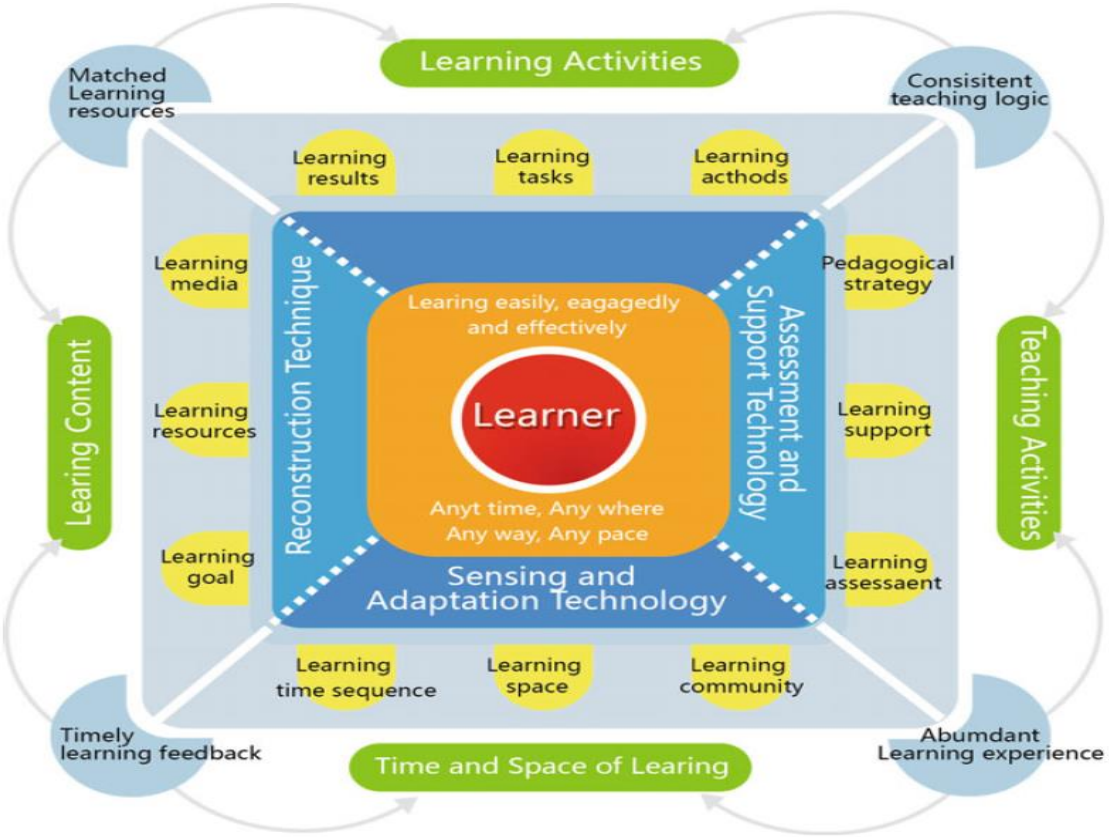


Figure 3. Cadre de l'apprentissage intelligent [3]

9. Défis des villes intelligentes

Comme défis, posons-nous quelques questions :

- Quelles sont les technologies habilitantes utilisées dans les plates-formes logicielles de pointe pour les villes intelligentes ?
- Quelles sont les exigences auxquelles doit répondre une plateforme logicielle pour les villes intelligentes ?
- Quels sont les principaux défis et les problèmes de recherche ouverts dans le développement de plates-formes logicielles robustes de prochaine génération pour les villes intelligentes ?
- Quels sont les besoins stratégiques pour l'amélioration de la qualité de vie pour une population qui croit rapidement dans nos pays ?
- Infrastructure informatique : infrastructures adaptées a chaque zone de services publics ou privés
- Coût : Réduction des coûts pour les consommateurs et avantages sociaux; réduction de coût pour les investisseurs
- Environnement hétérogène /Interopérabilité : service orienté architecture de façon à garantir un environnement hétérogène et une interopérabilité entres différents sites ou équipements
- Disponibilité et évolutivité : Services disponibles et qui s'adaptent aux changements de technologies
- Sécurité : sécurité des services déployés
- Protection de la vie privée : protection de la vie privée, des données personnelles
- Efficacité : Modèle d'informatique des nuages hiérarchique (Cloud) pour les applications, les infrastructures et le stockage
- Gestion des Big Data : gestion des gros volumes de données avec les outils appropriés tels que le machine Learning, le Deep Learning et produire une visualisation des tableaux de bords pour les prises de décision
- Adaptation sociale et développement d'applications : Modèle de données ouvert (Open data) et App Store de la ville intelligente.

10.Limites des villes intelligentes

Tout système présente des insuffisances [5] :

- les budgets élevés des investissements nécessaires aux nouvelles infrastructures ;
- le risque de recrudescence de la cybercriminalité, en raison de l'essor de la numérisation
- les possibles dérives d'un contrôle omniprésent des habitants ; des coûts à la charge financière des citadins, comme les compteurs électriques intelligents;
- La non-implication suffisante des citoyens ;
- l'existence de la ville "froide" ou sur-technologique (cameras) à l'insu de tous, de sa culture ;
- la sous-estimation de la complexification de la gestion des multiples réseaux urbains;
- Le manque d'anticipations des usages non convenables des big data collectées (dérive vers la surveillance de masse) ;
- dépendance à la technologie; l'impact direct des TIC ;
- l'information ne suffit pas à changer les comportements.

11.Quelques mesures de sécurité des villes intelligentes

- L'intégration de la surveillance de la sécurité : acquérir et d'analyser les données au fur et à mesure que les menaces sont détectées, et d'isoler les dispositifs touchés;
- Exigences de sécurité propres à l'IoT des villes intelligentes :
    - Prévention des fuites et des falsifications de données
    - Prévention et détection de la falsification des appareils
- Fonctions de sécurité requises pour la ville intelligente :
    - Plateforme de distribution de données sécurisée empêchant les fuites/falsifications de données
    - Sécurité des communications : Cryptage authentifié léger,

chiffrement des données;

- Technologie de détection de l'altération pour prévenir/découvrir l'altération des dispositifs

- Sécurité dans la coopération interurbaine:
  - Authentification fédérée
  - Protection des informations transmises par d'autres systèmes
  - Réponses aux incidents entre systèmes

12. Quelques solutions pour les villes intelligentes

Tableau 2. Résumé de quelques solutions smart [4]

Solutions	Caractéristiques
Green Buildings and Roofwtop Wind Turbines	Bâtiments écologiques et éoliennes de toit
Smart Traffic Control and Real Time	Contrôle intelligent du trafic et mises à jour en temps réel
Smart Automatic Parking	Stationnement automatique intelligent
Full Wi-Fi Coverage	Couverture Wi-Fi complète
Electric Transport and Vehicles	Transport et véhicules électriques
Smart Street Lighting	Éclairage public intelligent
Waste Management	Gestion des déchets
Automatic Leakage Management	Gestion automatique des fuites
Building Integrated Photovoltaics	Photovoltaïque intégré au bâtiment
Perimeter Access control	Contrôle d'accès au périmètre
Data driven traffic management air quality	Gestion du trafic basée sur les données Qualité de l'air

13. Management de projet de villes intelligentes

- **Fondamentaux** : Administration saine, garantie d'inclusion de la ville (fracture numérique ou sociale)
- **Infrastructures** :

- Les réseaux : gaz, eau, électricité et télécommunications ;
- Les infrastructures publiques : domotique, bâtiments, mobiliers urbains ;
- Les transports : les routes, les transports publics, les voitures intelligentes, le covoiturage et les mobilités dites douces (à vélo, à pied) ;

- Les e-services et e-administrations…

➢ **Critères**

- Smart Economy : Esprit innovateur ; Esprit d'entrepreneuriat ; Image de la ville ; Productivité ; Marché du travail ; Intégration internationale ;
- mart people : Education ; Apprentissage à long terme ; Pluralité ethnique ; Ouverture d'esprit ;.
- Smart environnement : Qualité de l'air ; Conscience écologique ; Gestion durable des ressources ;
- Smart Gouvernance : Conscience politique ; Services publics et sociaux ; Administration efficace et transparente ;
- Smart mobility : Système de transport local ; Accessibilité (inter-) nationale ; Infrastructure des technologies de l'information et des communications ; Durabilité du système de transport ;
- Smart living : Centre culturel et de loisirs; Conditions de santé ; Sécurité individuelle ; Qualité des logements ; Installations éducatives ; Attractivité touristique ; Cohésion sociale.,

➢ **Quelques modèles de smart cities** : Singapour, Dubai, Copenhague, Boston, Oslo, Londres, New York, Barcelone,
Amsterdam, Hong Kong,

14.Conclusion

Notre étude a montré :

- □ La nécessité de mer un projet de ville intelligente en se basant sur l'existant en TIC, infrastructures, gouvernance, énergie, etc,
- □ L'intérêt des TIC notamment de l'Intelligence artificielle comme enjeu de développement.

Travaux scientifiques :

Articles utilisant IA : agriculture (café, cacao, mangue, environnement), biométrie(empreintes digitales, visages), santé (cancer), SIG.

Recommandations :

- □ Vulgarisation des TIC
- □ Formation des experts en intelligence artificielle pour développer des solutions intelligentes

Collaboration :

Milieu de la recherche et les décideurs

Bibliographie :

1. ALBINO, Vito, BERARDI, Umberto, et DANGELICO, Rosa Maria. Smart cities: Definitions, dimensions, performance, and initiatives. Journal of urban technology, 2015, vol. 22, no 1, p. 3-21.:
2. ANG, Li-Minn, SENG, Kah Phooi, IJEMARU, Gerald K., et al. Deployment of IoV for smart cities: Applications, architecture, and challenges. IEEE access, 2018, vol. 7, p. 6473-6492. [IoV]
3. LIU, Dejian, HUANG, Ronghuai, WOSINSKI, Marek, et al. Characteristics and framework of smart learning. Smart learning in smart cities, 2017, p. 31-48. Learning]
4. https://parklio.com/en/blog/top-10-smart-solutions-for-the-cities-of-future |Solution]
5. RINGENSON, Tina, ERIKSSON, Elina, BÖRJESSON RIVERA, Miriam, et al. The limits of the smart sustainable city. In : Proceedings of the 2017 Workshop on Computing within Limits. 2017. p. 3-9. [limits]
6. T. Nam and T. A. Pardo, ``Conceptualizing smart city with dimensions of technology, people, and institutions," in Proc. 12th Annu. Int. Digit. Government Res. Conf., Digit. Government Innov. Challenging Times, 2011, pp. 282291,
7. « World Urbanization Prospects », 2014 : https://esa.un.org/unpd/wup/publications/files/wup2014-highlights.Pdf [1]

8.7. Nouvelle approche de sélection des valeurs optimales d'hyperparamètre

Kopoin ND. Charlemagne[1], Koffi Dagou[1] and Zouneme Boris[2]

[1] Ecole Superieure Africaine des TIC, Abidjan, CI
[2] Université Nangui Abrogoua, Abidjan, CI
Correspondent author mail : charlemagnekopoin@gmail.com

Abstract. Le problème que nous traitons dans cet article est un problème de sélection de modèle. Nous examinons la technique de validation croisée *k*-fois (KCV), appliquée à l'algorithme ede classification des machines à vecteur de support (SVM) de type gaussien. Dans le processus de la validation croisée, la valeur de *k* du nombre de sous ensemble est choisie et fixée de manière aprioristique (sans aucune expérience). Cependant, la valeur de

k agit sur le choix du meilleur compromis entre l'erreur d'estimation et l'erreur d'approximation du modèle. Ainsi, la valeur *k* du nombre de sous-ensembles peut influencer sévèrement les valeurs optimales d'hyperparamètres du classificateur SVM et par conséquent agir sur la performance du modèle sélectionné et sa capacité de généralisation.

Dans ce travail, nous proposons une approche rigoureuse de recherche d'hyperparamètres du SVM noté SVOH (Sélection de Valeurs Optimales d'Hyperparamètres). L'approche proposée considère la valeur *k* du nombre de sous-ensembles comme un paramètre influent du modèle et fait donc l'apprentissage pour retrouver une valeur optimale de *k*.

Keywords: apprentissage automatique, validation croisée, sélection de modèle

1. Introduction

La machine à vecteur de support (SVM) est l'une des techniques de pointe pour les tâches de classification. Elle appartient au domaine des réseaux neuronaux artificiels (ANN) [1] mais se caractérise par les fondements solides de la théorie de l'apprentissage statistique. L'apprentissage des SVM s'effectue par la recherche d'un ensemble de paramètres obtenus par la résolution d'un problème de programmation quadratique convexe sous contrainte (CCQP), pour lequel de nombreuses techniques efficaces ont été développées. La recherche des paramètres optimaux n'achève cependant pas le processus d'apprentissage, car il existe un ensemble de variables supplémentaires, les hyperparamètres, qui doivent être réglées pour atteindre la performance de classification optimale, comme pour les ANN, où l'hyperparamètre est le nombre de nœuds cachés. Pour le cadre des SVM gaussien, il s'agit des paramètre régulateur C et γ. Ce réglage n'est pas trivial et constitue un problème de recherche ouvert [2]–[5]. Le processus de recherche des meilleurs hyperparamètres est généralement appelé phase de sélection du modèle dans la littérature sur l'apprentissage automatique et est strictement lié à l'évaluation de la capacité de généralisation du SVM ou, en d'autres termes, du taux d'erreur que le SVM peut atteindre sur de nouvelles données (données inconnues). En fait, il est courant de sélectionner le SVM optimal (c'est-à-dire les hyperparamètres optimaux) en choisissant celui dont l'erreur de généralisation est la plus faible. Les méthodes permettant de réaliser la phase de sélection des modèles peuvent être divisées en deux catégories [6]. Les méthodes théoriques [7] fournissent des informations approfondies sur les algorithmes de classification, mais sont souvent inapplicables et incalculables pour être d'une quelconque utilité pratique. D'autre part, comme le mentionnent [2], les praticiens ont trouvé plusieurs procédures, qui fonctionnent bien dans la pratique mais n'offrent aucune garantie théorique sur l'erreur de généralisation. Certaines d'entre elles reposent sur des techniques de rééchantillonnage [8].

L'une des techniques de rééchantillonnage les plus populaires est la procédure de validation croisée *k*-Fold (KCV) [4], qui est simple, efficace et fiable. La technique KCV consiste à diviser un ensemble de données en *k* sous-ensembles indépendants : à leur tour, tous ces sous-ensembles sauf un sont utilisés pour former un classifieur, tandis que le sous-ensemble restant

est utilisé pour évaluer l'erreur de généralisation. Après la formation, il est possible de calculer une limite supérieure de l'erreur de généralisation pour chacun des classificateurs formés.

Le choix d'une valeur fixe de sous-ensembles pour la validation croisée peut produire un modèle avec un biais et une variance élevée [6]. La validation croisée fait en effet la moyenne de plusieurs estimateurs du risque de retenue correspondant à différents fractionnements de données. Dans [2] nous pouvons vérifier que la valeur *k* influence la stabilité de la moyenne des erreurs. Toujours selon [6], les performances de sélection de modèles avec la validation croisée sont généralement optimales lorsque la variance est aussi faible que possible. Cette variance diminue généralement lorsque le nombre *k* de sous-ensembles augmente, avec une taille d'échantillon d'entraînement fixe *n*. Lorsque *k* est fixe, la variance de la validation croisée dépend également de n. en effet, dans [4], nous pouvons voir que la valeur de γ dépend fortement de l'ensemble d'apprentissage utilisé. Le choix de *k* influe donc sur la variance de l'estimateur de la validation croisée et selon [3], [9], elle peut avoir un impact significatif sur la recherche des valeurs optimales d'hyperparamètres.

2. Approche proposée

2.1.Problème à résoudre

Pour ne pas fixer la valeur de *k* dans le cadre de la validation croisée *k*-fois, nous proposons d'utiliser la procédure qui consiste à considérer un certain nombre de combinaisons possibles de sous-ensembles dans lesquels l'ensemble d'apprentissage original peut être divisé. L'objectif est de choisir une meilleure procédure d'estimation de la validation croisée, celle qui présente le biais et la variance les plus faibles, permettant d'identifier une bonne combinaison d'hyperparamètres (C,γ) afin que le classificateur SVM puisse présenter une faible erreur de généralisation et prédire des données inconnues avec un taux de justesse supérieur.
Pour l'approche proposée, nous considérerons le nombre *k* en tant qu'hyperparamètre comme dans [3], qui peut prendre n'importe quelle valeur dans l'ensemble $k \in \{i,...,10\}$, $i \geq 3$. La plus petite valeur de *k* est fixée à 3 car pour chaque sous-ensemble, les données d'entrainement doivent être supérieures à 60% de l'ensemble d'apprentissage comme le montre [10]. Ici, nous fixons la plus grande valeur de k à 10 pour rester dans la marge fixée par les méthodes empiriques. Ce choix limité des valeurs test de k à 10 permet également, dans les cas où l'ensemble d'apprentissage est large, d'éviter que la technique soit beaucoup gourmande en calcul. En effet, en supposant qu'il y a q paramètres, et que chacun d'entre eux a m valeurs distinctes, sa complexité de calcul augmente exponentiellement à un taux de $O(m^q)$ [11], [12]. De plus, dans [3], nous pouvons voir que plus de 10 bases de données différentes ont produit une valeur optimale k inférieure à 10. Le nombre de paramètres à optimiser pour notre cas devient donc le triplé (C,γ,k), étant donné que notre fonction de décision f utilise un noyau gaussien qui lui-même fonctionne avec le couple (C,γ).

2.2. Fonctionnement de l'algorithme SVOH

Soient {C} et {γ} correspondant respectivement à l'ensemble des valeurs pour le paramètre C et l'ensemble des valeurs pour le paramètre γ. Posons D_Z notre ensemble d'apprentissage de Z observations et f notre modèle de SVM obtenu avec les hyperparamètres (C, γ), D_{ZE}, les Z(k-1)/Z sous-ensembles de l'ensemble d'apprentissage après subdivision en k-sous-ensembles et D_{ZS} , les Z/k sous-ensemble restant réservé pour le test après subdivision. L'algorithme prend en entrée D_Z, {C} et {γ}. Pour chaque k subdivision (k∈{3,10}) de l'ensemble d'apprentissage D_Z,, l'algorithme entraine un classifieur f à l'aide des valeurs de {C} et {γ} sur D_{ZE}, puis évalue sur D_{ZS} le taux de justesse de f. Pour finir, l'algorithme sélectionne le meilleur triplé (C,γ,k) ayant donné un taux de justesse supérieur. Nous présentons ci-dessous un pseudo-code de l'algorithme SVOH.

Algorithme SVOH

Entrées : D_Z : ensemble d'apprentissage
{C}: ensemble des valeurs pour C,
{γ}: ensemble des valeurs pour γ
Sorties : {k*, C*, γ*}
1 : **pour tout** C ∈ {C}, γ ∈ {γ}, k ∈{3,10} **faire** :
2 : f=∅
3 : D_{ZE}, D_{ZS}= subdivision (D_Z, k)
4 : f_E= SVM (D_{ZE}, C, γ)
5 : E_r = Evaluer le taux de justesse (f_E, D_{ZS})
6 : f = f ∪{ E_r }
7 : **fin pour**
8 : {k*, C*, γ* }= le meilleur taux de justesse de f
9 : **Retourner** {k*,C*,γ* }

3. Materiels et méthodes

3.1.Données d'apprentissage

Dans ce travail, les données utilisées pour l'expérimentation proviennent des travaux de kopoin et al. [13]. Kopoin et al. Ont utulisé une technique d'extraction de caractéristiques appelée *BP* (Bigram physicochemical) pour produire des données numériques à partir de trois ensembles de données de référence d'interaction protéine-protéine (IPP). L'IPP correspond au fait que deux protéines interagissent ou non. Dans le cas de l'interaction on parle de IPP positive et dans le cas contraire, IPP négativve. En premier, les données IPP HPRD [14] constituées d'un ensemble de 10000 échantillons repartis en 5000 paires IPP positives et 5000 paires IPP négatives comme dans [15]–[17]. Les ensembles de données IPP S. Cerevisiae [18] constitués de 11188 échantillons (avec 5594 paires positives et 5594 paires négatives) et IPP H. Pylori [19] constitué de 2496 échantillons (1458 paires positives et 1458 paires négatives) ont également été utilisés.

3.2.Classifieur SVM

La phase de prédiction des IPP qui passe tout d'abord par un SVM optimal est obtenu en

sélectionnant les hyperparamètres optimaux, c'est-à-dire ceux qui permettent au SVM de présenter l'erreur de généralisation la plus faible.
Considérons un ensemble d'apprentissage $\mathcal{Z} = \{(x_i, y_i), i \in [1, n]\}$ où à chaque vecteur $x \in \mathbb{R}^p$ est associé une valeur $y \in \{-1, +1\}$. La relation entre $\mathcal{X}$ et $\mathcal{Y}$ est encapsulée dans une distribution inconnue $P(\mathcal{X}, \mathcal{Y})$, qui est à l'origine des données. Le but de l'apprentissage est de trouver une fonction $f: \mathbb{R}^d \to \mathcal{Y}_f \subset \mathbb{R}$ qui se rapproche de cette relation. L'algorithme SVM [20] peut être exploité à cette fin, où le classificateur est identifié pendant la phase de recherche des hyperparamètres en résolvant le problème quadratique convexe suivant :

$$\begin{aligned} Maximiser \quad & \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j \cdot y_i y_j \cdot h(x, x_j) \\ sous\ réserve\ de \quad & 0 \le \alpha_i \le \mathcal{C}, \quad i = 1, .. n \\ & \sum_{i=1}^{n} \alpha_i y_i = 0 \end{aligned}$$

où les α_i sont les multiplicateurs de Lagrange et $\mathcal{C}$ un des hyperparamètres, qui contrôle le compromis entre la marge et l'erreur de mauvais classement et $h(x_i, x_j)$, la fonction noyau. Le noyau considéré ici est le noyau gaussien. Le noyau gaussien est dérivé de la fonction RBF (Radial Basic Function) et dépend de la distance euclidienne entre les vecteurs x_i, x_j dans l'espace de départ. Il est défini de la manière suivante :

$$h(x_i, x_j) = \exp\left(\frac{\| x_i - x_j \|^2}{2\gamma^2}\right)$$

avec γ un hyperparamètre additionnel qui détermine l'étendue de l'influence d'un seul exemple d'entraînement [15]. La résolution du problème quadratique convexe permet d'avoir un classifieur comme défini à l'équation 2-1 du chapitre 2 de la manière suivante :

$$f(x) = \sum_{i=1}^{n} \alpha_i y_i h(x_i, x_j) + b$$

où b représente le biais.

Les deux hyperparamètres $\mathcal{C}$ et γ sont donc les paramètres influents du classifieur SVM lui permettant d'estimer l'erreur de généralisation.

4. Resultas

Les ensembles de données IPP HPRD, S. Cerevisiae et H. Pylori, ont servi de données d'apprentissage pendant que les deux autres ont servi de tests.

4.1. Métriques d'évaluation utilisées

Pour évaluer l'approche proposée, nous avons utilisé les métriques généralement utilisées pour mesurer la performance d'un classifieur. Nous avons utilisé le taux de justesse (taux de bonne prédiction), la précision, la sensibilité et la courbe ROC [21].

4.2.Resultats d'entrainement

L'entrainement a été mené sur les données HPRD et a consisté à rechercher à l'aide de l'algorithme SVOH une valeur de k*, C*,γ* parmi une grille de valeurs potentielles renseignées dans la table 1. Nous nous sommes servis du taux de justesse comme métrique d'évaluation de performance pour retrouver les valeurs optimales d'hyperparamètres. La capacité de généralisation du modèle formé est évaluée sur les ensembles de données IPP S. Cerevisiae et H. Pylori.

Table 5: Rangée de valeurs d'hyperparamètres

Hyperparameter	Grid values
C	{1;3;10;32;50;100}
γ	{10^(-4),10^(-3),10^(-2),10^(-1),1}
k	{3,4,5,6,7,8,9,10}

L'application de l'algorithme SVOH a donné les valeurs optimales d'hyperparamètres suivantes : (C,γ,k)=(32;0.01;7). La table 2 présente les meilleures valeurs de performances du taux de justesse moyen pour différentes combinaisons du triplé (C,γ,k).

Table 6: Résultats du taux de justesse de de l'application de l'approche SVOH

k	(C^*, γ^*)	Justesse **(%)**
3	(10 ; 0.1)	91.92
4	(50 ; 0.01)	92.36
5	(100 ; 0.001)	92.70
6	(10 ; 0.001)	91.13
7	**(32 ; 0.01)**	**93.69**
8	(32 ; 0.001)	92.36
9	(100 ; 0.1)	92.21
10	(100 ; 0.01)	92.49

Les résultats de la table 2 montrent que pour des valeurs de $k \in \{3;4;5;6\}$, les taux de justesse se situent entre 92% et 93%. À partir de $k = 7$, les taux de justesse sont nettement supérieurs et se situent entre 92% et 94%. Dans l'ensemble, les taux de justesse sont sensiblement égaux, cependant, pour un nombre $k = \{5 ; 7 ; 10\}$ où les valeurs 5 et 10 représentent les valeurs à priori, le modèle formé avec un nombre $k = 7$ de sous-ensembles obtient le meilleur score de justesse avec 93.69% contre 92.70% pour $k = 5$ et 92.49% pour $k = 10$. Ces premiers résultats démontrent que les meilleures performances du modèle SVM sont obtenues sur le triplé (k,C, γ)=(7, 32, 0.01).

Dans la table 3, nous comparons sur les autres métriques précision, sensibilité et AUC, les scores obtenus pour les valeurs de subdivision $k = 7$ déterminés par l'approche SVOH contre ceux obtenus pour les valeurs $k = \{5;10\}$ qui sont les valeurs généralement appliquées.

Table 7: Résultats obtenus sur les autres métriques

k	Précision (%)	Sensibilité (%)	AUC (%)
5	92.90	92.15	96.36
7	**94.09**	**93.15**	**97.88**

10	92.87	92.67	95.58

Les scores obtenus pour des valeurs à priori du nombre de sous-ensembles sont sensiblement les mêmes dans toutes les métriques. Pour une subdivision k= 5 de l'ensemble d'apprentissage, nous obtenons comme valeurs d'hyperparamètres (C,γ)=(100;0.001). Les scores obtenus dans les métriques précision, sensibilité et AUC sont respectivement 92.90% ; 92.15% et 96.36%. Pour une subdivision k=10, nous obtenons comme valeurs d'hyperparamètres (C,γ)=(10;0.01). Les scores obtenus dans les différentes métriques sont respectivement 92.87% ;92.67% et 95.38%. Par-contre, les taux obtenus pour une subdivision k=7 avec pour valeurs d'hyperparamètres (C,γ)=(32;0.01) sont respectivement 94.09% ; 93.15% et 97.88%. En outre, bien vrai que l'écart entre les différents taux ne soient pas très grand, nous retenons tout de même qu'une subdivision de l'ensemble d'apprentissage en 7 sous-ensembles améliore le taux de justesse d'environ 1% par rapport aux taux obtenus avec les valeurs de subdivision à priori (voir tableau 4-3). Aussi, nous constatons également un meilleur score dans les métriques précision et sensibilité avec une performance moyenne supérieure à 0.7% que celles obtenues par les valeurs à priori. Les résultats montrent que les meilleurs taux sont obtenus avec une subdivision k = 7, c'est-à-dire celle déterminée par l'approche SVOH.

5. Discussions

La principale technique utilisée dans cette étude est SVOH pour la recherche rigoureuse des valeurs optimales d'hyperparamètres du classifieur SVM gaussien. La particularité de cette approche est qu'elle considère le nombre *k* de subdivision de l'ensemble d'apprentissage comme un hyperparamètre. Les résultats d'expérimentation avec le classifieur SVM tout comme le classifieur RAN (ANN) ont confirmé la pertinence du choix de la valeur du nombre k car, après des tests sur les ensembles de données IPP HPRD, H. Pylori et S. Cerevisiae, le taux de justesse affiché en utilisant SVOH s'est avérée supérieur aux taux de justesse affichés en utilisant les valeurs habituelles (5 ou 10). Enfin, l'outil SVM-BP formé dans les données HPRD en utilisant l'approche SVOH présente de meilleurs taux dans la plupart des métriques utilisées comparés à l'outil SVM-BP formé avec une valeur de k= 5 dans le chapitre précédent. Ces résultats démontrent bien que l'approche développée permet de retrouver de façon rigoureuse les valeurs optimales d'hyperparamètres du classifieur utilisé

6. Conclusion

Dans ce travail, nous avons proposé une approche modifiée de recherche sur grille combinée à la validation croisée *k*-fois. Cette approche permet d'arbitrer automatiquement entre le pourcentage de données utilisées pour entraîner un classifieur et la rigueur de l'erreur estimée, en considérant le nombre de sous-ensembles comme un hyperparamètre à ajuster pendant la phase de sélection de modèles. Alors que le nombre de sous-ensembles *k*, dans la pratique est

généralement fixé, nous avons montré, par le biais de tests sur des ensembles de données de référence bien connus, que l'approche proposée permet d'obtenir des limites d'erreur de généralisation plus légère sur l'ensemble de données de test.

References

[1] L. Shi, S. T. K. Lin, Y. Lu, L. Ye, et Y. X. Zhang, « Artificial neural network based mechanical and electrical property prediction of engineered cementitious composites », *Constr. Build. Mater.*, vol. 174, p. 667-674, juin 2018, doi: 10.1016/j.conbuildmat.2018.04.127.

[2] D. Anguita, A. Ghio, S. Ridella, et D. Sterpi, « K-Fold Cross Validation for Error Rate Estimate in Support Vector Machines. », in *DMIN*, 2009, p. 291-297.

[3] D. Anguita, L. Ghelardoni, A. Ghio, L. Oneto, et S. Ridella, « The'K'in K-fold Cross Validation. », in *ESANN*, 2012, p. 441-446.

[4] Y. Bengio et Y. Grandvalet, « No unbiased estimator of the variance of k-fold cross-validation », *J. Mach. Learn. Res.*, vol. 5, nº Sep, p. 1089-1105, 2004.

[5] J. D. Rodriguez, A. Perez, et J. A. Lozano, « Sensitivity Analysis of k-Fold Cross Validation in Prediction Error Estimation », *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, nº 3, p. 569-575, mars 2010, doi: 10.1109/TPAMI.2009.187.

[6] S. Arlot et A. Celisse, « A survey of cross-validation procedures for model selection », *Stat. Surv.*, vol. 4, nº none, p. 40-79, janv. 2010, doi: 10.1214/09-SS054.

[7] V. Vapnik, *The Nature of Statistical Learning Theory*. Springer Science & Business Media, 2013.

[8] J. L. Rodgers, « The bootstrap, the jackknife, and the randomization test: A sampling taxonomy », *Multivar. Behav. Res.*, vol. 34, nº 4, p. 441-456, 1999.

[9] C.-W. Hsu, C.-C. Chang, et C.-J. Lin, *A practical guide to support vector classification*. Taipei, 2003.

[10] G. A. Y. Laura, « Algorithme de descente du gradient stochastique », 2015.

[11] J. A. A. Brito, F. E. McNeill, C. E. Webber, et D. R. Chettle, « Grid search: an innovative method for the estimation of the rates of lead exchange between body compartments », *J. Environ. Monit.*, vol. 7, nº 3, p. 241-247, févr. 2005, doi: 10.1039/B416054A.

[12] L. Yang et A. Shami, « On hyperparameter optimization of machine learning algorithms: Theory and practice », *Neurocomputing*, vol. 415, p. 295-316, nov. 2020, doi: 10.1016/j.neucom.2020.07.061.

[13] C. N. Kopoin, N. T. Tchimou, B. K. Saha, et M. Babri, « A Feature Extraction Method in Large Scale Prediction of Human Protein-Protein Interactions using Physicochemical Properties into Bi-gram », in *2020 IEEE International Conf on Natural and Engineering Sciences for Sahel's Sustainable Development - Impact of Big Data Application on Society and Environment (IBASE-BF)*, Ouagadougou, Burkina Faso: IEEE, févr. 2020, p. 1-7. doi: 10.1109/IBASE-BF48578.2020.9069594.

[14] X.-Y. Pan, Y.-N. Zhang, et H.-B. Shen, « Large-Scale Prediction of Human Protein−Protein Interactions from Amino Acid Sequence Based on Latent Topic Features », *J. Proteome Res.*, vol. 9, nº 10, p. 4992-5001, oct. 2010, doi: 10.1021/pr100618t.

[15] Y. E. Göktepe et H. Kodaz, « Prediction of Protein-Protein Interactions Using An Effective Sequence Based Combined Method », *Neurocomputing*, vol. 303, p. 68-74, août 2018, doi: 10.1016/j.neucom.2018.03.062.

[16] W. Ma, Y. Cao, W. Bao, B. Yang, et Y. Chen, « ACT-SVM: Prediction of Protein-Protein Interactions Based on Support Vector Basis Model », *Sci. Program.*, vol. 2020, p. e8866557, juill. 2020, doi: 10.1155/2020/8866557.

[17] Z.-H. You, Y.-K. Lei, L. Zhu, J. Xia, et B. Wang, « Prediction of protein-protein interactions from amino acid sequences with ensemble extreme learning machines and principal component analysis », *BMC Bioinformatics*, vol. 14, nº S8, p. S10, mai 2013, doi: 10.1186/1471-2105-14-S8-S10.

[18] Z.-H. You, S. Li, X. Gao, X. Luo, et Z. Ji, « Large-Scale Protein-Protein Interactions Detection by Integrating Big Biosensing Data with Computational Model », BioMed Research International. Consulté le: 5 janvier 2019. [En ligne]. Disponible sur: https://www.hindawi.com/journals/bmri/2014/598129/abs/

[19] J. Martin, « Prédiction de la structure locale des protéines par des modeles de chaˆınes de Markov cachées », PhD Thesis, Citeseer, 2005.

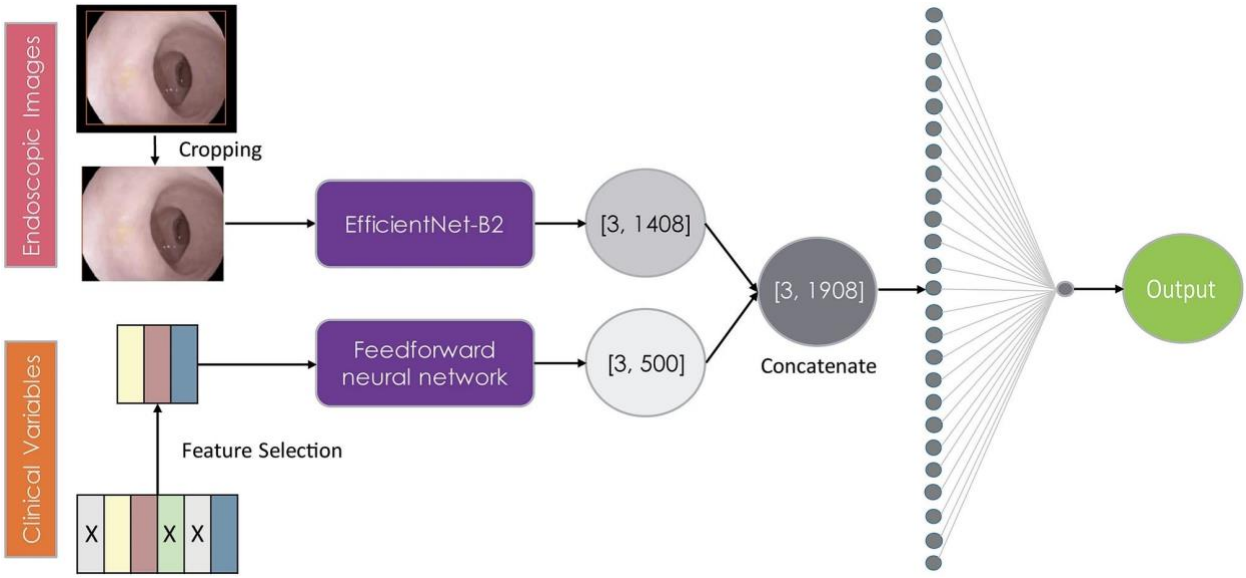
[20] V. N. Vapnik, « An overview of statistical learning theory », *IEEE Trans. Neural Netw.*, vol. 10, nº 5, p. 988-999, 1999.

[21] X. Zhang et C.-A. Liu, « Model averaging prediction by K-fold cross-validation », *J. Econom.*, vol. 235, nº 1, p. 280-301, juill. 2023, doi: 10.1016/j.jeconom.2022.04.007.

9. Conférence

Endoscopic images and clinical data to assess the response of radiation therapy

By: Selam Waktola, PhD



Short Bio

SUMMARY

My interests lie in solving real-world problems using computer vision and artificial intelligence techniques for analyzing and extracting insights from large datasets in interdisciplinary fields including medical, financial, insurance, material science, and driverless car systems. Published author of 10+ manuscripts that appear in peer-reviewed journals and several scientific international conference publications.

WORK EXPERIENCE

Sept 2022 – Present **Senior Data Scientist,** Mutual of Omaha, USA

- Building advanced machine learning models to extract impactful insights and making better business decisions.

Jul 2022 – Present **Clinical Research Associate,** Northwestern University, USA

- Developing explainable machine learning(ML) methods for computer-aided diagnosis systems.

Jul 2021 – Jul 2022 **Research Fellow in AI for Medical Image Analysis,** Johns Hopkins University, USA

- Developed AI methods for medical image analysis in image-guided surgery based on CT images.

Feb 2020 – Jul 2021 **Postdoc - AI in Radiology,** The Netherlands Cancer Institute, Netherlands

- Developed multiple ML methods to assist clinicians in stratifying cancer patients for personalized treatment.

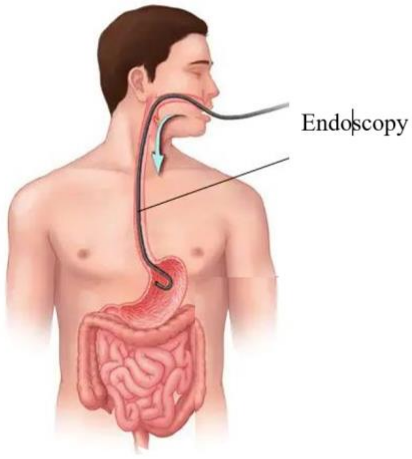
EDUCATION

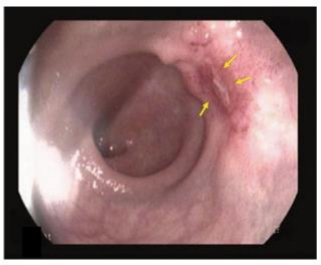
Oct 2013 – Sep 2019 **PhD in Computer Science,** Lodz University of Technology, Poland

- Developed new image processing and ML algorithms to analyze, extract insights and visualize flows based on X-ray CT data.

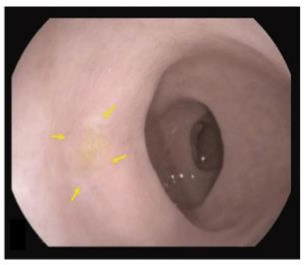
BACKGROUND

Endoscopic images in rectal cancer patients after radiation therapy

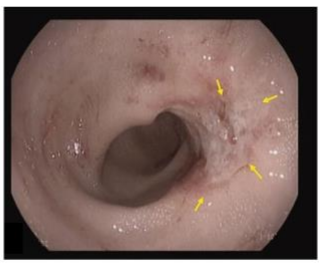




Doubtful response with a small ulcer



Complete response

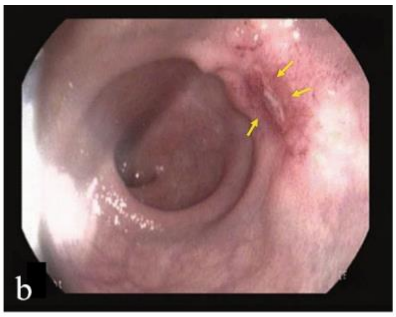


Doubtful response with a medium ulcer

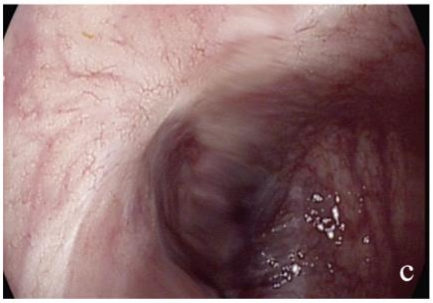
Incomplete response

Example doubtful complete responders and non-complete responders

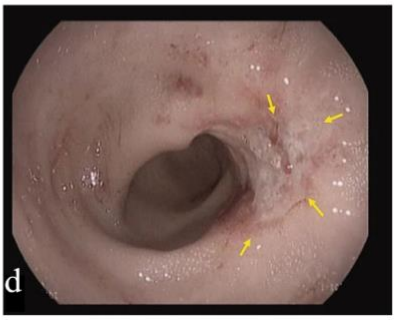
Doubtful response with complete response



Doubtful response with a small ulcer



Doubtful response with non-complete response



Doubtful response with a small-medium sized ulcer

BACKGROUND

Motivation:

- Experienced Endoscopists: AUC of ~86%
- AI Model: AUC of?

Aim:

- Accurate response evaluation is necessary to select complete responders from rectal cancer patients diagnosed treated with radiotherapy.
- The aim is to evaluate the accuracy to assess response with deep learning methods based on endoscopic images and clinical features.

DATASET

Clinical Features + **Endoscopy images**

Number of patients: 226(~4 images)
- Complete response: 109
- Non-complete response:117

	A	B	C	D	E	F	G	H	I	J	K
1	id	PatientSex	Age	cT	cN	Distance_t	Distance_c	adj_ctx	RT_scop_i	Surg_scop	response
2	./Sc	0	72	2	1	0	0	0	11.71	2.14	1
3	./Sc	0	72	2	1	0	0	0	11.71	2.14	1
4	./Sc	0	72	2	1	0	0	0	11.71	2.14	1
5	./Sc	0	72	2	1	0	0	0	11.71	2.14	1
6	./Sc	0	72	2	1	0	0	0	11.71	2.14	1
7	./Sc	0	72	2	1	0	0	0	11.71	2.14	1
8	./Sc	0	74	3	2	0	0	1	8.43	3.14	0
9	./Sc	0	74	3	2	0	0	1	8.43	3.14	0
10	./Sc	0	74	3	2	0	0	1	8.43	3.14	0
11	./Sc	0	76	3	1	4	0	0	12.29	1.71	1
12	./Sc	0	76	3	1	4	0	0	12.29	1.71	1

+

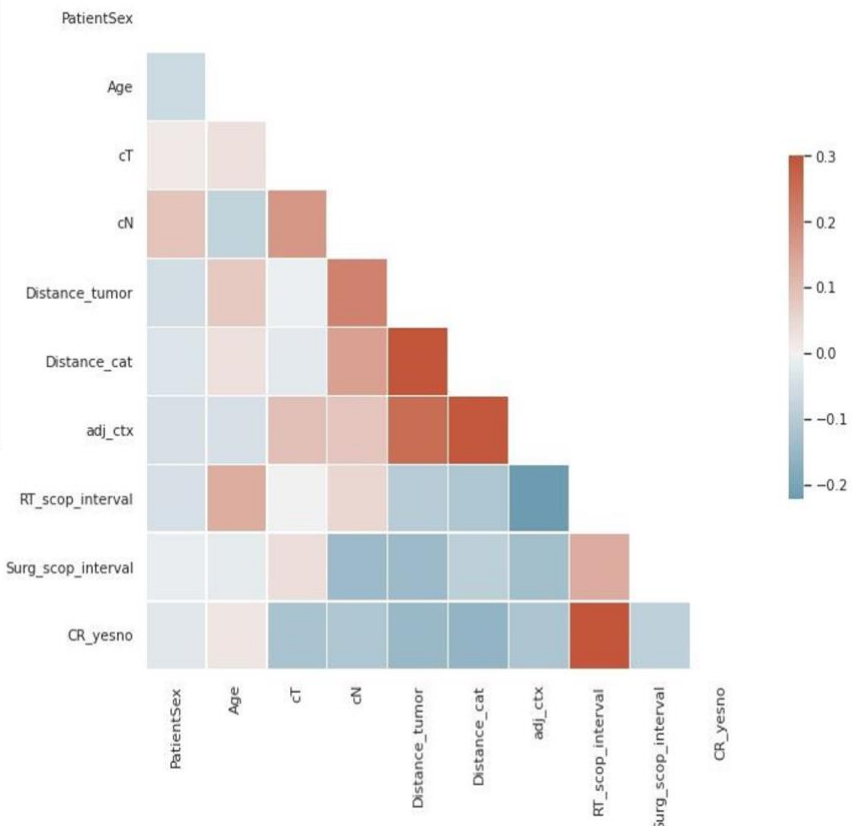
Patient demography

Variables	All (*n* = 226)	Non-CR (*n* = 117)	CR (*n* = 109)	*P*
Age, median (IQR), year	65 (58–73)	65 (58–74)	66 (59–73)	0.952
Sex, *n* (%)				
Male	153 (68)	78 (67)	75 (69)	0.731
Female	73 (32)	39 (33)	34 (31)	
Clinical T-stage, *n*(%)				
1–2	49 (22)	21 (18)	28 (26)	0.095
3	161 (71)	84 (72)	77 (70)	
4	16 (7)	12 (10)	4 (4)	
Clinical N-stage, *n*(%)				
0	54 (24)	26 (22)	28 (26)	0.038
1	64 (28)	26 (22)	38 (35)	
2	108 (48)	65 (56)	43 (39)	
Neoadjuvant treatment, *n*(%)				
5 × 5 Gy + prolonged waiting interval	20 (10)	16 (14)	4 (4)	<0.001
CRT	206 (90)	101 (86)	105 (96)	
Adjuvant chemotherapy, *n*(%)				
Yes	41 (18)	22 (19)	19 (17)	0.227
No	185 (82)	95 (81)	90 (83)	
Time between last radiotherapy and endoscopy, median (IQR), weeks	10 (8–15)	8 (8–12)	12 (9–18)	<0.001

Correlation coefficients between clinical features

Clinical Features

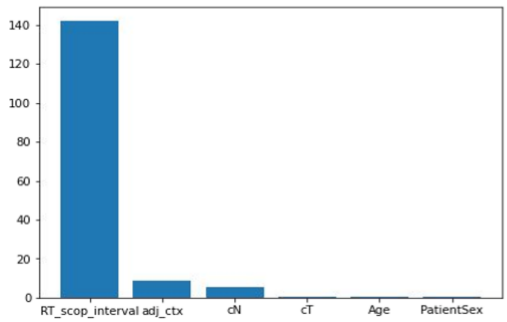
- age
- sex
- clinical T-stage
- clinical N-stage
- neoadjuvant treatment
- time between restaging endoscopy and surgery



Clinical variables model

Clinical feature selection using SelectKBest

Feature importance



MODEL 1: All clinical variables:

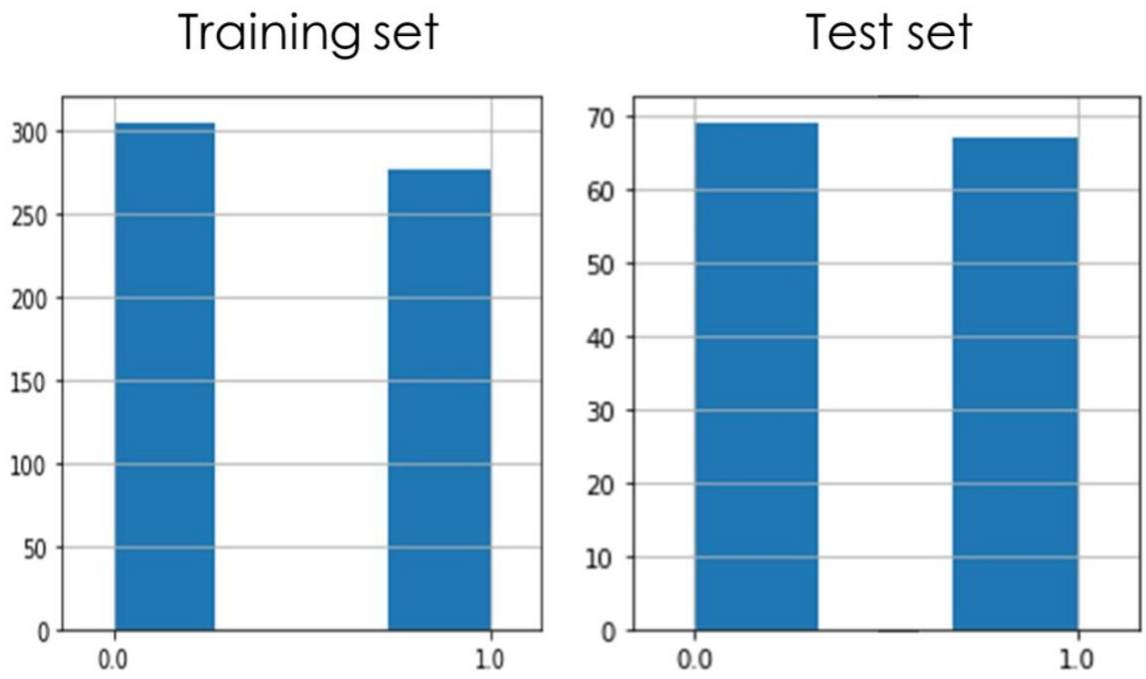
['PatientSex', 'Age', 'cT', 'cN', 'adj_ctx', 'RT_scop_interval']

MODEL 2: Selected clinical variables:

['cN', 'adj_ctx', 'RT_scop_interval']

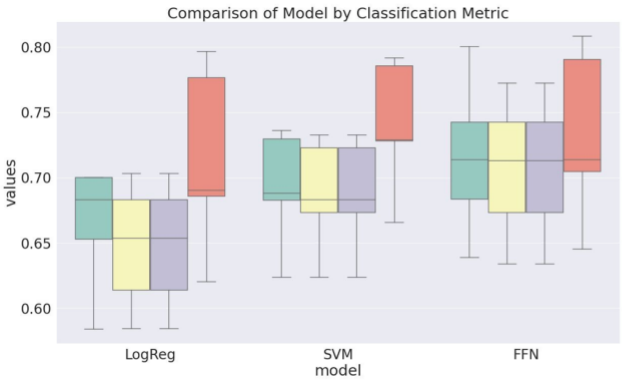
SelectKBest feature selection technique was used to choose the best predicting clinical variables for the outcome (CR or non-CR).

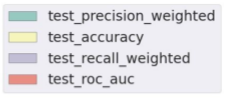
Class distribution



70% for training

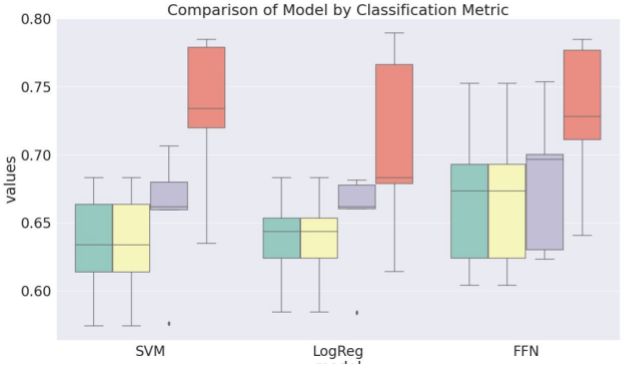
30% for test

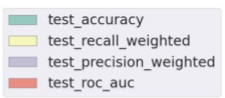




MODEL 1: All clinical variables:

['PatientSex', 'Age', 'cT', 'cN', 'adj_ctx', 'RT_scop_interval']





MODEL 2: Selected clinical variables:

['cN', 'adj_ctx', 'RT_scop_interval']

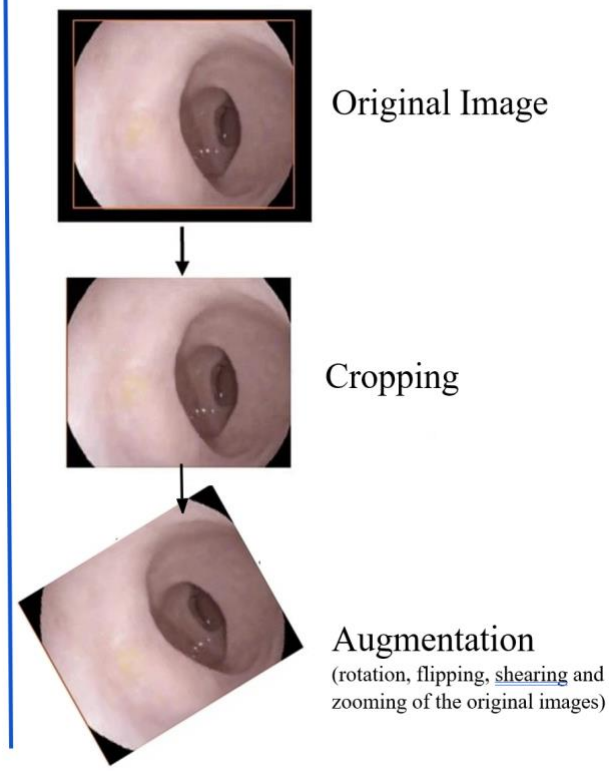
- Feedforward neural network (FFN)
- Support vector machine (SVM)
- Logistic regression (LogReg)

Deep Learning Model

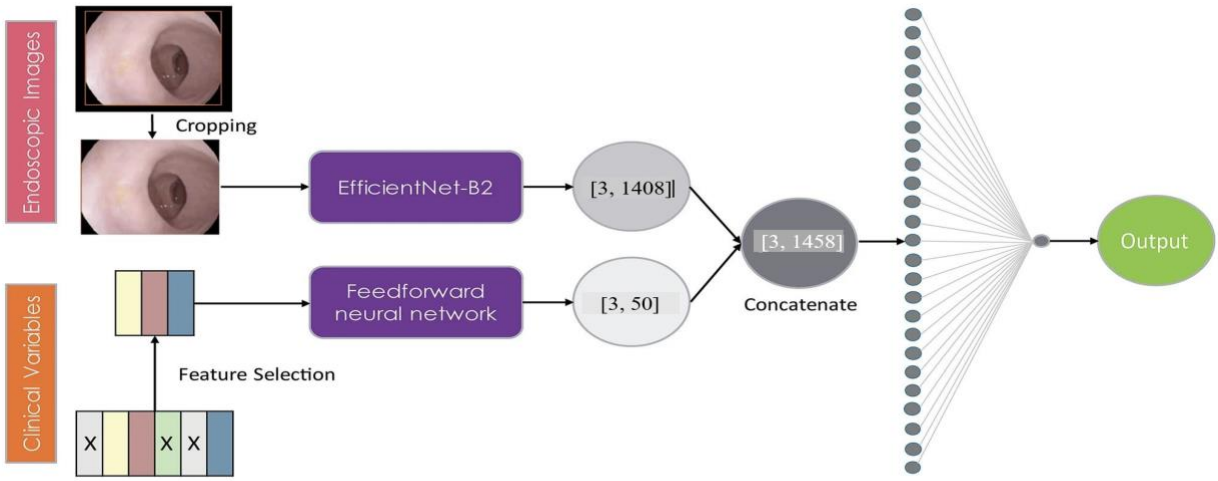
Selected models

- Top results achieved models in natural object recognition in ImageNet Large Scale Visual Recognition Challenge(ILSVRC) competition.
  - **EfficientNet** is one of the recent convolutional neural network architectures that achieve much better accuracy and efficiency as compared to the previous ConvNets.
  - **ResNet** design is able to stabilize gradient and relieve the gradient explosion or vanishing problem.
  - **MobileNet** separates convolution into depth wise and pointwise convolutions, compresses the network and also keeps the accuracy level.
- **Deep learning model construction:**
  - **Endoscopic image only**
  - **Endoscopic image and selected clinical features**

Image preprocessing

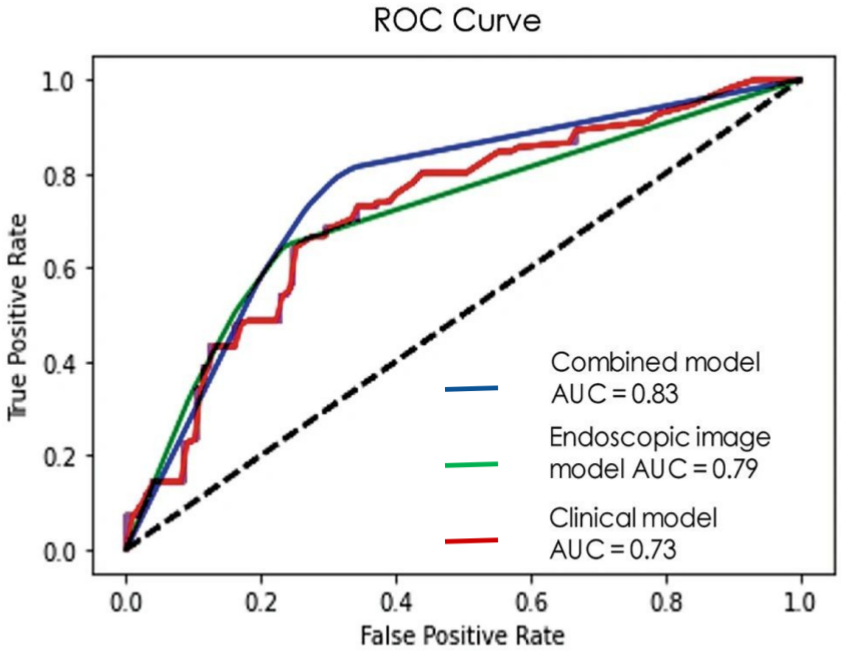


Combined model architecture



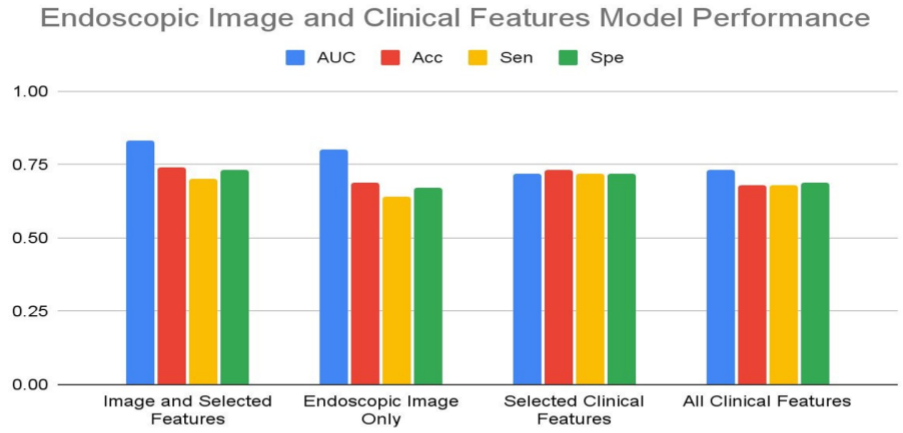
Combined model architecture: [3, 1408] denotes RGB channel and the last channels in EfficientNet-B2. [3, 50] denotes selected clinical features and the number of neurons in feedforward neural network.

Performance of the models



ROC curve of EfficientNet-B2 for the endoscopic image model and combined model and ROC curve of feedforward neural network model for selected clinical variables. AUC Area under the ROC curve.

RESULTS AND SUMMARY



Results:

- Experienced Endoscopists: AUC of ~86%
- AI Model : AUC of ~83%

Future works:

- Train with large patient dataset.
- Deploy model on AWS.

Conclusion

- Deep learning on endoscopic images to assess the response of rectal cancer after chemoradiation.

- This study shows that deep learning has a modest accuracy (AUCs 0.76-0.83).

- Deep learning models can however be further improved and may become useful to assist endoscopists in evaluating the response. More well designed prospective studies are required.

- The trained models could be deployed on the internet for remote hospitals where medical experts are scarce.

RFE: Haak, H.E., Gao, X., Maas, M., **Waktola, S.**, Benson, S., Beets-Tan, R.G.H., Beets, G.L., Leerdam, M. van., and Melenhorst, J. 'The use of deep learning on endoscopic images to assess the response of rectal cancer after chemoradiation', 2021 Surgical Endoscopy.