

Mise en place d'outils de détection d'anomalies pour une réduction du taux de prévalence de la bilharziose

Fiorenantsoa Santatra, Bassirou Kassé, Bamba Gueye, Moussa Diallo
Université Cheikh Anta Diop, Dakar, Senegal

Résumé. La bilharziose (urinaire et intestinale) constitue un problème majeur de santé publique au Sénégal avec un taux de prévalence globale variant de 0,3 à 1%. Les paramètres physico-chimiques tels que le pH et la température des points d'eau, l'ensoleillement jouent un rôle prédominant dans le développement des parasites et hôtes intermédiaires responsables de la transmission de la bilharziose lorsque certains seuils sont atteints. Une architecture basée sur les réseaux de capteurs capables de prélever ces informations a été déployée. Afin de détecter et de prédire les anomalies dans les points d'eau, nous avons élaboré des modèles d'apprentissage automatique. Les résultats obtenus montrent que le modèle de détection « Support Vector Machine » fournit les meilleurs résultats.

Mots clés : bilharziose, détection d'anomalies, apprentissage automatique, réseaux de capteurs.

1 Introduction

Plus de 2 milliards de personnes dans le monde sont affectées par la bilharziose ou schistosomiase [1] qui est la 2ème cause de consultation hospitalière après le paludisme [2][3]. En outre, le taux de prévalence a énormément augmenté au niveau du fleuve Sénégal (Mali, Mauritanie) avec un taux d'infestation moyen atteignant 71,8% [3][4]. Les méthodes de détection de la bilharziose existantes actuellement consistent à faire des analyses sur les selles ou bien les urines des patients pour y déterminer la présence des œufs, c'est-à-dire les schistosomes afin de prouver qu'effectivement le patient héberge des schistosomes adultes ou bien infecté par la maladie [1]. Le traitement des malades présentant des symptômes se fait par chimiothérapie en recourant aux médicaments existants tels que le « Praziquantel » et l' « Arthemeter », qui ne sont pas efficaces contre la bilharzie aiguë et présente un taux de guérison très faibles, car ne possédant pas d'impact sur le schistosome et les œufs immatures [5]. Il peut exister des personnes qui ont une sérologie positive et ne présentant pas de symptômes [2]. Ils continuent à infecter les points d'eau (fleuves, marigots, mares, lacs) lors des besoins naturels. Les solutions de prévention existantes consistent à utiliser des produits chimiques tels que le sulfate de cuivre pour tuer les mollusques (les hôtes intermédiaires) mais qui entraîne des troubles dans l'écosystème [3][4].

De nos jours, la préoccupation majeure est de trouver comment réduire le taux de prévalence de la bilharziose au Sénégal. C'est dans ce sens que les auteurs de [2] ont mis en place une architecture basée sur les réseaux de capteurs afin de détecter en temps réel si les conditions aquatiques sont favorables au développement du cycle de transmission de la bilharziose. En effet, le pH et la température de l'eau au niveau des points d'eau de même que l'irradiation solaire ou ensoleillement sont des paramètres physico-chimiques qui concourent au développement des hôtes intermédiaires et des parasites responsables de la transmission de la bilharziose [6]. Une architecture distribuée basée sur l'Internet des Objets a été déployé pour la prévention du risque d'infestation des populations [2]. Les informations prélevées par les nœuds capteurs (pH, température, irradiation solaire) sont envoyées par (« Short Message Service ») (*SMS*) ou via le protocole « *ZigBee* » [7] à un point de collecte. Les travaux présentés dans [2] ne gèrent pas la détection des anomalies présente dans les points d'eau mais s'arrête seulement dans la phase de collecte des informations physico-chimiques du milieu aquatique. La détection d'anomalies se définit usuellement par l'action de discriminer dans un jeu de données composé d'observations et d'attributs non connus à l'avance, caractérisant un système cible [8]. En effet, les observations ne correspondent pas au comportement normal représenté par la majorité des observations. Par anomalie, nous désignons l'atteinte des seuils (pH, température, irradiation solaire) favorables au développement du cycle de transmission de la bilharziose. Nous avons entraîné quelques algorithmes d'apprentissage automatique et nous avons ensuite développé un modèle optimal capable de détecter en temps réel les anomalies présentes dans les points d'eau.

Le reste du papier s'articule comme suite. La section 2 présente un bref état de l'art sur le cycle de transmission de la bilharziose. La section 3 illustre les différentes techniques de détection d'anomalies supervisées ainsi que les techniques utilisées dans la prévision de phénomènes. La section 4 décrit la mise en place des outils de détection d'anomalies appliqués à notre jeu de données. La section 5 montre les résultats obtenus par les modèles de détection et de prévision. Enfin, la section 6 conclut ce travail.

2 Bref aperçu sur la bilharziose

2.1 Le cycle de transmission de la bilharziose

Le cycle commence avec une personne infectée qui héberge un schistosome adulte noté « A » sur Fig. 1 [1]. Au cours de la première étape, les œufs sont excrétés dans l'urine ou les matières fécales d'une personne infectée. Ils sont viables jusqu'à

une semaine. Il existe trois types d'œufs qui sont représentés sur la Fig. 1 avec l'étiquette « B ». La transformation des écosystèmes crée des habitats favorables au développement des mollusques qui hébergent le miracidium une fois libéré par les œufs au contact des points d'eau. Le miracidium est illustré par l'étiquette « C » et les mollusques par l'étiquette « D ». Le miracidium devient une larve de cercaires à l'intérieur des mollusques. Il convient de noter que 4 à 6 semaines après l'infection, les cercaires sont libérées dans le point d'eau (étiquette « E »). Elles peuvent pénétrer activement la peau d'un hôte en 72 h. Par la suite, les cercaires se transforment en schistosomes mâles ou femelles (l'étiquette intitulée « A ») et le cycle de transmission recommence. La durée de vie des schistosomes adultes varie entre 3 et 5 ans, néanmoins, elle peut durer jusqu'à 30 ans [1]. Ainsi, l'hôte humain héberge le parasite et participe activement à la propagation de la schistosomiase.

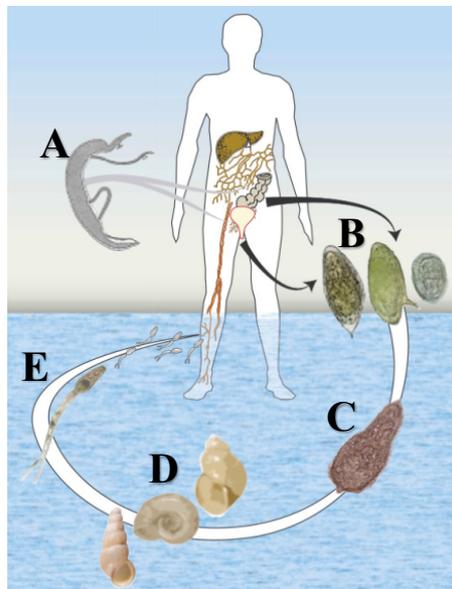


Fig. 1. Cycle de transmission de la bilharziose (source [1])

2.2 Conditions de développement des parasites et hôtes intermédiaires

Le cycle de transmission de la bilharziose repose sur des facteurs physico-chimiques qui ont une incidence importante sur la dynamique des populations des mollusques d'eau douce. L'irradiation solaire joue un rôle majeur lorsque les œufs sont libérés, par une personne infectée, au contact des points d'eau [1] [6]. En effet, l'intensité lumineuse diurne a un impact important sur la maturation des œufs mais aussi une action stimulante sur la reproduction des mollusques adultes, la croissance des larves et des jeunes mollusques. L'irradiation solaire moyenne au

Sénégal est approximativement égale à $242\mu\text{mol}/\text{m}^2/\text{s}$, ce qui représente une irradiation solaire de 5, 8 kWh/m²/jour. Par conséquent, cette valeur illustre un degré élevé d'intensité lumineuse.

La température des points d'eau joue un rôle important dans la solubilité des gaz tels que l'oxygène qui sont nécessaires à l'équilibre de la vie aquatique. L'activité métabolique des organismes aquatiques est également accélérée en fonction des fluctuations de la température. Si la température de l'eau est assez chaude pendant une longue période, les mollusques d'eau douce qui transmettent la schistosomiase sont susceptibles de se développer dans des conditions appropriées. La température optimale qui permet des développements favorables est mesurée entre 25°C à 28 °C [6]. Ensuite, lorsque la température varie de 28 °C à 30 °C, les conditions sont défavorables. À ce point, la reproduction cesse. Néanmoins, une température supérieure à 42 °C, pendant deux heures, provoque la mort des mollusques [6].

Quant au pH, des valeurs autour du neutre c'est à dire de 6,2 à 8.5 sont favorables au bon développement des créatures aquatiques [1] [6]. Le Tableau 1 résume les seuils à détecter, sur les paramètres considérés, pour prédire si les conditions aquatiques sont favorables au développement du cycle de transmission de la bilharziose. Dès lors, il faudra alerter pour interdire les points d'eau momentanément et faire les traitements idoines.

Tableau 1. Conditions favorables au développement du cycle de transmission de la bilharziose.

Paramètres	Seuils
pH de l'eau	Entre 6,5 et 8.2
Température de l'eau	Entre 25°C et 28°C
Irradiation solaire	$\geq 242 \mu\text{mol}/\text{m}^2/\text{s}$

Sur la base des seuils du Tableau 1, on peut déterminer en temps réel grâce à des algorithmes de détection d'anomalies : (i) si les conditions sont favorables à la maturation des œufs ; (ii) si les hôtes intermédiaires peuvent vivre et se développer de manière optimale ; (iii) si une reproduction efficace des mollusques est possible.

3 Techniques de détection et de prévision d'anomalies

La détection d'anomalies supervisée décrit les modèles dans lesquels les données comprennent des ensembles de données d'entraînement et de données de tests entièrement étiquetées. Ce scénario est très similaire à la classification traditionnelle, à l'exception que les classes sont généralement fortement déséquilibrées. Tous les algorithmes de classification ne conviennent donc pas

parfaitement à cette tâche. Par exemple, les arbres de décision ne peuvent pas bien gérer les données déséquilibrées, alors que les machines à vecteurs support (SVM) ou les réseaux neuronaux artificiels (ANN) devraient mieux fonctionner [9]. Quant à la détection d'anomalies semi-supervisée, elle utilise également des ensembles de données de formation et de test. Cependant, les données d'entraînement ne sont constituées que de données normales sans anomalie. La difficulté ici réside dans la création d'un modèle exhaustif qui représente l'ensemble des données normales. La détection d'anomalies non supervisée est la configuration la plus flexible qui ne requiert aucune étiquette. En outre, il n'y a pas non plus de distinction entre une formation et un ensemble de données de test.

3.1 Les types d'anomalies

Il existe deux possibilités pour les sorties d'un algorithme de détection d'anomalies. Une étiquette peut être utilisée pour indiquer si une instance est une anomalie ou non. Un score ou une valeur de confiance peut être un résultat plus informatif indiquant le degré d'anomalies. Pour la détection d'anomalies supervisée, souvent une étiquette est utilisée en raison des algorithmes de classification disponibles. D'autre part, pour les algorithmes de détection d'anomalies semi-supervisés et non supervisée, les scores sont plus fréquents. Les anomalies peuvent être réparties en trois groupes : les anomalies ponctuelles, contextuelles et collectives [9]. Dans notre étude, nous visons à déterminer si les conditions aquatiques sont favorables au développement du cycle de transmission de la bilharziose. Ce cycle dépend de certains seuils où nous pouvons mettre les étiquettes par rapport aux anomalies et les données normales sur les données historiques en notre possession. Etant donné que chaque observation du jeu de données peut être une anomalie, nous sommes en présence d'un problème de détection d'anomalies supervisées avec des anomalies de type ponctuelles.

3.2 Evaluation des techniques de détection d'anomalies supervisée

La détection d'anomalies s'évalue par la capacité du modèle à discriminer avec précision les anomalies des données normales et également par le taux d'erreurs réalisées par le modèle. Les anomalies qui ont été identifiées correctement par le modèle de détection sont appelées des « Vrai Positif » (V.P). Similairement, les données normales qui ont été classées correctement par le modèle s'appellent « Vrai Négatif » (V.N). Par contre, les erreurs réalisées par le modèle dans le cas où il a échoué pour identifier des anomalies s'appellent « Faux Négatif » (F.N). Si le modèle identifie des données normales comme des anomalies, c'est-à-dire qu'il a levé des fausses alarmes, on parle de « Faux Positif » [8][9]. Les quelques métriques illustrées au niveau du Tableau 2 vont être utilisées pour mesurer l'efficacité des algorithmes de détection d'anomalies par rapport à nos données.

Tableau 2. Quelques mesures de performance des algorithmes d'apprentissage supervisé

Mesures	Description	Formules
Précision	Taux de détection des données normales	$\frac{VP}{VP + FP}$
Rappel	Taux de vrais positifs	$\frac{VP}{VP + FN}$
TFP	Taux de faux positifs	$\frac{FP}{VN + FP}$

Caruana et al. ont considéré 10 modèles de détection supervisée appliquées à 11 jeux de données de caractéristiques différentes et dans plusieurs domaines (médical, télédétection, business, ...) [10]. Les performances de chacun des modèles ont été évaluées par 8 critères différents appartenant à trois groupements. Les résultats obtenus montrent que les algorithmes Random Forest (RF) ou Forêt Aléatoire, Artificial Neural Network (ANN) ou Réseau de Neurones Artificiels (RNA), et Support Vector Machines (SVM) ou Machine à Vecteur Support donnent des performances supérieures à 85 %. C'est ainsi que dans notre étude, nous avons choisis ces trois modèles pour tester notre jeu de données.

4 Mise en place des outils de détection d'anomalies et de prévision

4.1 Le jeu de données collecté par le réseau de capteurs

Le jeu de données stocké dans une base de données « *MySQL* » contient des échantillons (pH, température, irradiation solaire) collectés pendant 15 jours durant le mois de juillet 2016 par le réseau de capteurs déployé par [2]. Le pH et la température de l'eau (respectivement irradiation solaire) sont collectés chaque 1 minute (respectivement 2 minutes). Compte tenu des données disponibles, nous allons prendre une période journalière, une fenêtre d'échantillonnage des données de 1 heure et l'horizon de notre prévision se limitera à une journée.

Souvent les données collectées ne sont pas encore dans la forme voulue pour pouvoir les exploiter correctement. Il est nécessaire de passer par une étape de prétraitement des données. Après les mises en forme des données historiques disponibles, on les divise en deux parties dont la première partie, appelé données d'entraînement, sert à entraîner un algorithme d'apprentissage afin de développer un modèle de prédiction, et dont le reste des données (données de test) sert à évaluer la performance du modèle synthétisé. Dans le cas où le modèle présente

des performances suffisantes, le développement du modèle de prédiction est donc terminé avec succès. Si ce n'est pas le cas, il faut passer à l'étape de diagnostic du modèle pour trouver des solutions d'amélioration et revenir à l'étape de modélisation ou peut-être même parfois continuer à collecter beaucoup plus de données jusqu'à obtenir de bon résultat de performance.

Les données considérées viennent d'un réseau de capteurs et il peut y avoir des risques qu'elles soient bruitées. Une des meilleurs techniques pour avoir une bonne estimation des signaux bruités est l'utilisation des filtres de Kalman [12]. Après avoir appliqué les filtres de Kalman, nous avons trouvé une absence de bruit dans le jeu de données. Ensuite, nous avons normalisé les données en utilisant la normalisation min-max dans l'intervalle [0,1]. L'objectif de la normalisation est de réduire le temps d'entraînement ou de traitement par les algorithmes qui sont assez gourmands en ressources [8].

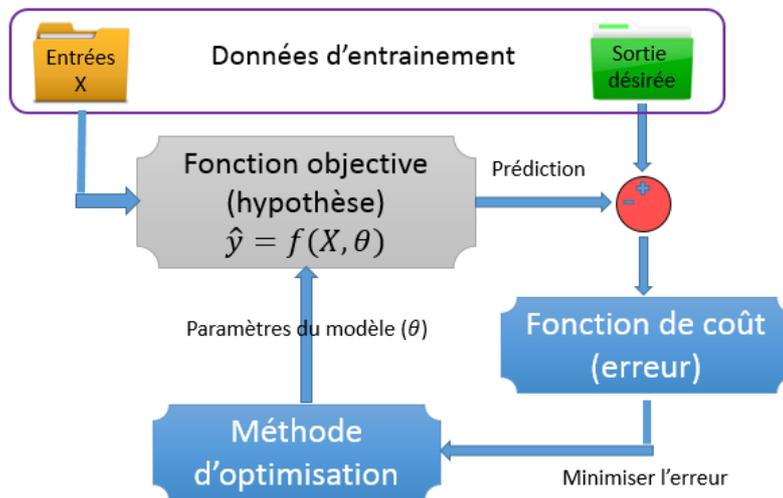


Fig. 2. Entraînement d'un algorithme d'apprentissage automatique

4.2 Processus d'apprentissage

La Fig. 2 résume l'entraînement des algorithmes d'apprentissage automatique. Nous devons définir la fonction mathématique appelée fonction hypothèse [12] qui est souvent dépendante de certains paramètres θ_i qui doivent être initialisés au début du processus d'entraînement du modèle. Le but de l'apprentissage est de trouver les valeurs de ces paramètres θ_i qui permettent de décrire au mieux les phénomènes à prédire. Pour y arriver, les valeurs des paramètres θ_i sont changées à chaque itération jusqu'à ce que la valeur de la fonction erreur soit minimisée en utilisant une méthode d'optimisation. Un des algorithmes de minimisation fréquemment utilisé est la descente de gradient mais il existe d'autres méthodes

plus complexes et plus élaborées [12]. Chaque valeur différente des paramètres θ_i correspond à un modèle différent. Le processus d'entraînement de l'algorithme peut aboutir à plusieurs modèles candidats et pour choisir le meilleur modèle on utilise des critères d'information qui mesure la qualité des modèles synthétisés. Ce sont des critères basés sur la vraisemblance et dont les valeurs minimales de ces critères indiquent le modèle le plus approprié.

Tableau 3. Résumé des données d'entraînement, échantillonnées par période de 10 minutes

Noms	Variables	Type	Echantillons	Description
« Training set »	date_time	POSIXct	2160	Temps d'échantillonnage des données par pas de 10 minutes
	Ph	Numeric		Moyenne des pH tous les 10 minutes
	Temper	Numeric		Moyenne des températures tous les 10 minutes
	ensol	Numeric		Moyenne des ensoleillements tous les 10 minutes
	detection	Factor		<ul style="list-style-type: none"> • « 1 » si (6.5 < ph < 8.2) et (25 < temper < 28) et (ensol > 442) ; • « 0 » sinon
« Scaled training set »	date_time	POSIXct	2160	Temps d'échantillonnage des données par pas de 10 minutes
	Ph	Numeric		Valeurs normalisées des pH, des températures et de l'ensoleillement entre 0 et 1 du jeu de données « training_set ».
	Temper	Numeric		
	ensol	Numeric		
	Detection	Factor		Garde les mêmes valeurs

4.3 Mise en œuvre

De nombreux outils de développement d'algorithmes d'apprentissage automatique existent mais notre choix s'est porté sur le langage R et son environnement de développement [13]. En effet, le pH et la température de l'eau présentent 18000 échantillons tandis que l'ensoleillement a 12000 échantillons. Afin d'effectuer nos analyses sur des périodes de temps, nous avons choisi une fenêtre de traitement moyennant les données dans un intervalle de temps égal pour les trois variables que nous utilisons (ph de l'eau, température de l'eau et l'irradiation solaire). La plus petite fenêtre de temps choisie est 10 minutes. Nous avons calculé la moyenne des données toutes les 10 minutes pour obtenir notre jeu de données. Au niveau de « R », la date et l'heure est de type « POSIXct » tandis que le pH de l'eau, la température de l'eau, l'irradiation solaire sont de type « numeric ». Enfin, une variable qualitative nommée « detection » de type « factor » est la sortie que nous voulons prédire. Les valeurs de cette variable sont déterminées sur la base des

seuils de détection fixés niveau du Tableau 1. Le Tableau 3 résume les données utilisées pour entraîner nos modèles d'apprentissage automatique.

Les tests de performance de détection de nos modèles ont été effectués sur plusieurs intervalles de temps qui ne dépassent pas une journée pour être suffisamment précis mais qui n'est pas moins de 1 heure pour éviter une période trop courte. Une prévision de plus d'une journée risque d'être imprécise. Les données que nous avons utilisées comme données de test sont donc échantillonnées sur des fenêtres de temps de 1 heure, 2 heures, 3 heures et 6 heures. La Fig. 3 présente notre jeu de données d'entraînement de nos modèles sur une fenêtre de 10 minutes. Les points en bleu représentent la présence des conditions favorables au développement du cycle de transmission de bilharziose, c'est-à-dire les anomalies qui sont observées, tandis que les points en noir montrent les données normales.

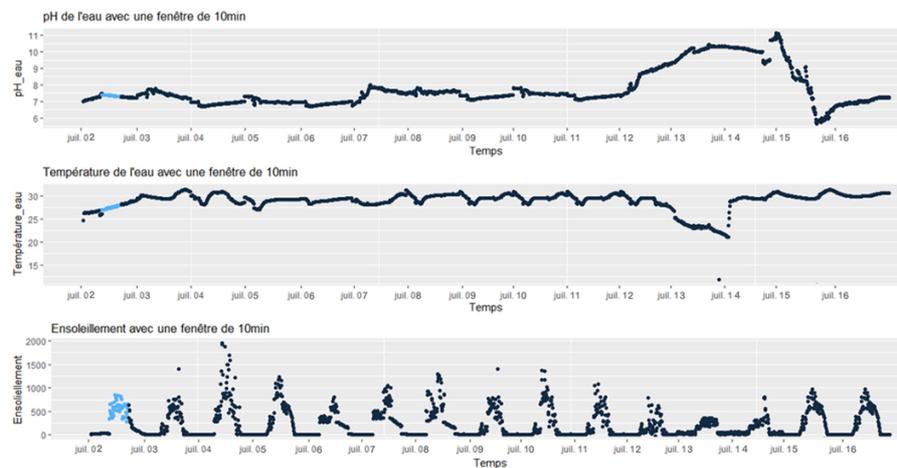


Fig. 3. Jeu de données d'entraînement de nos modèles

Le fait d'échantillonner les données dans une fenêtre plus grande ou plus petite peut masquer l'existence d'anomalies qui est identifié dans un intervalle de temps plus court ou bien introduire des fausses alertes supplémentaires causées par une agrégation par moyenne des données. Le Tableau 4 présente l'impact des fenêtres de 1 h, 2 h, 3 h et 6 h sur les données.

Sur une fenêtre de 1 heure, nous introduisons dans la variable « detection » une demi-heure de fausse alerte. Ce résultat est acceptable car les fausses alertes ne représentent pas de danger mais seulement quelques gênes dans les activités des villageois et des surplus de coût d'intervention. Le choix d'utiliser une fenêtre de 6 heures nous met dans une situation quelque peu similaire au précédent cas. Sauf que le nombre d'heure de fausse alerte augmente jusqu'à 3 heures et demi. Par contre, une fenêtre de 2 heures masque une demi-heure de risque potentiel de

transmission de la maladie et une fenêtre de 3 heures masque encore 2 heures et demi qui présentent des dangers non détectés.

Tableau 4. Impact des fenêtres de 1 h, 2 h, 3 h et 6 h sur les données.

Fenêtres	Nombre d'anomalies	Nombres d'heures	Observations
Référence (10mn)	51/2160	8,5	Toutes les anomalies réelles
1 heure	9/360	9	Une demi-heure de fausse alerte
2 heures	4/180	8	Une demi-heure non détecté
3 heures	2/120	6	2 heures et demi non détectés
6 heures	2/60	12	3 heures et demi de fausse alerte

5 Résultats

Le Tableau 5 compare les taux de réussite du modèle dans la détection des anomalies et des données normales en fonction des tailles de fenêtre. Nous remarquons que notre modèle « ANN » a réussi à détecter 100 % des observations d'anomalies présentes dans les jeux donnés de test quelle que soit la fenêtre choisie. Par contre, le modèle a commis des erreurs dans la classification des observations normales. RNA a levé comme anomalie 5,7 % des données normales pour la fenêtre de 1 heure qui présente le pourcentage la plus petite de fausse alarme suivi de la fenêtre de 6 heures avec un taux d'erreur de 6.9 %. Pour la fenêtre de 3 heures, le pourcentage d'erreur de classification des données normales est de 9.32 %. La fenêtre la moins efficace est celle de 2 heures qui compte une proportion de 11.93 % de fausses alarmes.

Tableau 5. Taux de détection réalisé par « Artificial Neural Network (ANN) »

Fenêtres	1 h	2 h	3 h	6 h
Taux de V.P détectées (%)	100	100	100	100
Taux de F.N détectées (%)	0	0	0	0
Taux de V.N détectées (%)	94,3	88,07	90,68	93,1
Taux de F.P levées (%)	5,7	11,93	9,32	6,9

Le Tableau 6 montre le taux de réussite et d'erreur dans la détection des anomalies et des données normales selon les différentes fenêtres de traitement utilisées. Nous remarquons, que le modèle « RF » présente de bons résultats sur toutes les fenêtres sauf pour celle de 1 heure où il n'arrive pas à détecter 11,11 % des anomalies. Il faut rappeler que la fenêtre de 1 heure est le meilleur choix car n'introduit que seulement une demi-heure d'erreur (fausse alerte) qui reste très peu significatif dans les données de test. Cette fenêtre est handicap pour la technique « RF »

contrairement à la technique « ANN » même si elle commet énormément d'erreur en matière de fausses alertes. Si on enlève la fenêtre de 1 heure, le modèle « RF » devient très performant.

Tableau 6. Taux de détection réalisé par « Random Forest (RF) »

Fenêtres	1 h	2 h	3 h	6 h
Taux de V.P détectées (%)	88,89	100	100	100
Taux de F.N détectées (%)	11,11	0	0	0
Taux de V.N détectées (%)	100	100	100	100
Taux de F.P levées (%)	0	0	0	0

On remarque qu'au niveau du Tableau 7, la technique SVM est très performante par rapport aux différentes fenêtres d'échantillonnage. Elle détecte 100 % des anomalies dans les jeux de données de test mais encore il n'y a presque pas de fausses alarmes levées. Notons quand même que pour la fenêtre de 3 heures, SVM a levé 0,84 % des données normales comme des anomalies. Cependant ce pourcentage d'erreur est négligeable car la fenêtre de 3 heures a été jugée non adéquate vue qu'elle masque des anomalies. En résumé, SVM est plus efficace sur les fenêtres de temps optimales (fenêtre de 1 heure et 6 heures pour l'échantillonnage des données) que les techniques RF et ANN. En outre, SVM est le moins coûteux en termes de ressources sur les trois modèles si on en juge par les temps d'entraînement.

Tableau 7. Taux de détection réalisé par « Support Vector Machine (SVM) »

Fenêtres	1 h	2 h	3 h	6 h
Taux de V.P détectées (%)	100	100	100	100
Taux de F.N détectées (%)	0	0	0	0
Taux de V.N détectées (%)	100	100	99,16	100
Taux de F.P levées (%)	0	0	0,84	0

6 Conclusion

Nous avons appliqué sur les mesures de pH, température, et irradiation solaire collectés par un réseau de capteurs déployé sur un point d'eau trois techniques de détections d'anomalies. La technique SVM a réussi à détecter correctement 100 % des anomalies avec aucune fausse alarme dans les deux fenêtres d'échantillonnage les plus pertinentes. Nous visons à valider la performance des modèles développés en collectant beaucoup plus de données et à mettre en place des modèles de prévisions d'anomalies. La prévision est un outil particulièrement utile dans la prise

de décisions. Dans le cas de notre étude, nous visons à prévoir les potentiels risques de transmission de la bilharziose au niveau des points d'eau. Il faudra aussi intégrer nos outils de détection dans la plateforme d'Internet des objets déployée.

Références

1. B. Gryseels, K. Polman, J. Clerinx, and L. Kestens, "Human schistosomiasis", *The Lancet*, vol. 368, no. 9541, pp. 1106-1118, 2006.
2. B. Kassé, M. Diallo, B. Gueye, H. Elbiaze, "Sensor-Based Bilharzia Detection for More Sensitive Disease-Prediction and Control-Model", *InterSol 2017, LNICST 204*, Dakar, Senegal.
3. Sy I., Diawara L., Ngabo D., Barbier D., Dreyfuss G., Georges P., "Bilharzioses au Sénégal oriental prévalence chez les enfants de la région de Bandafassi", *Med Trop* 2008 : 267-271.
4. Oudabdallahi M., Ouldbezyd M., Diop C., Dem E., Lassana K., "Epidémiologie des bilharzioses humaines en Mauritanie. L'exemple de la droite du fleuve Sénégal", *Bull. Soc. Pathol. Exot.* (2010).
5. L. Pratio, B. Mariani, E. Brunetti, R. Maserati, B. Antonella, S. Novati, and G. Chichino, "Failure of repeated treatment with praziquantel and arthemeter in four patients with acute schistosomiasis", *Journal of Travel Medicine*, vol. 21, no. 2, pp. 133-136, 2014.
6. N. McCreesh M. Arinaitwe, W. Arineitwe, E. M. Tukahebwa, and M. Booth, "Effect of water temperature and population density on the population dynamics of schistosoma mansoni intermediate host snails", *Parasites & Vectors*, vol. 7, 2014.
7. ZigBee Alliance, "New zigbee pro feature: Green power connecting battery free devices", White Paper, 2012. [Online: Available: <http://www.zigbee.org/zigbee-for-developers/network-specifications/zigbeepro/>].
8. C. Sauvanaud, "Monitoring et détection d'anomalies par apprentissage dans des infrastructures virtualisées", *Informatique [cs]*, INSA Toulouse, France 2016.
9. V. Chandola, A. Banerjee, V. Kumar, "Anomaly detection: A survey", *ACM Comput. Surv.* 41, 3, Article 15 (July 2009), 58 pages.
10. R. Caruana, A. Niculescu-Mizil, "An Empirical Comparison of Supervised Learning Algorithms", *ACM ICML*, pp 161-168, 2006, Pennsylvania, USA.
11. A. Niculescu-Mizil, R. Caruana, "Predicting good probabilities with supervised learning", *ACM ICML*, pp 625-632, 2005, Bonn, Germany.
12. P.J. Brockwell, R.A. Davis. "Time Series: Theory and Methods", Second Edition, *Springer Series in Statistics*, 1991.
13. R Development Core Team (2008), "R: A Language and Environment for Statistical Computing", R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.