Toward a Data Fusion Based Framework to Predict Schistosomiasis Infection

Teegwende Zougmore*, Sadouanouan Malo†, Bamba Gueye‡, Stanislas Ouaro§

*Université Alioune Diop de Bambey, Bambey, Senegal, Email: teeg-wende@uadb.edu.sn
† Université Nazi BONI, Nasso, Burkina Faso, Email: sadouanouan@yahoo.fr
‡ Université Cheikh Anta Diop, Dakar, Senegal, Email: bamba.gueye@ucad.edu.sn
§ Université Joseph Ki Zerbo, Ouagadougou, Burkina Faso, Email: ouaro@yahoo.fr

***Abstract*—We propose a conceptual framework to predict the risk of freshwater source infestation by Schistosomiasis parasites. Our approach aims to combine two sources of information which are outputs of prediction models. The proposed framework is broken down into three Y-shaped branches. The left branch is a water quality prediction model built on the basis of machine learning algorithms applied on data collected by an IoT platform. These data represent physical and chemical parameters of a freshwater source which affect the development of snails and parasites that cause Schistosomiasis. The branch on the right is a non autonomous mathematical model which through its derived reproduction number R_0 determines the density evolution of all actors involved in Schistosomiasis transmission life cycle. In the middle branch happens a fusion process which combines the two information by taking into account their uncertainty and complementary.The output of the fusion is the final decision about the risk of infestation. This work has focused on the identification of applicable machine learning algorithms for water quality prediction and the identification of a mathematical model.The work has consisted also to give the characteristics of the fusion problem to handle.**

I. Introduction

In the area of health, early warning can help to save lives. With recent developments in computer technology, it is possible to build effective earlier warning systems for healthcare facilities thanks to the capabilities of the internet of things and artificial intelligence. Our objective is to propose as part of this work a system capable of alerting health structures in real time of the risk of infestation of fresh water sources (surface water) by parasites causing Schistosomiasis. Schistosomiasis is an acute and chronic parasitic disease caused by trematodes of the genus Schistosoma [1]. The larvae of the parasite, released by freshwater snails, enter a person's skin when in contact with infested water. The life cycle of the disease transmission involves freshwater sources, humans (final host), parasites and snails (vectors).

The cycle starts with infected people who contaminate freshwater sources with feces or urine containing parasite eggs. These eggs hatch in water, penetrate and develop in the bodies of snails until the formation of a parasite. After these parasites are released by snails and enter the body of other humans by skin contact with infested water. In the body, the larvae develop and pass to the stage of adult schistosomes that live in the blood vessels. Females lay eggs, some of which come out of the body through feces or urine. Others are trapped in the body's tissues, causing an immune response. People are mainly infected during routine agricultural, domestic, professional and recreational activities, which expose them to infested water.

Schistosomiasis is diagnosed by detecting parasite eggs in feces or urine [1]. The diagnosis can also be made by observation of the symptoms of the disease which can however take time before appearing on an infected person. Meanwhile, the undiagnosed infected person can continue to contaminate freshwater sources. The fight against Schistosomiasis is based on pharmaceutical treatment, the fight against snails, improved sanitation and health education. In our work we are interested in the fight against snails.

For this, we monitor the quality of surface freshwater to predict the conditions favorable for the reproduction and development of snails and parasites. For water quality monitoring we have deployed sensors that measure certain water parameters. The parameters monitored are those which favor the development of snails and parasites [2], [3]. These are namely PH, temperature, dissolved oxygen, electrical conductivity, flow, turbidity and total dissolved solid. We then observe the evolution of the snails population according to the variations of the water parameters. Taking into account the difficulties for a real estimation of the density of snails and parasites, we use a mathematical model which provides a qualitative evolution of this density.

Water quality prediction model can provide information about the favorable conditions of freshwater for the reproduction and development of snails and parasites but cannot assess the presence of the latter ones. Schistosomiasis mathematical model can respond to this limit but is not able to assess in real time the values of water parameters. To leverage the two sources of information we formulate in this paper a data fusion based approach to combine the results of two mentioned models.

The rest of the paper is organized as follows: section II presents some related works about water quality monitoring through sensor networks and estimating of snails and schistosoma parasites densities by mathematical models. We present also in this section some machine learning algorithms employed to predict water quality. Section III describes our elaborated approach based on IoT platform, mathematical modeling, machine learning, and data fusion. Section IV gives a conclusion and some perspectives to be addressed in future work.

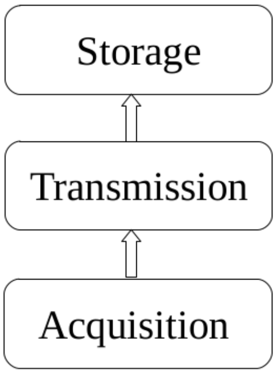


Fig. 1. Main components of a wireless sensor network

II. Related works

The conceptual framework proposed in the section III was inspired by some research areas that we have explored. It is about wireless sensor networks ($WSNs$) and machine learning (ML) employed in the monitoring and the prediction of water quality and mathematical modeling of infectious diseases especially models formulated for the water-related disease Schistosomiasis.

A. Deployment of wireless sensor network for monitoring water quality

Water quality can be defined as the suitability of water for a particular application based on its chemical, biological and physical characteristics. Monitoring water quality involves detecting its characteristic parameters and comparing them with set of standards and guidelines [4]. WSN is a possible technology employed to collect data for monitoring water quality as stressed in [5].

Before $WSNs$, conventional monitoring processes involve manual collection of samples from various points in the distribution network, followed by laboratory testing and analysis [4]. It is a laborious and time consuming process. $WSNs$ have since been considered as a promising alternative to complement conventional monitoring processes. These networks are relatively affordable and allow measurements to be taken remotely, in real-time and with minimal human intervention [4]. We can summarize this technology as a technology which involves three main components.

Fig. 1 shows the relationships between theses components. An acquisition component which consists of measuring water quality parameters in situ through a set of sensors. A transmission component which serves to the transmission of the measured data to a database directly or through a gateway which in its turn forwards the data to a database. A storage component which consists to store the measured data like a multivariate time series in a database.

[6] have deployed a WSN as an alternative way for the physicists at UCAD's Faculty of Science who still were using traditional means to collect water quality data from a pool in the faculty's Botanical Garden (dedicated to aquaculture).The built system is constituted of an acquisition node, a gateway and a database. The acquisition node is a set of four low cost sensors and a wireless transmission module. The 4 sensors measure pH, temperature, oxidation/reduction potential (ORP) and electrical conductivity (EC). The transmission module which is Lora-like sends measured data to a gateway (Dragino LG 01) which in its turn sends data through an Ethernet network to a database.

In [7] SmartCoast, a multi sensor for water quality monitoring has been designed. The parameters measured include temperature, phosphate, dissolved oxygen, conductivity, pH, turbidity and water level.

[8] have built a low cost WSN to address the water quality monitoring problem for drinking water distribution systems as well as for consumer sites. Five sensors interfaced to a PIC32 MCU based board measure turbidity, ORP, temperature, pH and EC and send data to a control node through zigbee technologoy. The control node stores received data in a local database and provides gateway to the Internet.

The WSN developped by [9] collects temperature, PH, turbidity, conductivity, dissolved oxygen. It is a low cost system built with a Raspberry PI B+ equipped with a WiFi module ($USR-WIFI232-X-V4.4$) which serves to transmit data into a online cloud. In the literature we have found many studies relative to WSN. A thorough survey on this area is provided in [4].

The main role of a WSN is data collecting as time series most of the time. Once the data are available, further analysis can be done such as detecting contamination event or predicting future values of water quality parameters. This can be ensured by other means such as machine learning algorithms.

B. Machine learning algorithms for predicting water quality

In [10], it is stressed that water quality prediction (WQP) is the forecasting of the variation trend of water quality at a certain time in the future. The main principle of WQP is the estimation of one or more water parameters values in a short or long term time followed by an evaluation of set of conditions as what has been done in [11], [12].

The prediction can be done by either conventional or traditional and artificial intelligence (AI) based methods [10], [13], [14]. The conventional or traditional methods are based on statistics. AI-based methods are relative to machine learning algorithms. In the literature, the benefits of AI-based methods are more highlight than conventional ones. [10] let know that traditional methods are not able to capture the non-linear and non-stationarity of water quality well due to their complex and sophisticated nature. They also let know that artificial neural network (ANN) models with their capacity to handle the non-linear and uncertain problems have become a hotspot in water quality research. One note that by ANN models, the authors [10] refer to four models architectures include feedforward (e.g multi-layer perceptron neural networks MLP), recurrent networks (e.g long short term memory $LSTM$), hybrid (e.g auto regressive integrated aoving average articificial neural network $ARIMA-ANN$) and emerging models (e.g depp belief network DBN).

In a review [14] on AI-based methods used to predict water quality, it is stressed that the limitations in the ability of traditional techniques such as multiple linear regression (MLR), multiple non-linear regression ($MNLR$), Mann-Kendall (MK) trend test and auto-regressive integrated moving average ($ARIMA$) models to accurately estimate the water quality in rivers due to the complexity and sophistication of the water quality time series led to the use of the black box methods such as the articial neural network ANN, neuro-fuzzy (NF), genetic programming (GP), and support vector machine (SVM).

In [15], three models have been developed to predict each three water quality parameters namely dissolved oxygen (DO), biochemical dissolved oxygen (BOD) and chemical dissolved oxygen (COD). The models include Multi-layer perceptron neural networks (MLP-ANN), ensemble neural networks ($E-ANN$) and support vector machine (SVM). SVM achieved more accurate predictions than $E-ANN$ and $MLP-ANN$.

[16] used SVM to predict separately weekly water quality parameters including DO, COD and ammonia nitrogen ($NH3-N$). The predictions have been after used as inputs of trained wavelet neural network (WNN) to forecast the whole status index of water quality. SVM achieved around 0.28 for DO mean average error (MAE) when used three 3 last weeks recording data and around 0.8 for 4 last weeks data.

Four ANN with non linear autoregressive ($ANN-NAR$) have been developed by [17] to predict four water quality namely chlorophyll, dissolved oxygen, turbidity and specific conductance. Based on performance measures such as regression, mean squared error (MSE) and root mean squared error ($RMSE$), the proposed model proves to be reliable with the lowest MSE being $3.7\text{x}10^{-4}$ for turbidity and the best regression value for EC (0.99).

Taking advantage of the good performance of $LSTM$ in time series prediction, a drinking-water quality model was designed and established to predict water quality by [18]. The dataset used to build this model is a two years collected dataset constituted of seven parameters namely PH, temperature, DO, EC, COD and $NH3-N$. The results of the study indicate that the predicted values of the model and the actual values were in good agreement and accurately revealed the future developing trend of water quality, showing the feasibility and effectiveness of using $LSTM$ deep neural networks to predict the quality of drinking water [18]. The $LSTM$ model performed better on two forecasting horizons ten days and six months than auto-regressive moving average ($ARIMA$) and support vector regression (SVR).

As we can note in [10], [14] many studies have been conducted about WQP. [10] and [14] respectively reviewed 151 papers published from 2008 to 2019 and 51 papers from 2000 to 2016. [14] presents a comprehensive investigation into the application of AI methods for modeling river water quality and offers a critical insight into the use and reliability of the various modeling approaches for modeling diverse water quality measurements. [10] have done an extensive investigation taking into account many architectures of ANN and many types of water quality parameters. They concluded that the ANN models are able to deal with different modeling problems in rivers, lakes, reservoirs, rivers, lakes, reservoirs, wastewater treatment plants ($WWTPs$), groundwater, ponds, and streams.

WSN associated to machine learning can help in the prediction of water quality. Detection or prediction of presence/density of certain species in a freshwater source such as snails and parasites is still difficult with the available sensors in the market to the best of our knowledge. We are interested however on the density evolution of snails and schistosoma which are involved in Schistosomiasis transmission life-cycle. To address this problem we refer to mathematical models which in the case of Schistosomiasis have been formulated to investigate the propagation of the disease and used to estimate the density evolution of snails and schistoma parasites.

C. Mathematical model for estimating the density evolution of snails and parasites

In [19] we can note that mathematical modeling of epidemics is the process of representing physical or biological phenomena in terms of equations or functions. In general the models are systems of differential or partial equations constituted of states variables and parameters. The state variables represent the different categories of populations studied most often subdivided in compartments and the parameters represent the transitions between the compartments or the interaction between them.

Many mathematical models are formulated using a various set of assumptions and can be resolved by numerical, graphical or analytic methods in the purpose to prove that they are reliable for predictions as [20] indicates in his paper. Since 1965 some mathematical models are formulated with the first one developed by MacDonalds to investigate the dynamics of Schistosomiasis [21]. [21] studied the relationship between the mean worm burden per person and the proportion of infectious snails.

[22] with a more realistic model has shown that the threshold parameter governing whether or not an endemic cycle can be established is closely related to the proportion of infected snails in a community. Until now many other researchers have also interested on mathematical modeling of Schistosomiasis with the aim of a preventive conclusion or control strategy as stressed by [19].

The models mainly differentiate themselves by the populations considered and the different factors (biological, environmental, etc) which govern Schistosomiasis' transmission. Some of them considered humans and snails populations [21]. Other in addition to these populations considered the worms populations subdivided into mirracidium and cercariae [23], [24]. In the attempt to be more realistic certain models take in account climate change as environmental factor [25], [26].

In the different models we have studied, a conceptual threshold known as the reproduction number R_0 is derived from the systems of differential equations. It permits to determinate the disease invasion and the plausibility to exterminate parasites. It also determines the density evolution of different populations studied. Indeed when its value exceeds 1, that means increasing is the variation of infected populations (humans and snails) and parasites (cercariae and mirracidium). Conversely when its value is less that 1, the variation of different populations considered is decreasing.

In [26], graphs obtained after numerical simulations showed that $R_0 < 1$ corresponds with decreasing populations

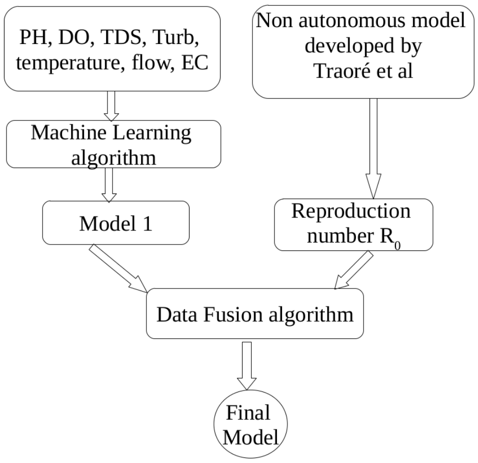


Fig. 2. The diagram of the conceptual framework.

of cercariae, infectious humans, miracidia and infectious snails. And $R_0 > 1$ corresponds with increasing populations of infectious humans, miracidia, infectious snails and cercariae. The derivation of R_0 and the determination of the density variation is done through parameters which can be obtained either through a laboratory experiment or through a statistical study or through a analytical resolution.

III. Presentation of our approach

WSN associated to machine learning and Schistosomiasis mathematical model are complementary. Indeed WSN collect data on which machine learning algorithms can be applied to predict water quality. Mathematical model provides information about snails and parasites density evolution. Our approach is to take advantage of the results of machine learning and mathematical models to build an efficient earlier warning system. The conceptual framework we propose involves internet of things (IoT), machine learning, mathematical modeling. To combine them we rely on data fusion methods. The framework is represented by the diagram shown in Fig. 2. We give it the name $FuMalMMO$ which means Fusion of Machine Learning and Mathematical' Models. .In the following sub sections we describe the different components of the conceptual framework.

A. IoT platform

The IoT platform is constituted of an acquisition device and a web platform. The acquisition device is an assembly of low cost sensors, Arduino-like microcontroller, and communication module. We rely on work done by [6] to realize the platform.

The sensors used serve to measure physical and chemical parameters namely pH, electrical conductivity, turbidity, temperature, water flow, dissolved oxygen, total solid dissolved. The communication module is an Arduino GSM shield 2 which allows the connection of the acquisition device to the Internet. Indeed the shield uses a radio modem $M10$ by Quectel which supports TCP/UDP and $HTTP$ protocols through a $GPRS$ connection [27]. We used a SIM card and a data plan provided by a local network operator to activate the access to the

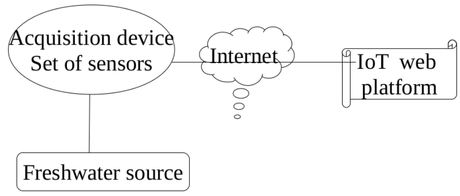
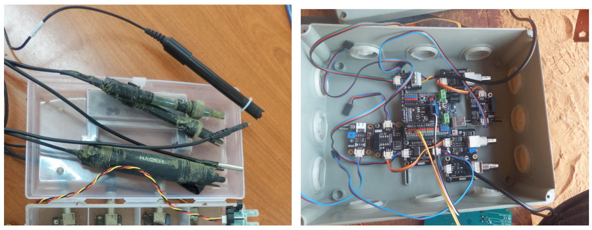


Fig. 3. IoT logical Architecture.



(a) Low cost sensors bought on a china market. (b) Sensors boards connected to an Arduino microcontroller.

(c) Panamasso backwater. (d) Installation realized.

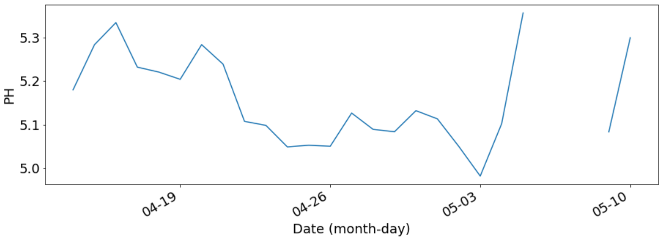
Fig. 4. Assembling and deployment of acquisition device.

Internet. Fig. 3 shows the logical architecture of the IoT platform.

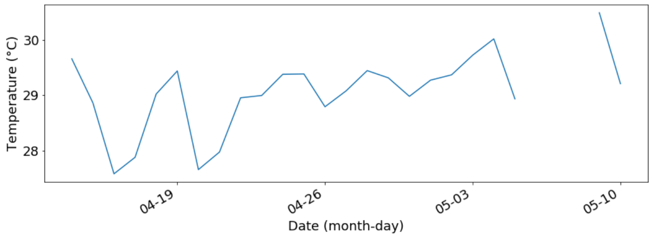
The device is placed since April 14 2020 in a backwater in Panamasso, a village located on Houet district in Burkina Faso. It is between latitude 11°23'0"North and longitude 4°12'0"West. It's an endemic zone of Schistosomiasis [26]. The device is equipped with solar panel which permits a 24h/24h working. The following pictures in Fig. 4a, Fig. 4b, Fig. 4c, Fig. 4d show respectively the different sensors used, the inside view of the box where they are connected and the installation realized.

The data measured by the device are stored regularly on a online web platform www.thinkspeak.com. The frequency of storage is 5 minutes. The platform offers a write and read API. Both API are RESTFUL. The write one serves to store data into the platform and the read API serves to retrieve data from the platform. It is also possible to export data to some specific files formats such as json, xml and csv (we chose this one). The collected data form a time series of j parameters an can be defined as:

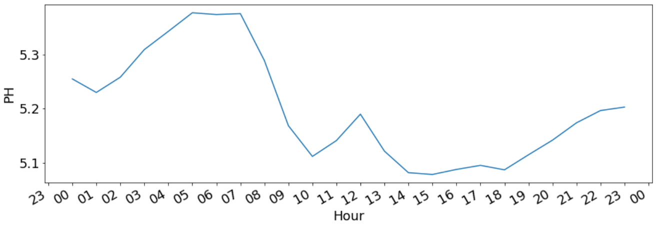
$$S_{i,n} = (\{x_{i,1,T_1}\}, ..., \{x_{i,n,T_n}\}) \quad (1)$$



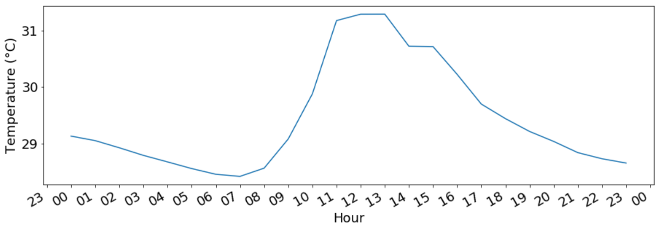
(a) Weekly Average PH of period April-May



(b) Weekly Average temperature of period April-May



(c) Hourly Average PH of day 2020-04-19



(d) Hourly Average temperature of day 2020-04-19

Fig. 5. Data collected of temperature and PH

where n represents the time series length, $x_{i,k}$ is the value of ith parameter measured at observation time T_k $(1 \leq i \leq j, 1 \leq k \leq n)$.

We show some graphics generated with these data. Fig. 5a and Fig. 5b represent respectively weekly average values of PH and temperature measured during April to May 2020. Fig. 5c, Fig. 5d represent respectively hourly average values of PH and temperature measured over 24h of day 2020-04-19. Despite the discontinuity noted in Fig. 5a and Fig. 5b, these graphics show the possibility to monitor on real time the freshwater source with the IoT platform. The discontinuity is probably due to a malfunctioning of the Arduino GSM shield. We didn't investigate deeply to track the reason of this malfunctioning. After a restarting of the acquisition device, things were working right again.

B. Machine Learning Algorithms

The machine learning algorithm purpose in our framework is to handle one part of the research problem which is to assess the suitability of a freshwater source to the development of parasites and snails based on a set of parameters. The general flow of steps to achieve this is as follow. Time series provided by IoT platform are used as input data. Not ignoring that missing data is often a common problem and same for noise which are present among collected data by WSN, a pre-processing step is needed. This step consists not only to handle noise (removing negative and null values) but also handle missing data through some methods such linear interpolation, improved mean value, etc. The pre-processing step consists also in normalization task which is necessary to equalize the importance of variables initially and to improve the interpretability of the network weights [14].

Model training is the following step after pre-processing. It consists to build the prediction model with an AI-based method using historical data of water parameters. [14] stressed that it is hard to draw a definitive conclusion about the performance of AI-based methods applied in WQP. Model structure selection discussed in [10] gave a similar conclusion. Nevertheless few recommendations have been formulated in both studies. As stated in section II-B these studies relative to the application of AI methods in WQP have concerned the reviewing of 51 papers published from 2000 to 2016 [14] and 151 papers from 2008 to 2019 [10]. The authors in [14] indicate that SVM can be considered as an appropriate method in WQP. Even it has already shown performing against certain ANN and SRC methods [15], [16], there is still a need of comparison of SVM with other AI methods.
Many studies have proved that the combined or hybrid models such as WANN are more accurate comparing to the single structure AI-models [28], [29], [30], [31], [32]. Based on this the authors in [14] recommended that more attention be given to the improvement of such models. In [10], it is stressed that RNN has good memory ability which allows its to make make full use of historical information and lay a solid foundation for realizing long-term prediction of water quality. LSTM which is a variant of RNN [18] showed to be more accurate and generalized than other methods in [33]. Based on these recommendations we are going to investigate SVM, LSTM and WANN.

The built prediction models are evaluated on the basis of root mean squared error ($RMSE$), mean absolute error (MAE) and coefficient of determination (R^2).$RMSE$ and MAE are errors measures which describe the difference between the model simulations and observations in the units of variables [34]. R^2 describes the proportion of the total variance in the observed data that can be explained by the model [34].

The general flow is ended by an inferring on the water suitability. This interfering is done through the step of the comparison of the predicted values with the thresholds of water parameters considered. [35] indicated that a PH range from 6.5 to 8.2, a temperature between 25°C and 28°C and irradiance upper $242umol/m^2/s$ are favourable breeding conditions for snails. [2] let know that some snails were collected in certain habitats with a DO concentration between $4.7mg/L$ and $11.4mg/L$. In [36] snails were collected in some locations with DO concentration between $2.7mg/L$ and $4.8mg/L$. Referring to the extreme values of DO mentioned in [2], [36], the tolerate values of DO for snails lay between $2.7mg/L$ and $11.4mg/L$.

The general flow of steps to predict water quality is shown

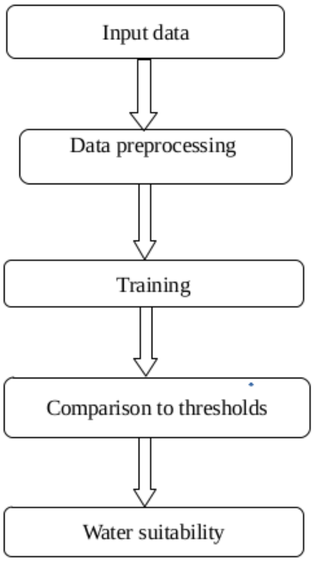


Fig. 6. WQP general flow of steps

in Fig. 6.

C. Mathematical model

The mathematical model in $FuMalMMO$ serves to complete the lack of information about snails and parasites densities that the IoT platform cannot provide. The model formulated by [26] provides this information. It is a non autonomous model which takes into account the effect of climate change on Schistosomiasis transmission. Climate change is reflected by the consideration of model parameters as time-dependent. It is an extension of the model formulated by [25] and wants to be more realistic though it is more difficult to analyze. The model is formulated following the transmission life-cycle of Schistosomiasis. Humans population is designated by H and M represents snails population. K and P designated respectively the density of cercariae and the density of mirracidia. Humans and snails populations are divided into two compartments : susceptible and infectious. H_s and M_s represent humans and snails in the susceptible compartment. H_i and M_i designate humans and snails in the infectious compartment.

A diagram showing the different compartments of populations and the interactions between them is presented in Fig. 7. The squares designate the different compartments of populations and the arrows designate the interactions. The solid arrows mean transitions between two compartments. When the arrow go to a compartment that means there is an adding in the compartment. Conversely the arrow going from a compartment means a withdraw. The blue dotted arrows indicate what the populations in compartment release or produce.

In the formulation of the model the squares represent the state variables and the arrows represent the parameters. We give a detailed description of different variables and parameters in the table I and the table II.

The authors have derived the basic reproduction ratio R_0 of the periodic model and have proved that it is the threshold between the extinction and the uniform persistence of the disease.

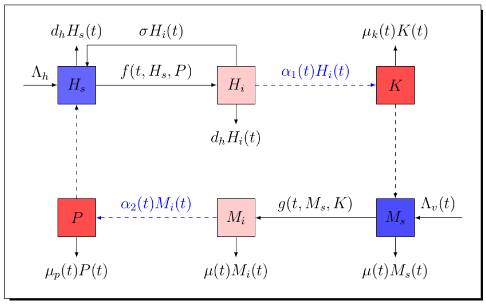


Fig. 7. Transmission life-cycle diagram

TABLE I. DESCRIPTION OF VARIABLES

Variable	Description
$H_s(t)$	Susceptible human host population
$H_i(t)$	Infectious human host population
$M_s(t)$	Susceptible snail population
$M_i(t)$	Infectious snail
$K(t)$	Density of miracidia
$P(t)$	Density of cercariae

TABLE II. DESCRIPTION OF PARAMETERS

Parameter	Description
Λ_h	Recruitment rate of human host
$B(t)$	Recruitment rate of snail
$\mu(t)$	Natural death rate of snails
$\mu_p(t)$	Natural death rate of cercariae
$\mu_k(t)$	Natural death rate of miracidia
$d_h(t)$	Natural death rate of human host
σ	Recovery rate of infectious human host
$\alpha_1(t)$	Production rate of miracidia by infected humans
$\alpha_2(t)$	Production rate of cercariae by infected snails
$f(t, H_s, P)$	Infection function of susceptible humans due to cercariae
$g(t, M_s, K)$	Infection function of susceptible snails due to miracidia

More precisely, they have shown that the periodic disease-free equilibrium E_p is globally asymptotically stable if $R_0 < 1$, whereas the disease is persistent if $R_0 > 1$. By analyzing graphics produced by the authors, it is showed that $R_0 < 1$ corresponds with low densities of cercariae, infectious humans, miracidia and infectious snails. And $R_0 > 1$ corresponds with increasing densities of infectious humans, miracidia, infectious snails and cercariae. One interesting thing in this study is the demonstration that basic reproduction ratio of non autonomous model is more accurate than autonomous model one.

D. Fusion description

In the literature many definitions of data fusion (DF) have been proposed. One of the popular and historical is given by the Joint Directors of Laboratories (JDL). JDL defines DF as “A process dealing with the association, correlation, and combination of data and information from single and multiple sources to achieve refined position and identity estimates, and complete and timely assessments of situations and threats as well as their significance” [37]. Others interesting definitions are given by some authors such as [38] which defines DF as a combination of information originating from several sources in order to improve decision making.

A DF process follows a certain scheme. The scheme determines the way the data are merged. It can be distributed or centralized [39]. In decentralized scheme, decisions are taken locally on each source and merge after into a global one. In the centralized scheme, no partial decision is taken. This scheme implies to gather all the information before taking the global one. There are some characteristics of information to be taken into in a fusion process. They are among others type (numeric versus symbolic), imperfection, redundancy and complementary of information. Imperfection can take different forms such as uncertainty, imprecision, incompleteness and so on [39]. We are not giving details of all them but just describe the ones which concern our fusion context. Uncertainty represents the adequation of an information to the reality [39]. Imprecision represents the quantitative defect of knowledge or measure [39]. Imprecision can cause uncertainty and reciprocally. Both are present simultaneous most often times. Information are complementary when each of them represent a specific part of the observed phenomenon.

Referring to our proposal framework, the fusion problem risen can be described as follow. The phenomenon observed is freshwater source state about its infestation by Schistosomiasis parasites. Two different aspects are considered. The first one is the water quality for the reproduction of snails and parasites. The second concerns the density evolution of these entities. Two independent sources are available and provide information on these aspects. The fusion scheme suited to the problem is the centralized one. The global decision is taken based on all the information provided by the two models. Uncertainty, numeric and complementary are the characteristics identified to take into account in the fusion process. Uncertainty is due in one hand to the fact that data collected by low cost sensors can contain noise caused by the instability of sensors. In other hand mathematical models in general are not intend to represent faithfully the reality and some parameters used in the numerical simulation are estimated. Complementary is due to the fact that none of two considered models can give at the same time information about water quality and density evolution of disease's involved actors.

The fusion problem can be posed as :

$$\begin{array}{cc} & \begin{array}{cc} S_1 & S_2 \end{array} \\ \begin{array}{c} d_1 \\ d_2 \end{array} & \begin{bmatrix} M_1^1(x) & M_2^1(x) \\ M_1^2(x) & M_2^2(x) \end{bmatrix} \end{array}$$

where $S_j \ {}_{1\leq j\leq 2}$ represents the two models considered. They provide information $M_j^i(x)$ which relate x (freshwater source state) to each possible decision $d_i \ {}_{1\leq i\leq 2}$ according to each source S_j. d_i belongs to a decision space $D = \{d_1, d_2\}$ with $d_1 = infested$ and $d_2 = notinfested$. The final decision on observation x is taken from the combination all of $M_j^i(x)$ ${}_{1\leq j\leq 2} \ {}_{1\leq i\leq 2}$.

$M_j^i(x)$ evaluates how much a source j supports the decision i [39]. It can be represented in different ways such as conditional probabilities, membership degrees, possibility degrees, plausibility or belief functions. This depends on the chosen approach. Many approaches have been proposed to handle fusion problems. They can be gathered mainly into probabilistic and fuzzy approaches. Probabilistic approaches can be subdivided into Bayesian and Evidential approaches. Fuzzy approaches come from Possibility theory. [40] have mentioned also Machine learning Algorithms as possible methods.

There are four main steps in a fusion process [39] to respect whatever the considered approach. The first one is the modeling of information and its imperfections. Modeling consists to find the right representation of $M_j^i(x)$. The second step is the estimation of information relating to each possible decision according to each source. The third step is combination. This consists to gather information by choosing an appropriate combination operator. The fourth step is relative to the choice of decision criteria. Classically it is about minimization or maximization of a function coming from combination step.

Once the fusion problem is characterized, it is necessary to experiment the mainly approaches mentioned above to figure out which one can handle better all its characteristics. Further work in this direction is called for and is currently being carried out.

IV. Conclusion and Perspectives

Despite many efforts employed to control Schistosomiasis, it still persists. But with the technologies of computer science, there exist some means that we can employ to propose a solution to this public health problem. In this work in progress, we propose a conceptual framework to predict a freshwater source suitability to the transmission of the Schistosomiasis.

Two models are involved in the framework. One serves to assess if a freshwater source is favorable or not to the reproduction and the development of snails and parasites. It is a prediction model to build with machine learning algorithm applied on data collected (physical and chemical parameters) with an IoT platform. To complement the lack of information on the presence of the snails and parasites, this model is fused to a mathematical model formulated by [26] which provides the trend of the variation of snails and parasites densities.

An IoT platform is deployed at Panamasso backwater to collect some physical and chemical parameters. The machine learning algorithms to build the prediction model are identified. It's about SVM, $LSTM$ and $WANN$. The mathematical model which provides the complementary information is also determined.

As next tasks, we are going to implement and compare models built with the machine learning algorithms. We are also going to elaborate the different stages of the fusion such as the modeling, the estimation, the combination and the decision. The modeling consists to the choice of the formalism to represent the data. The estimation depends on the modeling and serves to estimate the data. The combination consists to choose the right operator. And the decision consists to maximize or minimize the result of the combination.

References

[1] "Schistosomiasis." [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/schistosomiasis

[2] E. Abdel Malek, "Factors conditioning the habitat of bilharziasis intermediate hosts of the family Planorbidae," *Bulletin of the World Health Organization*, vol. 18, no. 5-6, pp. 785–818, 1958.

[3] S. Bakhoum, R. A. Ndione, C. J. E. Haggerty, C. Wolfe, S. Sow, C. T. Ba, G. Riveau, and R. R. Jason, "Influence des paramètres physico-chimiques sur la répartition spatiale des mollusques hôtes intermédiaires des schistosomes humains dans le delta du fleuve sénégal," *Médecine et Santé Tropicales*, vol. 29, no. 1, pp. 61–67.

[4] M. Pule, A. Yahya, J. Chuma, M. Pule, A. Yahya, and J. Chuma, "Wireless sensor networks: A survey on monitoring water quality," *Journal of applied research and technology*, vol. 15, no. 6, pp. 562–570, 2017.

[5] E. M. Dogo, N. I. Nwulu, B. Twala, and C. Aigbavboa, "A survey of machine learning methods applied to anomaly detection on drinking-water quality data," *Urban Water Journal*, vol. 16, no. 3, pp. 235–248, 2019.

[6] B. Ngom, M. Diallo, B. Gueye, and N. Marilleau, "LoRa-based Measurement Station for Water Quality Monitoring: Case of Botanical Garden Pool," in *2019 IEEE Sensors Applications Symposium (SAS)*. Sophia Antipolis, France: IEEE, Mar. 2019, pp. 1–4.

[7] B. O'Flynn, R. Martinez-Catala, S. Harte, C. O'Mathuna, J. Cleary, C. Slater, F. Regan, D. Diamond, and H. Murphy, "SmartCoast: A Wireless Sensor Network for Water Quality Monitoring," in *32nd IEEE Conference (LCN 2007)*, Oct. 2007, pp. 815–816.

[8] T. P. Lambrou, C. C. Anastasiou, C. G. Panayiotou, and M. M. Polycarpou, "A Low-Cost Sensor Network for Real-Time Monitoring and Contamination Detection in Drinking Water Distribution Systems," *IEEE Sensors Journal*, vol. 14, no. 8, pp. 2765–2772, Aug. 2014, conference Name: IEEE Sensors Journal.

[9] N. Vijayakumar and R. Ramya, "The real time monitoring of water quality in IoT environment," in *2015 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS)*, Mar. 2015, pp. 1–5.

[10] Y. Chen, L. Song, Y. Liu, L. Yang, and D. Li, "A review of the artificial neural network models for water quality prediction," *Applied Sciences*, vol. 10, no. 17, p. 5776.

[11] T. Jin, S. Cai, D. Jiang, and J. Liu, "A data-driven model for real-time water quality prediction and early warning by an integration method," *Environmental Science and Pollution Research*, vol. 26, no. 29, pp. 30 374–30 385.

[12] B. Kassé, B. Gueye, M. Diallo, F. Santatra, and H. Elbiaze, "IoT based Schistosomiasis Monitoring for More Efficient Disease Prediction and Control Model," in *2019 IEEE Sensors Applications Symposium (SAS)*, Mar. 2019, pp. 1–6.

[13] G. Kang, J. Z. Gao, and G. Xie, "Data-driven water quality analysis and prediction: A survey," in *2017 IEEE Third International Conference on Big Data Computing Service and Applications (BigDataService)*, pp. 224–232.

[14] T. Rajaee, S. Khani, and M. Ravansalar, "Artificial intelligence-based single and hybrid models for prediction of water quality in rivers: A review," *Chemometrics and Intelligent Laboratory Systems*, vol. 200, p. 103978.

[15] A. Najah, A. El-Shafie, O. A. Karim, O. Jaafar, and A. H. El-Shafie, "An application of different artificial intelligences techniques for water quality prediction," *International Journal of Physical Sciences*, vol. 6, no. 22, pp. 5298–5308.

[16] L. Zhang, Z. Zou, and W. Shan, "Development of a method for comprehensive water quality forecasting and its application in miyun reservoir of beijing, china," *Journal of Environmental Sciences*, vol. 56, pp. 240–246.

[17] Y. Khan and C. S. See, "Predicting and analyzing water quality using machine learning: A comprehensive model," in *2016 IEEE Long Island Systems, Applications and Technology Conference (LISAT)*, pp. 1–6.

[18] P. Liu, J. Wang, A. K. Sangaiah, Y. Xie, and X. Yin, "Analysis and Prediction of Water Quality Using LSTM Deep Neural Networks in IoT Environment," *Sustainability*, vol. 11, no. 7, pp. 1–14, 2019, publisher: MDPI, Open Access Journal.

[19] M. Diaby, "Global analysis of epidemic models: Schistosomiasis case," Theses, Université Gaston Berger de Saint-Louis (SENEGAL), Jan. 2016.

[20] M. L. Diagne, "Modélisation et étude mathématique de la dynamique de prolifération du Typha dans le Parc National des Oiseaux de Djoudj (PNOD)," p. 160.

[21] G. Macdonald, "The dynamics of helminth infections, with special reference to schistosomes," *Transactions of the Royal Society of Tropical Medicine and Hygiene*, vol. 59, no. 5, pp. 489–506, Sept. 1965.

[22] A. D. Barbour, "Macdonald's model and the transmission of bilharzia," *Transactions of the Royal Society of Tropical Medicine and Hygiene*, vol. 72, no. 1, pp. 6–15, Jan. 1978.

[23] E. T. Chiyaka and W. Garira, "Mathematical analysis of the transmission dynamics of schistosomiasis in the human-snail hosts," *Journal of Biological Systems*, vol. 17, no. 03, pp. 397–423, Sept. 2009.

[24] Z. Feng, A. Eppert, F. A. Milner, and D. J. Minchella, "Estimation of parameters governing the transmission dynamics of schistosomes," *Applied Mathematics Letters*, vol. 17, no. 10, pp. 1105–1112, Jan. 2004.

[25] Y. Li, Z. Teng, S. Ruan, M. Li, and X. Feng, "A mathematical model for the seasonal transmission of schistosomiasis in the lake and marshland regions of China," *Mathematical biosciences and engineering: MBE*, vol. 14, no. 5-6, pp. 1279–1299, Dec. 2017.

[26] B. Traoré, O. Koutou, and B. Sangaré, "Global dynamics of a seasonal mathematical model of schistosomiasis transmission with general incidence function," *Journal of Biological Systems*, vol. 27, no. 01, pp. 19–49, Feb. 2019, publisher: World Scientific Publishing Co.

[27] "Arduino GSM Shield 2 (Integrated Antenna)." [Online]. Available: https://store.arduino.cc/arduino-gsm-shield-2-integrated-antenna

[28] T. Partal and H. K. Cigizoglu, "Estimation and forecasting of daily suspended sediment data using wavelet–neural networks," *Journal of Hydrology*, vol. 358, no. 3, pp. 317–331.

[29] T. Rajaee, V. Nourani, M. Zounemat-Kermani, and O. Kisi, "River suspended sediment load prediction: Application of ANN and wavelet conjunction model," *Journal of Hydrologic Engineering*, vol. 16, no. 8, pp. 613–627, publisher: American Society of Civil Engineers.

[30] T. Rajaee, "Wavelet and ANN combination model for prediction of daily suspended sediment load in rivers," *Science of The Total Environment*, vol. 409, no. 15, pp. 2917–2928.

[31] M. Ravansalar and T. Rajaee, "Evaluation of wavelet performance via an ANN-based electrical conductivity prediction model," *Environmental Monitoring and Assessment*, vol. 187, no. 6, p. 366.

[32] M. Ravansalar, T. Rajaee, and M. Ergil, "Prediction of dissolved oxygen in river calder by noise elimination time series using wavelet transform," *Journal of Experimental & Theoretical Artificial Intelligence*, vol. 28, no. 4, pp. 689–706.

[33] Y. Wang, J. Zhou, K. Chen, Y. Wang, and L. Liu, "Water quality prediction method based on LSTM neural network," in *2017 12th International Conference on Intelligent Systems and Knowledge Engineering (ISKE)*, pp. 1–5.

[34] D. R. Legates and G. J. McCabe, "Evaluating the use of "goodness-of-fit" measures in hydrologic and hydroclimatic model validation," *Water Resources Research*, vol. 35, no. 1, pp. 233–241.

[35] B. Kasse, B. Gueye, M. Diallo, F. Santatra, and H. Elbiaze, "IoT based Schistosomiasis Monitoring for More Efficient Disease Prediction and Control Model," in *2019 IEEE Sensors Applications Symposium (SAS)*. Sophia Antipolis, France: IEEE, Mar. 2019, pp. 1–6.

[36] G. Y. Félicien, D. N. Rose, A. O. Faustin, and K. Eliézer, "Dynamique des populations de mollusques hôtes intermédiaires deSchistosoma haematobium et schistosoma mansoni dans le lac du barrage de taabo (sud côte d'ivoire)." no. 3, p. 15.

[37] F. E. White, "Data fusion lexicon," Tech. Rep., oct 1991.

[38] B. I. M. (H.), *01 - Fusion de données en traitement d'images: modèles d'information et décisions*. GRETSI, Saint Martin d'Hères, France, 1994, vol. 11.

[39] B. Isabelle, "Fusion of Image Information under Imprecision and Uncertainty: Numerical Methods," in *Data Fusion and Perception*, 2001, pp. 135 – 168, vol. 431.

[40] F. Alam, R. Mehmood, I. Katib, N. N. Albogami, and A. Albeshri, "Data fusion and iot for smart ubiquitous environments: A survey," *IEEE Access*, vol. 5, pp. 9533–9554, 2017.