An AI-based approach to the prediction of water points quality indicators for schistosomiasis prevention

Teegwende Zougmore [1] Bamba Gueye [2] Sadouanouan Malo [3]

[1]Université Alioune Diop de Bambey [2]Université Cheikh Anta Diop [3]Université Nazi Boni

Context

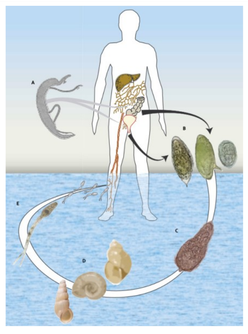


Figure 1. Life cycle of Schisotomiasis

Schistosomiasis

Parasitic disease affecting rural populations when they carry out certain domestic activities: laundry, cattle breeding, etc.

Neglected Tropical Disease

- Necessary reduction in prevalence nevertheless
- Proposal of control approaches
- Endemicity still observed
  - 251.4 million people needed preventive treatment in 2021
  - second cause of hospitalization in Senegal

Problem statement

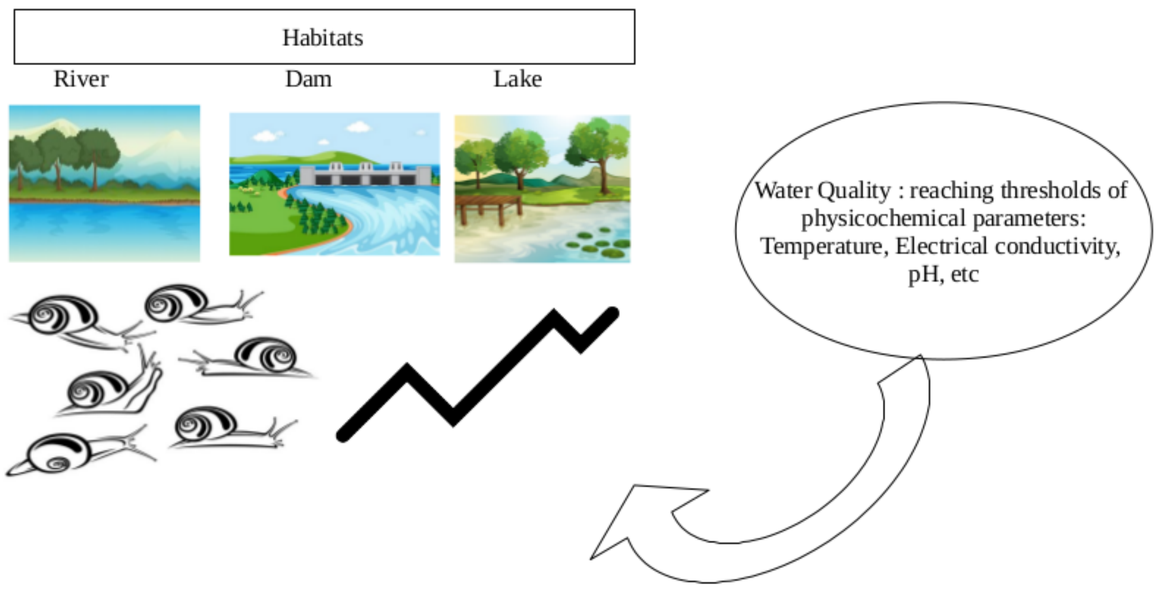


Figure 2. Influence of water quality on the biological cycle of intermediate hosts

How to forecast that a water point will be infested?

- How to forecast the water quality of the water point?
- How to forecast the evolution of the density of infected molluscs?
- How to take these two forecasts into account to infer the state of infestation from the water point?

Addressed Question

Forecast accurately water quality favourable to the development of snails and parasites which cause schistosomiasis

Structure of proposed system and used AI methods backgrounds

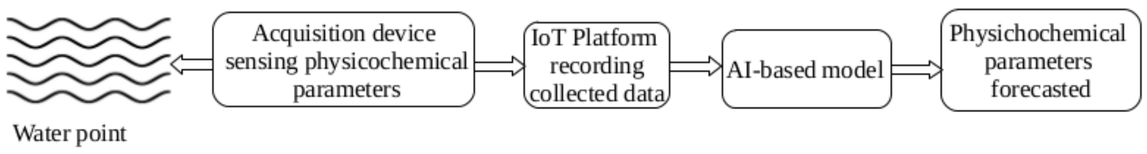
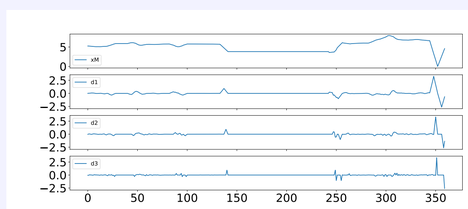


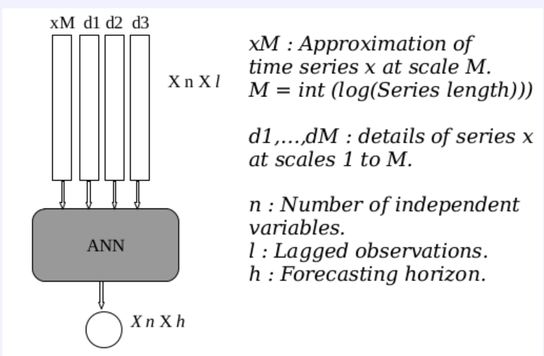
Figure 3. Structure of proposed system

Wavelet Artificial Neuronal Network (WANN)

Time series decomposition



Combination of wavelet transform and ANN

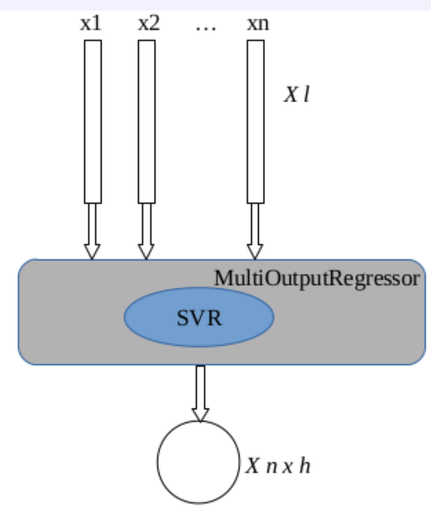


xM : Approximation of time series x at scale M.
M = int (log(Series length)))

d1,...,dM : details of series x at scales 1 to M.

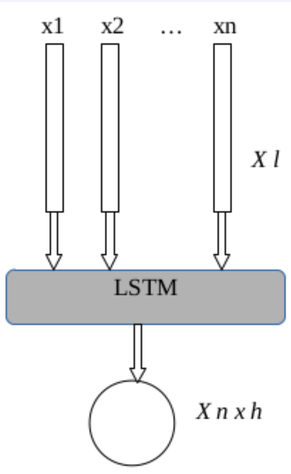
n : Number of independent variables.
l : Lagged observations.
h : Forecasting horizon.

Support Vector Regression (SVR)



x1,..,xn : Time series
n : Number of independent variables.
l : Lagged observations.
h : Forecasting horizon.

Long Short Term Memory (LSTM)



x1,..,xn : Time series
n : Number of independent variables.
l : Lagged observations.
h : Forecasting horizon.

Models training steps and hyperparameters determination

Models building steps

- Data pre-processing : Values removal and re-sampling
- Data preparation

DATE	TEMP	PH
2020-04-14 17:48:00	30.00	5.18
2020-04-14 17:53:32	29.94	5.20
2020-04-14 17:59:04	29.88	5.20
2020-04-14 18:04:36	29.88	5.18
2020-04-14 18:10:08	29.88	5.20

PH $(t-2)$	TEMP $(t-2)$	PH $(t-1)$	TEMP $(t-1)$	PH (t)	TEMP (t)	PH $(t+1)$	TEMP $(t+1)$
5.18	29.66	5.28	28.86	5.33	27.58	5.23	27.88
5.28	28.86	5.33	27.58	5.23	27.88	5.22	29.02
5.33	27.58	5.23	27.88	5.22	29.02	5.20	29.44

- Data splitting in sets
  - Training : April 2020 - February 2021
  - Validation : March 2021 - April 2021
  - Test : May 2021 - July 2021

Figure 4. Times series transformed to a supervised learning problem format

Hyperparameters

Table 1. Number of nodes of input and output layers

Method	Input Layer	Output Layer
$WANN$	$(i+1)*n$	$H*n$
$LSTM$	$L*n$	$H*n$

H: Forecasting horizon; L :Lag length; n : number of series; i : Level of decomposition.
Epochs and nodes of hidden layers = 100

Table 2. SVR hyperparameters

Kernel	ϵ	C	Γ
Gaussian	0.001	5	0.001

Γ : kernel's width; ϵ : tube's width; C : regularization parameter

Evaluation

Study area and data

Figure 5. Panamasso : Schistosomiasis's endemic area

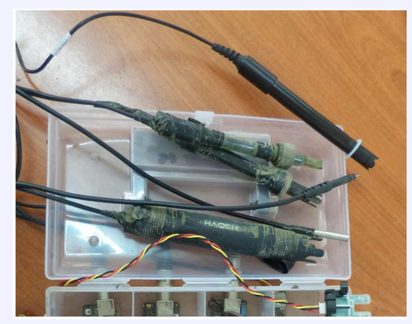


Figure 6. Acquisition device

Figure 7. Backwater view

Collected data

15 months (from April 2020 to July 2021) data of pH, temperature, dissolved oxygen, electrical conductivity

Results

Table 3. Models performances

Metric	PH	Temp	OD	EC	Model
RMSE	0.07	0.13	0.09	9.79	WANN L2_H1
MAE	0.05	0.06	0.06	7.15	
RMSE	1.04	1.00	0.34	37.40	LSTM L2_H1
MAE	0.62	0.73	0.22	27.63	
RMSE	0.91	1.06	0.02	11.40	SVR L2_H1
MAE	0.53	0.71	0.01	8.45	

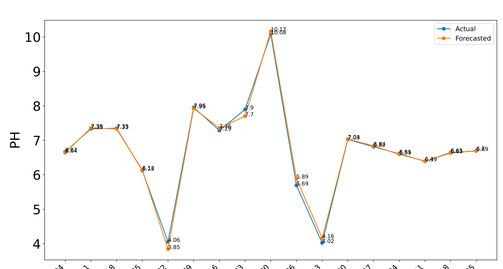


Figure 8. pH forecasting test with WANN L2_H1

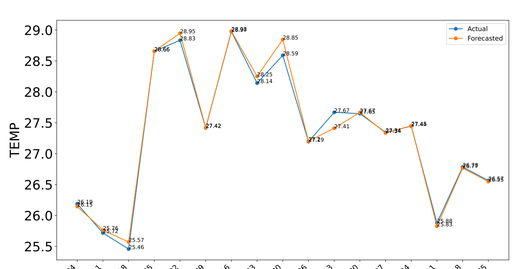


Figure 9. Temperature forecasting test with WANN L2_H1

References

[1] 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.

[2] 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.

teegwend@gmail.com