Infectious Disease Modelling 9 (2024) 1081–1094

Contents lists available at ScienceDirect

Infectious Disease Modelling

journal homepage: www.keaipublishing.com/idm

Schistosomiasis transmission in Zimbabwe: Modelling based on machine learning

Hong-Mei Li [a], Jin-Xin Zheng [a], Nicholas Midzi [b], Masceline Jenipher Mutsaka-Makuvaza [b, c], Shan Lv [a], Shang Xia [a], Ying-jun Qian [a], Ning Xiao [a], Robert Berguist [d], Xiao-Nong Zhou [a, *]

[a] National Institute of Parasitic Diseases, Chinese Center for Disease Control and Prevention (Chinese Center for Tropical Diseases Research), NHC Key Laboratory of Parasite and Vector Biology, WHO Collaborating Centre for Tropical Diseases, National Center for International Research on Tropical Diseases, Shanghai, 200025, China
[b] National Institute of Health Research, Ministry of Health and Child Care, Harare, Zimbabwe
[c] University of Rwanda, College of Medicine and Health Sciences, School of Medicine and Pharmacy, Department of Microbiology and Parasitology, Rwanda
[d] Ingerod 407, SE-454 94, Brastad, Sweden

ARTICLE INFO

Article history:
Received 19 November 2023
Received in revised form 2 June 2024
Accepted 8 June 2024
Available online 12 June 2024
Handling Editor: Dr Yijun Lou

Keywords:
Machine-learning
Transmission risk model
Schistosomiasis
Zimbabwe

ABSTRACT

Zimbabwe, located in Southern Africa, faces a significant public health challenge due to schistosomiasis. We investigated this issue with emphasis on risk prediction of schistosomiasis for the entire population. To this end, we reviewed available data on schistosomiasis in Zimbabwe from a literature search covering the 1980-2022 period considering the potential impact of 26 environmental and socioeconomic variables obtained from public sources. We studied the population requiring praziquantel with regard to whether or not mass drug administration (MDA) had been regularly applied. Three machine-learning algorithms were tested for their ability to predict the prevalence of schistosomiasis in Zimbabwe based on the mean absolute error (MAE), the root mean squared error (RMSE) and the coefficient of determination (R^2). The findings revealed different roles of the 26 factors with respect to transmission and there were particular variations between *Schistosoma haematobium* and *S. mansoni* infections. We found that the top-five correlation factors, such as the past (rather than current) time, unsettled MDA implementation, constrained economy, high rainfall during the warmest season, and high annual precipitation were closely associated with higher *S. haematobium* prevalence, while lower elevation, high rainfall during the warmest season, steeper slope, past (rather than current) time, and higher minimum temperature in the coldest month were rather related to higher *S. mansoni* prevalence. The random forest (RF) algorithm was considered as the formal best model construction method, with MAE = 0.108; RMSE = 0.143; and R^2 = 0.517 for *S. haematobium*, and with the corresponding figures for *S. mansoni* being 0.053; 0.082; and 0.458. Based on this optimal model, the current total schistosomiasis prevalence in Zimbabwe under MDA implementation was 19.8%, with that of *S. haematobium* at 13.8% and that of *S. mansoni* at 7.1%, requiring annual MDA based on a population of 3,003,928. Without MDA, the current total schistosomiasis prevalence would be 23.2%, that of *S. haematobium* 17.1% and that of *S. mansoni* prevalence at 7.4%, requiring annual MDA based on a population of 3,521,466. The study reveals that MDA alone is insufficient for schistosomiasis elimination, especially that due to *S. mansoni*. This study predicts a

* Corresponding author.
E-mail address: zhouxn1@chinacdc.cn (X.-N. Zhou).
Peer review under responsibility of KeAi Communications Co., Ltd.

https://doi.org/10.1016/j.idm.2024.06.001

moderate prevalence of schistosomiasis in Zimbabwe, with its elimination requiring comprehensive control measures beyond the currently used strategies, including health education, snail control, population surveillance and environmental management.

1. Introduction

Schistosomiasis is basically a chronic disease caused by schistosome parasite infections that were first discovered in Egypt by Theodore Bilharz in 1851 (Hagan, 2009). World Health Organization (WHO) classifies it as one of the neglected tropical diseases (NTD) (Utzinger et al., 2009). Depending on parasite species and site of infection, schistosomiasis appears as one of two types: intestinal or urinary schistosomiasis, with the former caused by infection either by *Schistosoma mansoni*, *S japonicum*, *S. mekongi* or *S. intercalatum*, whereas urinary schistosomiasis is only caused by *S. haematobium* (World Health Organization., 2023). The clinical signs of the former are abdominal pain, diarrhoea and bloody stools that if not treated eventually lead to hepatosplenomegaly and ascites by ongoing re-infections if not treated. The typical symptom of urinary schistosomiasis is haematuria, with serious complications in the long run, including kidney failure, bladder cancer and/or prostate cancer (Choto et al., 2020). Women infected with urinary schistosomiasis also are at increased risk of human immunodeficiency virus (HIV) infection (Bustinduy et al., 2014). In children, schistosomiasis can cause anaemia, stunted growth, and even death (Ekpo et al., 2012). WHO, in its second NTD Global Roadmap for 2021–2030 (World Health Organization., 2021), specifies elimination of schistosomiasis as one of its targets.

The life cycle of this parasite includes an intermediate freshwater snail host and transmission from snail to human that occurs when individuals encounter water contaminated with *Schistosoma* cercariae, the infective stage of schistosoma released from the intermediate snail host. This transmission mechanism may seem straightforward but is in fact complex as snail presence and level of infections are strongly influenced by the climate and other environmental factors (Hu et al., 2017; Yang & Bergquist, 2018; Zhou et al., 2008) as well as socioeconomic factors (Gong et al., 2021; Liu et al., 2023; Mutsaka-Makuvaza et al., 2018). Schistosomiasis is the second most prevalent parasitic disease in Africa after malaria (Isaiah et al., 2023). Approximately 250 million people worldwide required preventive treatment for schistosomiasis in 2021, but only 75.3 million actually received it, more than 90% of all patients in need of preventive chemotherapy live in Africa (World Health Organization., 2023).

Even if sporadic reports on *S. intercalatum* in Zimbabwe have been published (Kolodziej et al., 2023), *S. haematobium* and *S. mansoni* are in principle the only two species found there. The disease is of major public health concern (Chimbari, 2012) and two national surveys have been conducted in 1981 and 2012, respectively. The first one targeted school-aged children (8–10 years old), whose infection rates for *S. haematobium* ranged from 0% to 97%, with a mean of 41%, while for *S. mansoni* they ranged from 0% to 81%, with a mean of 8% (Taylor & Makura, 1985). The second survey targeted older school-aged children (10–15 years old) and reported infection rates for *S. haematobium* ranging from 0% to 76%, with a mean of 18%, while for *S. mansoni* they ranged from 0% to 64%, with a mean of 7.2% resulting in an overall schistosomiasis infection rate of 22.7% (Midzi et al., 2014). Since then, Zimbabwe has implemented a national schistosomiasis control programme, mainly involving a plan of six years of mass drug administration (MDA) once annually, by praziquantel as preventive treatment for children aged 6–15 years. Treatment is school-based, which means that only the youngest pupils received all six rounds of MDA, with the older ones dropping out of the scheme when leaving school after becoming 15 years old. After the six rounds of MDA, the prevalence of schistosomiasis among school-aged children in the project areas decreased to <1% (Mduluza et al., 2020). Although these data provide evidence of the distribution of schistosomiasis in Zimbabwe, the surveys and the treatment strategy mainly targeted school-aged children, so they do not reflect the disease burden in the entire population.

Identifying high-risk areas in the entire population and accurately predicting disease risk would allow precision interventions, including targeted or prompted treatment, improvements in snail control, quality monitoring activities of water, sanitation and hygiene (WASH) (Bartlett et al., 2022). Praziquantel was recommended by WHO to be the essential medicine for the treatment of schistosomiasis. Praziquantel is effective on adult schistosomas but less susceptible to juvenile worms (Agniwo et al., 2023; Diop et al., 2023). In addition, global donations of praziquantel to Africa are limited, so we need to better identify areas at risk and use preventive medicines more effectively. In many cases, this can be done by automatically analysing data to discover patterns and using these patterns to make predictions. The advantage with an approach based on both available and unseen data is that predictions can be made without necessarily having to engage in enlarged sampling. Machine-learning (ML) technology (Keshavamurthy et al., 2022; Mooney & Pejaver, 2018; Rampogu, 2023; Vilne et al., 2019) is a most useful way in this context. It involves designing and analysing algorithms that enable computers to learn automatically. Unlike statistical methods, in which variable relationships are explicitly defined, ML models can achieve accurate predictions more rapidly and have therefore become popular with regard to infectious diseases (Bergquist et al., 2024; Lu et al, 2022; Rampogu, 2023). This approach has been used to identify high-risk areas for *S. japonicum* transmission (Gong et al., 2021) and the parasite's intermediate host in China *Oncomelania hupensis* (Liu et al., 2023; Zheng et al., 2021).

The previous national surveys in Zimbabwe primarily focused on schistosomiasis among school-aged children, overlooking those of middle age and the elderly. In an effort to cover the entire country, we collected all available epidemiological data on the disease in Zimbabwe over the past 40 years. We also collected the corresponding environmental and socioeconomic indicators, and used this information to construct a model for transmission risk predicting the current epidemic risk under two different MDA implementation scenarios. We aimed to explore the distribution of both *S. haematobium* and *S. mansoni* to assess the disease burden and the needed, required treatment for the country as a whole.

2. Methods

2.1. Study area

This study focused on distribution of *S. haematobium* and *S. mansoni* covering whole country of Zimbabwe, a landlocked country in southern Africa. A digital map of Zimbabwe was downloaded from the Database of Global Administrative Areas (GADM, https://gadm.org) and used for orientation.

2.2. Data collection

2.2.1. Epidemiological data

Schistosomiasis data with human infection rates were collected through literature searches, personal information and from the Zimbabwe Public Health website (http://www.mohcc.gov.zw/). We conducted literature searches from four databases (PubMed, Web of Science, ScienceDirect, and African Journals Online) that were searched using the following keywords: '*Schistosoma mansoni*', '*Schistosoma haematobium*', 'schisto*', 'schistosomiasis', '*Schistosoma*', together with 'Zimbabwe'.

This study applied stringent inclusion criteria. References were only included if they featured epidemiological surveys conducted in Zimbabwe providing data on *S. mansoni* or *S. haematobium* infections within the population. Additionally, data extraction for all surveyed individuals including overall prevalence was required. Only references published in English were included. The earliest publication included was from 1 January 1980 and the final search was conducted on 17 April 2023. Articles lacking detailed survey site information, offering only national- or provincial-level data, were excluded and so were those not distinguishing between schistosome species.

Endnote X8 software was used to organize the references. After removing duplicate references, titles and abstracts were reviewed excluding irrelevant literature, with full texts read when in doubt. The following information was extracted from the selected references: literature identification, survey location, sex and age of study subjects, survey period, survey methods, schistosome species, detection methods, total number of tested individuals, number of infected individuals, infection rate and information on preventive chemotherapy for the population. Missing latitude and longitude information was geolocated using Google Maps. Excel software was used to record the data.

2.2.2. Environmental and socioeconomic data

The environmental and socioeconomic variables involved in the transmission of schistosomiasis and its intermediate host snails are listed in Table 1. They were all obtained from several public databases offering remotely sensed data (see bottom part of Table 1). The R software (version 4.2.1, R Foundation for Statistical Computing, Vienna, Austria) was used to collect and extract all environmental and socioeconomic raster data corresponding to the survey points and then resample and crop them into the same spatial areas. All data were resampled and cropped to the spatial resolution of 5×5 km.

The 19 global bioclimatic variables (Hijmans et al., 2005; Fick and Hijmans, 2017) were obtained from the Global Climate Database (http://www.worldclim.com). Elevation and slope were obtained from the Shuttle Radar Topography Mission Digital Elevation Model (SRTM) by the U. S. National Aeronautics and Space Administration (NASA). Water body data came from the global water body dataset that also originated from STRM and then the nearest water distance was calculated from each grid cell to the nearest water body in the rasterised representation by using the 'distance' function in the 'raster' package in R software. The normalized difference vegetation index (NDVI) was obtained from the Moderate Resolution Imaging Spectroradiometer (MODIS) onboard the Terra Satellite managed by NASA.

Population data were sourced from WorldPop (http://www.worldpop.org.uk/), while information on the gross domestic product (GDP) and the Human influence index (HII), a global 1-km^2 dataset created from different datasets related to population pressure, such as land use, land cover and general infrastructure data, were obtained from NASA's Socioeconomic Data and Applications Center (SEDAC) (https://sedac.ciesin.columbia.edu/).

In addition, we included three categorical variables, the time, divided into three periods: 1980–1990, 1991–2010 and 2011–2022; age group of the study subjects (<6 years old =1, 6–15 years =2 and >15 years=3) and information whether or not MDA had been carried out. MDA was assigned values of 0 or 1, where the former represented its absence and 1 that it had been executed.

Table 1
Variables used in the study of schistosomiasis distribution in Zimbabwe.

Type of factor	Variable	Period	Spatial resolution	Source	Link
Climate	Bio1-19	1970–2000	1 km	Worldclim	http://www.worldclim.org/
Environment	Elevation	2020	5 km	Worldclim	
	Distance to water body	2009	5 km	GRG[a] vector data	http://gis.ess.washington.edu/data/vector/
	NDVI[b]	2000–2020	5 km	NASA[c]	https://srtm.csi.cgiar.org/
	Slope	2020	1 km	NASA[c]	https://srtm.csi.cgiar.org/
Socio-economy	Population density	2000–2022	1 km	WorldPop	http://www.worldpop.org.uk/
	[d]GDP	1990, 2025	25 km	SEDAC[e]	https://sedac.ciesin.columbia.edu/
	[f]HII	1995–2004	1 km	SEDAC[e]	https://sedac.ciesin.columbia.edu/

[a] Geomorphological Research Group.
[b] Normalized difference vegetation index.
[c] U.S. national aeronautics and space administration.
[d] Gross domestic product (million USD).
[e] Socioeconomic data and applications centre.
[f] Human influence index.

2.3. ML model construction

2.3.1. Variable selection and correlation

This study employed ML to construct risk models built on transmission data. First, we examined the relationships among the continuous variables. Spearman's correlation was used to calculate the strength of this relationship, which ranged from −1 to 1. Values > 0 indicated positive correlations, whereas values < 0 indicated negative correlations. Based on the results, variables with correlation coefficients >0.9 were considered to have strong multicollinearity and were therefore removed, except that the variable with the highest correlation within this group, i.e. the coefficient with the dependent variable, was selected and included in the model. Combining the remaining, continuous variables with the three categorical variables: time period, age and MDA (whether it had been done or not) provided information for the model fitting. The 'corrplot' package in the R platform was used to visualize the correlation matrix.

2.3.2. Model evaluation

ML provides various fitting models based on the research objectives and data type. This study used the 'caret' package of the R software to find the best model and performance evaluation. All data were divided into training and testing sets at a 8:2 ratio. The training set data was used to train different models, including random forest (RF), gradient boosting machine (GBM) and eXtreme Gradient Boosting (XGB). Corresponding models were constructed using these algorithms. The testing dataset was used to evaluate the optimal performance of each model. For different models, hyperparameters in the training set were determined using grid search. The performance of each hyperparameter combination was assessed through 10-fold cross-validation to identify the combination that yielded the best performance. The details and hyperparameters of each ML model used, both for *S. haematobium* and *S.mansoni*, are shown in Appendix 1. Three evaluation metrics (mean absolute error (MAE), root mean squared error (RMSE), and the coefficient of determination (R^2) as used by Huang's team (Huang et al., 2020) were applied to assess the model's optimal performance on using the testing set. The first two reflect the predictive performance error of the model. Lower values of these two parameters indicate a better model performance. R^2 represents the fitness of actual and predicted data with a range of 0–1; the closer to 1, the better the prediction.

2.4. Model prediction and risk analysis

The optimal ML model was selected to calculate the transmission risk for both schistosome species throughout Zimbabwe for the period 2011-2022. The variables for prediction were those from the model fitting: on the one hand the selected continuous variables related to climate, demography and socio-economy, and the other one the three categorical ones, i.e. period of investigation, age group and whether or not MDA had been done. To improve the stability of the models and to compute 95% confidence intervals (CI) for estimates, 100 fitting repetitions were performed using resampled the training set and testing set, with the average of which considered the final mean.

Prediction of the risk for *S. mansoni* or *S. haematobium* in Zimbabwe for the latest period was hypothesized for the two different MDA scenarios and for three different age groups. The overall infection rate for the entire population was adjusted based on the population distribution of Zimbabwe in 2022 when the proportions of Zimbabwe's population aged <5 years, 5–14 years and >14 years were 14%, 26% and 60%, respectively (PopulationPyramid.net ., 2023). The mixed infection rate of *S. haematobium* and *S. mansoni* was calculated using the formula presented as follows (Kokaliaris et al., 2022; Lai et al., 2015):

$$\text{prevalence}_{\text{mixed}} = \text{prevalence}_{\text{s.mansoni}} + \text{prevalence}_{\text{s.haematobium}} - \text{prevalence}_{\text{s.mansoni}} \times \text{prevalence}_{\text{s.haematobium}}$$

Where "mixed" refers to geographical areas where the two species of schistosomes are co-endemic. This means that the populations in these areas are exposed to both *S. mansoni* and *S. haematobium*, which could potentially lead to individuals being infected with both species simultaneously.

The regional- or provincial-level average infection rates for the period of 2011–2022 and their 95% CI were extracted from the predicted maps. Based on Zimbabwe's population at national and provincial level in 2022, the required amounts of praziquantel was estimated at the national and regional levels based on WHO's global guidelines (2022) for schistosomiasis elimination and control (World Health Organization., 2020). The burden of schistosomiasis was assessed by estimating the number of years lived with non-fatal disability (YLD), which was calculated as follows:

YLD = number of population x incidence x average duration x disability weight = number of population x prevalence x disability weight

The overall disability weight (Qian et al., 2011) were calculated by the probability and related sequelae caused by *S. haematobium* and *S. mansoni* infections. The probability (Van der Werf et al., 2003) and disability weight of each sequela for schistosomiasis were obtained from the Global Burden of Diseases (GBD) database (Global Burden of Disease Collaborative Network., 2020).

3. Results

3.1. Reference review and data extraction

A total of 1,893 references were collected for this study. Among these, 148 were retrieved from PubMed, 65 from Web of Science, 1,465 from ScienceDirect and 208 from African Journals Online. In addition, articles of particular relevance were included, one of which was the second national survey report on schistosomiasis in Zimbabwe. After removing 61 duplicate articles, 1,832 unique articles were obtained, from which 1,622 were excluded for not being relevant to the study. After thoroughly reading the full texts of the remaining 210 articles, a total of 52 met the inclusion criteria. They included data from 155 survey points for *S. haematobium* infection distributed across ten provinces and 56 districts in Zimbabwe, and data from 110 survey points for *S. mansoni* infection covering ten provinces and 55 districts. The spatial distribution of the survey points is shown in Fig. 1. The characteristics of the data extracted from the references regarding *S. haematobium* and *S. mansoni* infections in Zimbabwe are presented in Appendix 2, organized by the selected period of year, age groups, diagnostic methods, and MDA. The total number of individuals tested for *S. haematobium* infections was 40,896, with an average infection rate of 28.8% (standard deviation (SD) = 20.84%), and 26,304 for *S. mansoni* infection, with an average infection rate of 10.5% (SD = 10.01%). Urine filtration was the most commonly employed diagnostic method for *S. haematobium* infections, representing 92.3% (142/155) of all surveys. The Kato-Katz method was the predominant diagnostic method for *S. mansoni* infections, utilized in 96.4% (106/110) of the surveys.

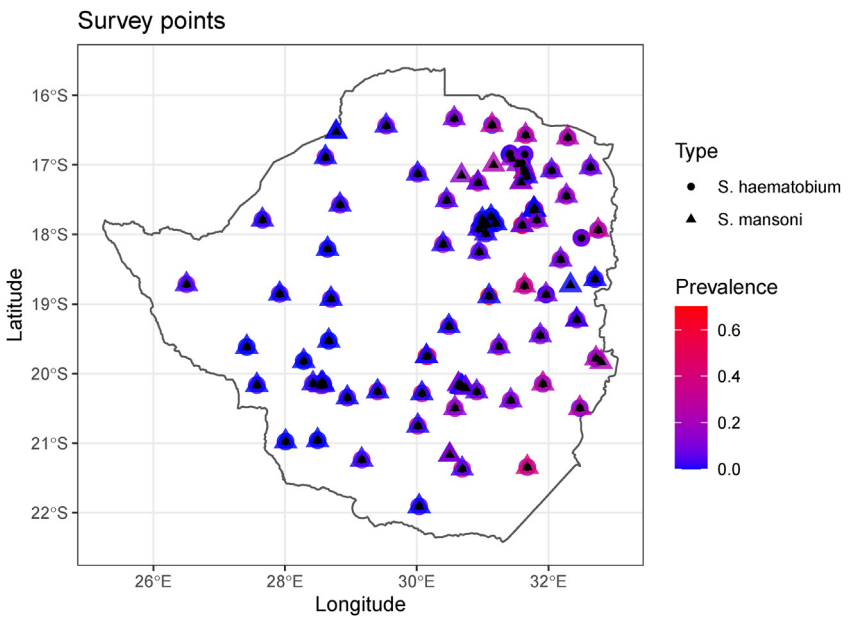

Fig. 1. Observed prevalence of *S. haematobium* and *S. mansoni* **in Zimbabwe**.

Table 2
Selected variables in the final model.

Type of factor	Variable	Schistosoma haematobium[a]	Schistosoma mansoni[a]
Time period	1980–1990	14.19% (8.70%–19.69%)	13.64 %(7.22%–20.05%)
	1991–2010	21.94 %(15.42%–28.45%)	14.55% (7.96%–21.13%)
	2011–2022	63.87% (56.31%–71.43%)	71.82% (63.41%–80.23%)
Age group	<6 years	15.48% (9.79%–21.18%)	3.64% (0.14%–7.13%)
	6–15 years	60.65% (52.95%–68.34%)	74.55% (66.41%–82.69%)
	>15 years	23.87% (17.16%–30.58%)	21.82% (14.1%–29.54%)
MDA situation	No (0)	87.10 %(81.82%–92.37%)	94.55% (90.3%–98.79%)
	Yes (1)	12.90% (7.63%–18.18%)	5.45 %(1.21%–9.70%)
Bio2	Mean diurnal range	13.38 (12.88–13.82)	/
Bio3	Isothermality	58.04 (57.40–58.95)	58.15 (57.39–58.87)
Bio4	Temperature seasonality	311.30 (286.80–327.30)	312.20 (288.70–328.80)
Bio5	Maximum temperature of the warmest month	29.70 (28.50–30.70)	/
Bio6	Minimum temperature of the coldest month	6.50 (6.10–7.10)	6.50 (6.10–7.08)
Bio7	Annual temperature range	23.10 (22.05–23.70)	22.95 (22.23–23.88)
Bio8	Mean temperature of the wettest quarter	/	22.27 (21.05–23.10)
Bio12	Annual precipitation	779.00 (657.00–851.00)	/
Bio13	Precipitation of the wettest month	/	178.00 (138.00–208.00)
Bio15	Precipitation seasonality	110.79 (102.20–117.46)	106.93 (±9.48)
Bio17	Precipitation of the driest quarter	5.00 (3.00–10.50)	6.50 (3.00–11.00)
Bio18	Precipitation of the warmest quarter	308.00 (268.00–372.00)	312.00 (261.20–358.50)
Slope	Grade of land tilt	1.82 (1.18–2.93)	1.71(1.04–2.99)
Elevation	Altitude	1124.00 (947.00–1320.00)	1136.00 (929.20–1347.00)
NDVI	Normalized difference vegetation index	0.11 (±0.38)	0.09 (±0.38)
HII	Human influence index	22.00 (16.00–26.00)	22.00 (14.00–29.00)
Water distance	Distance to closest water body	17153.00 (11201.00–25259.00)	16597.00 (11143.00–24406.00)
Population density	Number of persons/km^2	41.03 (24.79–143.60)	34.05 (18.54–167.69)
GDP	Gross domestic product	3.50 (2.00–5.00)	3.25 (1.00–5.00)

[a] The characteristics of categorical variables are represented by frequency and 95% CI. The characteristics of continuous variables conforming to normal distribution are described by mean (± standard deviation), while those not conforming to normal distribution are described by median (upper quartile - lower quartile).

3.2. Variable correlation and contribution

A total of 29 variables were used to construct the transmission model. Finally, 20 variables were selected in the *S. haematobium* ML model and 19 variables for the *S. mansoni* model (Table 2).

Fig. 2 shows the correlation matrix for the relationship between the variables in *S. haematobium* and *S. mansoni*. Nine factors, *i.e.* year period, age group, MDA distribution, economy, population density, HII, temperature, location, and precipitation, were correlated with infection rates of *S. haematobium* and *S. mansoni,* respectively.

3.2.1. Variables associated with S. haematobium infection rates

Five variables showed particularly high associations with either high or low infection rates of *S. haematobium*, *i.e.* year period, living in an area with MDA implementation (MDA=1), precipitation of the warmest quarter(Bio18), economical situation and annual precipitation.

The following variables were found to be associated with lower *S. haematobium* infection rates, *i.e.* the most recent period of years (2011-2022), having experienced MDA (MDA=1), being socio-economically well off, and normally experiencing high temperature (high Bio5, Bio2 and Bio7). Conversely, with respect to higher *S. haematobium* infection rates, we found high rainfall (Bio18, Bio12 and Bio17), and relatively steep slopes correlate positively with higher number of *S. haematobium* infections.

3.2.2. Variables associated with S. mansoni infection rates

The following variables were found to correlate positively with higher *S. mansoni* infection rates, *i.e.* precipitation of the warmest quarter (Bio18), slope, minimum temperature of the coldest month (Bio6), precipitation of the driest quarter (Bio17), isothermality (Bio3), mean temperature of the wettest quarter (Bio8), precipitation of the wettest month (Bio13), age, and NDVI. By contrast, factors were found to correlate negatively with higher rates of *S. mansoni* infections, *i.e.* elevation, year period (2011-2022), HII, population density, and economic conditions. Among these variables, five variables including elevation, Bio18, slope, year period and Bio6 had the strongest correlations.

3.3. ML and model construction

The results of Table 3 indicated that the XGB models had lower MAE and RMSE and higher R^2 values than the GBM and RF models for *S. haematobium* in the training dataset, but the RF model showed the best performance with regard to the testing dataset. With regard to the optimal risk model for *S. mansoni*, it should be said that the RF and XGB models showed high

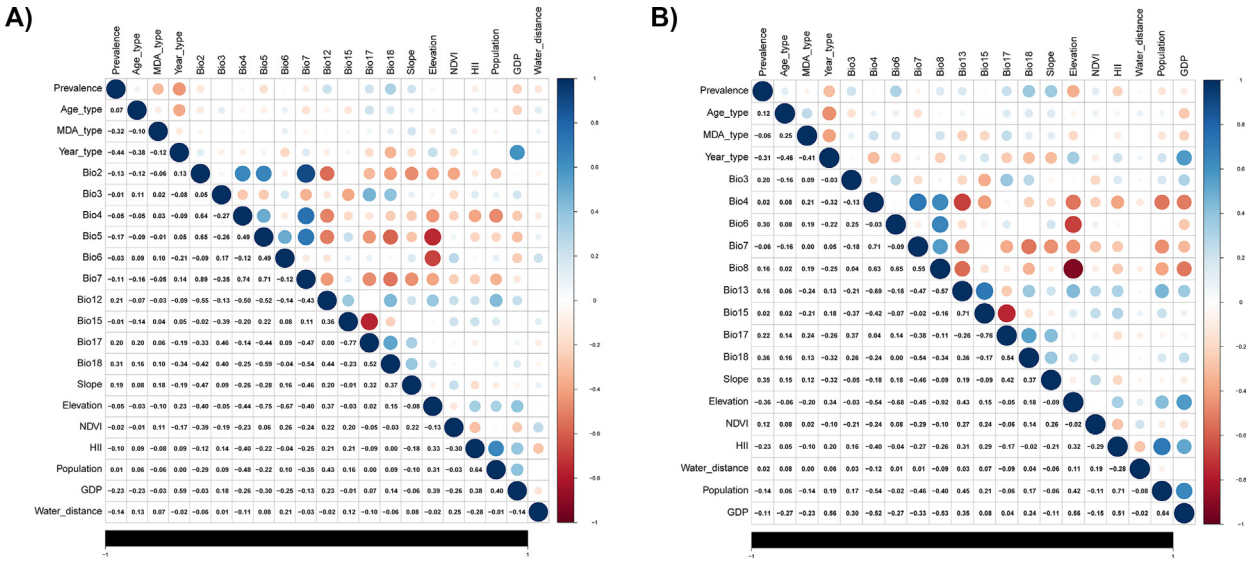

Fig. 2. Correlation matrix of the relationship between variables in *S. haematobium* and *S. mansoni*
A=*S. haematobium*, B=*S. mansoni;* Spearman's correlation among various variables in relation to the prevalence of *S. haematobium* and *S. mansoni*. The variables considered include climatic factors (Bio2, Bio3, Bio4, Bio5, Bio6, Bio7, Bio8, Bio12, Bio13, Bio15, Bio17, Bio18), environmental factors (Slope, Elevation, Water_distance, NDVI), and socio-economic factors (Age_type, MDA_type, Year_type, HII, Population, GDP). Positive correlations represented in blue and negative correlations in orange.

Table 3
Evaluation of three machine-learning regression models.

Data	Model	*S. haematobium*				*S. mansoni*			
		No. of survey points	MAE	RMSE	R^2	No. of survey points	MAE	RMSE	R^2
Training set	RF	124	0.047	0.063	0.937	89	0.034	0.049	0.822
	GBM		0.052	0.067	0.900		0.042	0.063	0.629
	XGB		0.027	0.040	0.964		0.038	0.053	0.749
Testing set	RF	31	0.108	0.143	0.517	21	0.053	0.082	0.458
	GBM		0.106	0.148	0.487		0.063	0.089	0.269
	XGB		0.104	0.145	0.505		0.058	0.088	0.282

RF = random forest; GBM = gradient boosting machine; XGB = eXtreme Gradient Boosting; MAE = mean absolute error; RMSE = root mean squared error; R^2=determination coefficient.

consistency in the actual and predicted values in the training set, but the RF model again showed the best performance with regard to the testing dataset. Based on the models' performance, the RF model was the most optimal both for *S. haematobium* ($R^2 = 0.517$) and *S. mansoni* ($R^2 = 0.458$).

3.4. Model prediction

Using the optimal models, predictions were conducted to assess the transmission risk of schistosomiasis under different MDA scenarios. Infection rates were accounted for adjusted by population distribution, and risk maps for generated for *S. haematobium* and *S. mansoni* separately as well as for mixed infections for the entire population for the period 2011–2022. Fig. 3 shows the schistosomiasis risk when MDA was not implemented. Predictive results for *S. haematobium* infections indicated a higher prevalence in the central, southern, and south-eastern regions of Zimbabwe, whereas the western area had a lower prevalence. In contrast, the *S. mansoni* map predicted severe prevalence presented in the eastern, south-eastern, and north-eastern areas of the country, with low infection rates in the western region. Without MDA, the population-adjusted prevalence in 2022 was predicted to be 17.12% (95%CI: 17.02%–17.23%) for *S. haematobium*, 7.43% (95%CI: 7.36%–7.50%) for *S. mansoni*, and 23.20% (95%CI: 23.08%–23.32%) for mixed infections.

Additionally, a risk map for schistosomiasis in Zimbabwe for 2022 was predicted under the MDA implementation scenario (Fig. 4). The prediction map clearly demonstrated that the infection rates for *S. haematobium* were significantly lower under the MDA scenario than without MDA. However, this was not observed for *S. mansoni*. Under the MDA implementation scenario, the population-adjusted prevalence in 2022 was predicted to be 13.75% (95% CI: 13.65%–13.85%) for *S. haematobium*, 7.07% (95%CI:7.01%–7.13%) for *S. mansoni*, and 19.79% (95%CI: 19.67%–19.91%) for mixed infections.

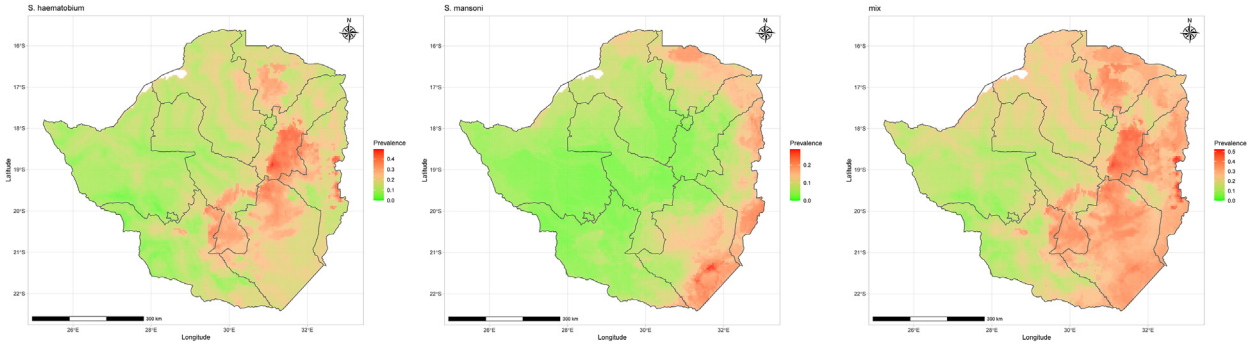

Fig. 3. Predictive risk map of infections with *Schistosoma haematobium* and *Schistosoma mansoni* and mixed infections in 2022 without MDA implemented A=*S. haematobium*; B=*S. mansoni*; C=Mixed infections.

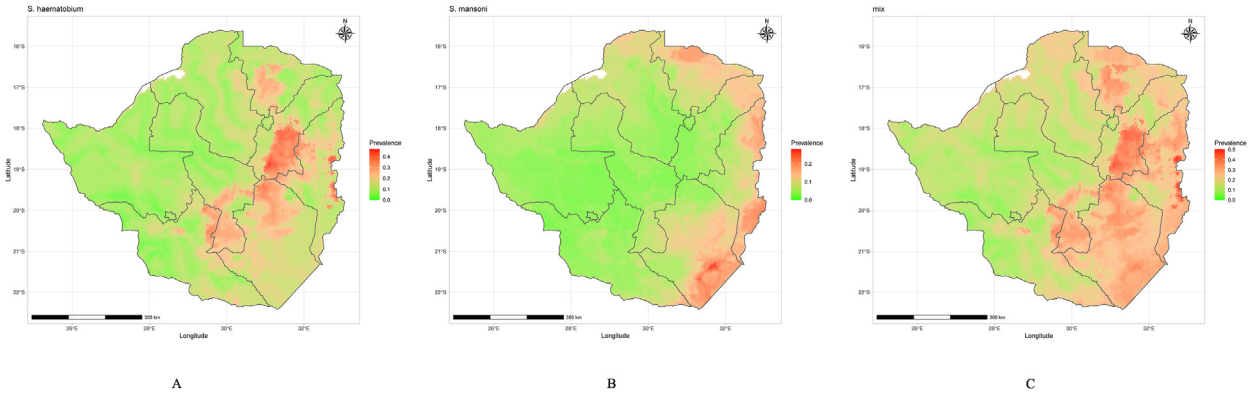

Fig. 4. Predicting risk map of infections with *Schistosoma haematobium* and *Schistosoma mansoni* and mixed infections in 2022 with MDA implemented A=*S. haematobium*;B=*S. mansoni*; C=Mixed infections.

The required quantity of praziquantel for MDA for the entire country and regional levels was estimated according to WHO guidelines. In total, it required an annual MDA population of 3,521,466 (95%CI:3502580–3540352) without MDA scenario and 3,003,928 (95%CI: 2986151–3021705) under the MDA scenario (Table 4).

The overall estimated disability weights of total schistosomiasis as well as of *S. haematobium,* and *S. mansoni,* were0.023, 0.025, and 0.017, respectively. The estimated disease burden using YLD (Table 5) in the national level was 80,993 DALYs (95% CI: 80559–81428) for the non-MDA implementation scenario and 69,090 DALYs (95%CI: 68681–69499) for the MDA implementation scenario.

4. Discussion

This is the first study to predict the prevalence of schistosomiasis in Zimbabwe using ML algorithms. By comparing three different algorithms, the RF one emerged as the best performing method for constructing the modelling, considering the corresponding environmental and socioeconomic factors. This model confirmed that the average prevalence of schistosomiasis in Zimbabwe is moderate. Still, it is a significant public health problem, and the country is moving towards controlling and eliminating targets. However, Zimbabwe has not yet conducted a nationwide survey based on the whole population, as a results of unclear schistosomiasis prevalence and distribution yet in the country, although two national surveys based on the school-aged children were conducted in last 4 decades (Midzi et al., 2014; Taylor & Makura, 1985). The findings of the study that estimated the prevalence of schistosomiasis in Zimbabwe using machine-learning algorithms revealed that schistosomiasis remains a highly prevalent risk in Zimbabwe, particularly in the central, southern, south-eastern, north-eastern, and eastern regions. Continued attention is essential for effective management and control of the disease in these areas of Zimbabwe.

The correlation between the prevalence of schistosomiasis and the environmental and socioeconomic factors revealed that factors such as extreme climatic conditions (rainfall and temperature), NDVI, distance to water bodies, elevation, slope, and economic and population factors had various effects on the transmission risk of schistosomiasis. This is consistent with

Table 4
Predicted, required quantity of praziquantel need for MDA in 2022 by region in Zimbabwe.

Region	Population 2022	*S. haematobium*		*S. mansoni*		Mixed infection		MDA implanted Yes/No
		Prevalence (%) (95% CI)	MDA population (95% CI)	Prevalence (%) (95% CI)	MDA population (95% CI)	Prevalence (%) (95% CI)	MDA population (95% CI)	
Midlands	1811908	16.87 (16.57–17.16)	305580 (300292–310868)	5.02 (4.93–5.12)	90996 (89248–92745)	20.96 (20.60–21.32)	379732 (373200–386265)	No
Matabeleland South	760345	14.03 (13.78–14.28)	106687 (104799–108576)	4.82 (4.72–4.93)	36683 (35861–37506)	18.18 (17.87–18.50)	138252 (135873–140631)	No
Matabeleland North	827626	10.58 (10.49–10.67)	87545 (86816–88274)	4.17 (4.11–4.24)	34536 (33989–35083)	14.30 (14.17–14.43)	118350 (117239–119462)	No
Masvingo	1638539	22.22 (22.01–22.44)	364131 (360624–367637)	12.08 (11.84–12.31)	197877 (194034–201719)	31.76 (31.54–31.98)	520368 (516743–523994)	No
Mashonaland West	1893578	16.04 (15.92–16.15)	303675 (301532–305817)	6.70 (6.59–6.82)	126912 (124696–129127)	21.67 (21.52–21.83)	410378 (407422–413334)	No
Mashonaland East	1731181	24.42 (23.92–24.93)	422811 (414084–431538)	7.20 (6.95–7.45)	124624 (120305–128943)	30.04 (29.52–30.55)	519999 (511095–528903)	No
Mashonaland Central	1384891	20.05 (19.74–20.35)	277618 (273423–281812)	11.96 (11.72–12.20)	165654 (162306–169002)	29.59 (29.25–29.93)	409781 (405087–414475)	No
Manicaland	2037762	20.98 (20.66–21.30)	427595 (421094–434095)	12.20 (11.92–12.48)	248569 (242909–254229)	30.82 (30.50–31.14)	628080 (621598–634562)	No
Harare	2427209	12.46 (11.83–13.08)	302321 (287152–317490)	3.77 (3.41–4.13)	91522 (82687–100357)	15.96 (15.22–16.70)	387393 (369416–405370)	No
Bulawayo	665940	10.14 (9.51–10.77)	67532 (63313–71750)	3.29 (3.16–3.41)	21883 (21044–22722)	12.97 (12.26–13.68)	86399 (81665–91133)	No
Zimbabwe	15178979	17.12 (17.02–17.23)	2599277 (2583213–2615340)	7.43 (7.36–7.50)	1127509 (1117315–1137703)	23.20 (23.08–23.32)	3521466 (3502580–3540352)	No
Midlands	1811908	12.91 (12.62–13.20)	233946 (228660–239231)	4.79 (4.70–4.88)	86788 (85228–88348)	17.04 (16.74–17.35)	308822 (303358–314286)	Yes
Matabeleland South	760345	12.02 (11.79–12.25)	91380 (89640–93120)	4.63 (4.53–4.73)	35173 (34408–35939)	16.04 (15.80–16.29)	121989 (120136–123842)	Yes
Matabeleland North	827626	8.64 (8.56–8.71)	71480 (70856–72105)	4.06 (4.00–4.12)	33603 (33087–34120)	12.34 (12.25–12.43)	102111 (101350–102872)	Yes
Masvingo	1638539	18.60 (18.39–18.81)	304750 (301374–308127)	11.31 (11.08–11.53)	185263 (181542–188985)	27.94 (27.75–28.12)	457747 (454658–460836)	Yes
Mashonaland West	1893578	11.41 (11.29–11.52)	215986 (213834–218138)	6.46 (6.35–6.57)	122327 (120184–124471)	17.12 (16.98–17.26)	324158 (321509–326807)	Yes
Mashonaland East	1731181	20.30 (19.76–20.85)	351510 (342085–360935)	6.89 (6.66–7.13)	119364 (115381–123347)	25.97 (25.50–26.44)	449625 (441524–457725)	Yes
Mashonaland Central	1384891	15.17 (14.83–15.51)	210081 (205420–214741)	11.36 (11.13–11.58)	157296 (154172–160419)	24.88 (24.56–25.20)	344526 (340089–348962)	Yes
Manicaland	2037762	17.38 (17.06–17.70)	354190 (347638–360743)	11.50 (11.23–11.77)	234382 (228914–239850)	26.99 (26.70–27.29)	550069 (544096–556041)	Yes
Harare	2427209	10.99 (10.33–11.65)	266717 (250776–282658)	3.73 (3.39–4.07)	90448 (82170–98726)	14.31 (13.49–15.12)	347236 (327538–366935)	Yes
Bulawayo	665940	8.69 (8.07–9.30)	57844 (53746–61943)	3.28 (3.16–3.41)	21849 (21012–22685)	11.66 (11.02–12.31)	77659 (73371–81947)	Yes
Zimbabwe	15178979	13.75 (13.65–13.85)	2086544 (2071209–2101879)	7.07 (7.01–7.13)	1073044 (1063462–1082627)	19.79 (19.67–19.91)	3003928 (2986151–3021705)	Yes

Table 5
Prediction of disease burden for schistosomiasis in Zimbabwe in 2022.

Schistosome species	Disability weight	YLD (95% CI)	
		Non-MDA scenario	MDA scenario
S. haematobium	0.025	64,981 (64,580–65,383)	52,163 (51,780–52,546)
S. mansoni	0.017	19,167 (18,994–19,340)	18,241 (18,078–18,404)
Mixed infection	0.023	80,993 (80,559–81,428)	69,090 (68,681–69,499)

previous studies (John et al., 2008; Pedersen et al., 2014; Stensgaard et al., 2016). However, these factors correlated differently with the two schistosomiasis species. For *S. haematobium*, three of the five was most closely associated socioeconomic factors, whereas two were related to environmental factors. Over time, five impact factors, including a later time period of investigation, implementation of MDA, low rainfall during the warmest season, higher GDP level, and low annual precipitation, were the top five factors closely related to the decline in the prevalence of *S. haematobium*. In contrast, among the five most relevant factors for *S. mansoni* prevalence, four were environmental and one was economic factors. The lower prevalence of *S. mansoni* was closely related to five impact factors, such as higher elevation, low rainfall during the warmest season, lower slope, a later

time period of investigation, as well as lower minimum temperature of the coldest month. These results are in line with a prior study that the limited impact of climate change on the prevalence of schistosomiasis was presented in Zimbabwe (Pedersen et al., 2017). This study also found that although climatic factors are important in the transmission of schistosomiasis, the infection of *S. haematobium* is also associated with socio-economic factors, whereas *S. mansoni* infection is closely related to topographical and environmental factors in addition to climatic factors. These findings provide valuable information for formulating strategies to control and eliminate schistosomiasis eventually, especially in the low endemic areas. Thus, a tailored strategy considering the various characteristics of different schistosome species and the related impact factors are of importance and very crucial.

Although the distance to the nearest water body correlated with the prevalence of schistosomiasis, it was not among the top-ranking associations. This observation could be related to Zimbabwe's climate pattern, which is characterised by distinct rainy and dry seasons. Seasonality could influence the abundance of the intermediate snail host, but not stop their presence (Gouvras et al., 2017). Water bodies have extensive coverage during the rainy season, resulting in easy access to water, whereas they tend to dry up during the dry season, leading people to search for water sources, thus concentrating at just a few water bodies or sites. These water sources may be used for swimming, bathing, gardening, fishing, washing, etc (Mutsaka-Makuvaza et al., 2019; Mutsaka-Makuvaza et al., 2019). Although such unprotected water sources can lead to schistosomiasis infection (Nyati-Jokomo & Chimbari, 2017), their distance from residential areas might not the most significant factor (Mwai et al., 2021).

ML algorithms have been employed to predict the risk of transmission of schistosomiasis. Although those algorithms are often likened to a black box, as they can be challenging to decipher through traditional mathematical calculations, they have been utilized in previous studies that focused on predicting discrete or continuous results. The predicted prevalence in this study was consistent with the quantified prevalence at any point obtained from the prediction map. Although the predicted results of this study were higher than those of a previous study conducted elsewhere (Kokaliaris et al., 2022), they were similar to those of recent field surveys conducted in Zimbabwe (Mutapi et al., 2021; Mutsaka-Makuvaza et al., 2018; Mutsaka-Makuvaza et al., 2019).

This study used a scenario analysis to predict the prevalence of schistosomiasis with and without MDA implementation. Zimbabwe has been implementing six-year MDA programs targeting school-aged children since 2012. The regions where the program was implemented have witnessed a significant decrease in the prevalence of schistosomiasis among school-aged children, and MDA continues to be a major schistosomiasis control measure in Zimbabwe. However, it is difficult to obtain exact information about where MDA was implemented and where it was not. Scenario analysis was conducted including the implementation of MDA as a factor in the model and exploring schistosomiasis transmission risk under different MDA scenarios. The results indicated a lower transmission risk where MDA was used, especially for *S. haematobium,* which may relate to *S. haematobium* being more sensitive to praziquantel than *S. mansoni* (Kabuyaya et al., 2018). These results provide a reference for predicting the risk of schistosomiasis based on the actual implementation of the MDA in different regions.

Given the lack of prevalence survey data, estimates of the number of individuals requiring treatment on different administration level can serve as a reference for local implementation of MDA, which will improve the feasibility of MDA administration and optimize resource allocation. Effective MDA should prevent both the potential overuse of praziquantel, which could lead to drug resistance, and under-treatment, which may increase the disease burden. Local authorities are also recommended to continuously monitor and evaluate the effectiveness of MDA and adjust strategies as needed. Furthermore, the study reveals that MDA alone is insufficient for schistosomiasis elimination, especially with respect to *S. mansoni*. Local control efforts should therefore integrate other measures such as WHO-recommended health education to raise awareness about disease and preventive practices, and safe water and sanitation and hygiene (WASH) strategies. In addition, snail control should be revived tackle the presence of the intermediate host, though environmental modifications to reduce snail habitat suitability, and warnings at common points of water contact to reduce exposure. These comprehensive control efforts are expected to enhance overall prevention and control outcomes.

5. Limitations

Despite collecting over 40 years of publicly available schistosomiasis epidemiological data in Zimbabwe, the actual number of data points was limited, which might have led to biases in the model's predictive results. Secondly, most of the data obtained were from relatively high prevalence areas, with less information obtained from low-prevalence or non-epidemic areas, which might have led to the overestimated of the predictions. Thirdly, the numbers of environmental and socioeconomic variables used in this study could not reflect all influential factors. The ones used were derived from public databases, and potential measurement and accuracy errors in these data could have contributed to model prediction biases.

6. Conclusion

The RF algorithm emerged as the most formal method for constructing the model. We also found that the impact of various factors on the prevalence of *S. haematobium* and *S. mansoni* differed. Socioeconomic and environmental factors were closely associated with *S. haematobium*, whereas topographical and environmental factors were more closely associated with *S. mansoni* prevalence. The study reveals that MDA alone is insufficient for schistosomiasis elimination, especially *S. mansoni*. These findings provide insights into tailoring different schistosomiasis control strategies during periods of low prevalence.

Additionally, this study revealed that schistosomiasis control measures in Zimbabwe have primarily focused on MDA in school-aged children, with few other control strategies implemented. Based on the latest WHO schistosomiasis control guidelines and China's schistosomiasis control experience, achieving schistosomiasis elimination goals requires a more comprehensive approach. It is recommended that Zimbabwe adopt an integrated control strategy that includes not only population-wide MDA, but also snail control by environmental modification, better access to clean water sources and sanitation facilities, health education. Systematic monitoringwould be important to follow up.

Funding

This study was supported by the program of the Chinese Center for Tropical Diseases Research (No. 131031104000160004) and the China-Africa Cooperation Project on Schistosomiasis Control and Elimination (2020-C4-0001-2).

Availability of data

For access to the data obtained from public databases, please contact the corresponding author.

Ethics approval and consent to participate

All the data utilized in this study were sourced from the publicly available databases. No personal or biological sample were utilized in this research.

CRediT authorship contribution statement

Hong-Mei Li: Writing – original draft, Software, Methodology, Data curation. **Jin-Xin Zheng:** Software, Methodology, Data curation. **Nicholas Midzi:** Writing – review & editing, Resources. **Masceline Jenipher Mutsaka- Makuvaza:** Writing – review & editing, Resources. **Shan Lv:** Writing – review & editing, Supervision, Conceptualization. **Shang Xia:** Writing – review & editing, Methodology, Formal analysis. **Ying-jun Qian:** Investigation, Data curation. **Ning Xiao:** Writing – review & editing, Supervision, Conceptualization. **Robert Berguist:** Writing – review & editing. **Xiao-Nong Zhou:** Writing – review & editing, Writing – original draft, Supervision, Investigation, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:Xiao- Nong Zhou reports financial support was provided by National Health Commission of the People's Republic of China.

Appendix 1. The hyperparameters of each machine learning model for *S. haematobium* and *S.mansoni*

For *S. haematobium*:
The hyperparameters used in this Random Forest (RF) model for *S. haematobium* are as follows.

1. mtry (Number of variables tried at each split): 22
2. ntree (Number of trees): 500 (default)
3. nodesize (Minimum size of terminal nodes): 5 (default)
4. maxnodes (Maximum number of terminal nodes): NULL (no limit, default)

The final hyperparameters for the XGBoost model for *S. haematobium* are are as follows.

1. eta (learning rate): 0.3
2. max_depth (maximum depth of a tree): 3
3. gamma (minimum loss reduction required to make a further partition on a leaf node): 0
4. colsample_bytree (subsample ratio of columns when constructing each tree): 0.6
5. min_child_weight (minimum sum of instance weight needed in a child): 1
6. subsample (subsample ratio of the training instances): 1
7. nrounds (number of boosting rounds): 100

The hyperparameters for the GBM model for *S. haematobium* are are as follows.

1. shrinkage: 0.1
2. n.minobsinnode (minimum number of observations in each node): 10

3. n.trees (number of trees in the model): 100
4. interaction.depth (depth of variable interactions): 2

For *S.mansoni*:
The hyperparameters used in the Random Forest (RF) model for *S.mansoni* are as follows.

1. mtry (Number of variables tried at each split): 2
2. ntree (Number of trees): 500 (default)
3. nodesize (Minimum size of terminal nodes): 1 (default)
4. maxnodes (Maximum number of terminal nodes): NULL (no limit, default)

The hyperparameters used in this XGBoost model for *S.mansoni* are as follows.

1. eta (Learning rate): 0.3
2. max_depth (Maximum depth of a tree): 1
3. gamma (Minimum loss reduction required to make a further partition on a leaf node): 0
4. colsample_bytree (Subsample ratio of columns when constructing each tree): 0.6
5. min_child_weight (Minimum sum of instance weight needed in a child): 1
6. subsample (Subsample ratio of the training instances): 1
7. nrounds (Number of rounds for boosting): 50

The hyperparameters for the GBM model for *S.mansoni* are as follows.

1. shrinkage: 0.1
2. n.minobsinnode (minimum number of observations in each node): 10
3. n.trees (number of trees in the model): 100
4. nteraction.depth (depth of variable interactions): 2

Appendix 2. Characteristics for *S. haematobium* (A) and *S. mansoni* (B) in Zimbabwe during 1980–2022 extracted from references

A

Characteristic	No. of survey points	No. of participants	No. of positives	Infection rate	
				Mean (±standard deviation)	Median (upper quartile-lower quartile)
Year period					
1980~1990	22	5257	2149	40.88% (±25.07%)	33.69% (15.90%–50.58%)
1991~2010	34	12051	5081	42.17% (±21.37%)	39.84% (24.27%–59.69%)
2011~2022	99	23588	4542	19.26% (±14.86%)	12.07% (4.26%–26.30%)
Age group					
<6	24	5643	993	17.60% (±14.82%)	13.53% (4.63%–30.33%)
6~15	94	22335	7094	31.76% (±23.25%)	18.05% (5.90%–40.86%)
>15	37	12918	3686	28.53% (±16.63%)	20.29% (8.45%–35.32%)
Diagnostic method					
Urine filtration	141	36082	9615	26.65% (±20.28%)	16.30% (5.90%–34.90%)
Urinary dipstick	2	1005	163	16.22% (±2.62%)	16.85% (15.00%–18.70%)
Urine filtration and urinary dipstick	1	551	329	59.71% (±2.09%)	59.71% (59.71%–59.71%)
Sedimentation-centrifugation and miracidial hatching	11	3258	1666	51.14% (±20.98%)	40.12% (26.68%–62.70%)
MDA implementation					
No	135	37278	11490	30.82% (±21.15%)	20.90% (9.00%–39.80%)
Yes	20	3618	283	7.82% (±7.09%)	6.31% (3.35%–9.16%)
Total	155	40896	11773	28.79% (±20.84%)	18.60% (6.08%–36.62%)

B

Characteristic	No. of survey points	No. of participants	No. of positives	Infection rate	
				Mean (±standard deviation)	Median (upper quartile-lower quartile)
Year period					
1980~1990	15	2743	368	13.41% (±11.01%)	8.10% (2.60%–19.00%)
1991~2010	16	10061	1376	13.71% (±8.61%)	14.01% (0.70%–15.95%)
2011~2022	79	13500	1013	7.51% (±9.81%)	0.80% (0.00%–6.50%)
Age group					
<6	4	937	5	0.51% (±2.71%)	0.26% (0.00%–4.33%)
6~15	82	18427	1787	9.71% (±10.71%)	1.50% (0.00%–12.15%)
>15	24	6940	965	13.91% (±8.21%)	4.65% (0.18%–15.22%)
Diagnostic method					
Kato Katz	34	7279	716	9.81% (±9.21%)	3.25% (0.00%–14.26%)
Formol-ether concentration	2	2030	349	17.21% (±4.11%)	18.72% (15.79%-.)
Kato Katz and formol-ether concentration	72	15444	1420	9.21% (±10.31%)	0.85% (0.00%–8.93%)
Sedimentation-centrifugation and miracidial hatching	2	1551	272	17.51% (±6.71%)	18.20% (13.50%–18.20%)
MDA implementation					
No	104	25092	2732	10.91% (±10.21%)	2.29% (0.00%–13.00%)
Yes	6	1212	25	2.11% (±3.61%)	1.30% (0.00%–7.20%)
Total	110	26304	2757	10.51% (±10.01%)	2.05% (0.00%–12.33%)

References

Agniwo, P., Sidibe, B., Diakite, A., Niare, S. D., Guindo, H., Akplogan, A., et al. (2023). Ultrasound aspects and risk factors associated with urogenital schistosomiasis among primary school children in Mali. *Infect Dis Poverty, 12*(1), 40. https://doi.org/10.1186/s40249-023-01071-6

Bartlett, A. W., Sousa-Figueiredo, J. C., van Goor, R. C., Monaghan, P., Lancaster, W., Mugizi, R., et al. (2022). Burden and factors associated with schistosomiasis and soil-transmitted helminth infections among school-age children in Huambo, Uige and Zaire provinces, Angola. *Infect Dis Poverty, 11*(1), 73. https://doi.org/10.1186/s40249-022-00975-z

Bergquist, N. R., Zhen, J. X., & Zhou, X. N. (2024). Synergistic integration of climate change and zoonotic diseases by artificial intelligence (AI): A holistic approach for sustainable solutions. *Science in One Health, 3*, Article 100070. https://doi.org/10.1016/j.soh.2024.100070

Bustinduy, A., King, C., Scott, J., Appleton, S., Sousa-Figueiredo, J. C., Betson, M., et al. (2014). HIV and schistosomiasis co-infection in African children. *The Lancet Infectious Diseases, 14*(7), 640–649. https://doi.org/10.1016/S1473-3099(14)70001-5

Chimbari, M. J. (2012). Enhancing schistosomiasis control strategy for Zimbabwe: Building on past experiences. *J Parasitol Res, 2012*, Article 353768. https://doi.org/10.1155/2012/353768

Choto, E. T., Mduluza, T., Mutapi, F., & Chimbari, M. J. (2020). Association of schistosomiasis and risk of prostate cancer development in residents of Murehwa rural community, Zimbabwe. *Infectious Agents and Cancer, 15*, 59. https://doi.org/10.1186/s13027-020-00327-2

Diop, B., Sylla, K., Kane, N. M., Boh, O. K., Gueye, B., Ba, M., et al. (2023). Schistosomiasis control in Senegal: Results from community data analysis for optimizing preventive chemotherapy intervention with praziquantel. *Infect Dis Poverty, 12*(1), 106. https://doi.org/10.1186/s40249-023-01155-3

Ekpo, U. F., Oluwole, A. S., Abe, E. M., Etta, H. E., Olamiju, F., & Mafiana, C. F. (2012). Schistosomiasis in infants and pre-school-aged children in sub-saharan Africa: Implication for control. *Parasitology, 139*(7), 835–841. https://doi.org/10.1017/S0031182012000029

Global Burden of Disease Collaborative Network. (2020). *Global burden of disease study 2019 (GBD 2019) disability weights*. Seattle, United States of America: Institute for Health Metrics and Evaluation (IHME). https://ghdx.healthdata.org/gbd-2019.

Gong, Y. F., Zhu, L. Q., Li, Y. L., Zhang, L. J., Xue, J. B., Xia, S., et al. (2021). Identification of the high-risk area for schistosomiasis transmission in China based on information value and machine learning: A newly data-driven modeling attempt. *Infect Dis Poverty, 10*(1), 88. https://doi.org/10.1186/s40249-021-00874-9

Gouvras, A. N., Allan, F., Kinung'hi, S., Rabone, M., Emery, A., Angelo, T., et al. (2017). Longitudinal survey on the distribution of Biomphalaria sudanica and B. Choanomophala in mwanza region, on the shores of lake victoria, Tanzania: Implications for schistosomiasis transmission and control. *Parasites & Vectors, 10*(1), 316. https://doi.org/10.1186/s13071-017-2252-z

Hagan, P. (2009). Schistosomiasis–a rich vein of research. *Parasitology, 136*(12), 1611–1619. https://doi.org/10.1017/S003118200999093X

Hu, Y., Xia, C., Li, S., Ward, M. P., Luo, C., Gao, F., et al. (2017). Assessing environmental factors associated with regional schistosomiasis prevalence in Anhui Province, Peoples' Republic of China using a geographical detector method. *Infect Dis Poverty, 6*(1), 87. https://doi.org/10.1186/s40249-017-0299-x

Huang, J. C., Tsai, Y. C., Wu, P. Y., Lien, Y. H., Chien, C. Y., Kuo, C. F., et al. (2020). Predictive modeling of blood pressure during hemodialysis: A comparison of linear model, random forest, support vector regression, XGBoost, LASSO regression and ensemble method. *Computer Methods and Programs in Biomedicine, 195*, Article 105536. https://doi.org/10.1016/j.cmpb.2020.105536

Isaiah, P. M., Solveig Palmeirim, M., & Steinmann, P. (2023). Epidemiology of pediatric schistosomiasis in hard-to-reach areas and populations: A scoping review. *Infect Dis Poverty, 12*(1), 37. https://doi.org/10.1186/s40249-023-01088-x

John, R., Ezekiel, M., Philbert, C., & Andrew, A. (2008). Schistosomiasis transmission at high altitude crater lakes in western Uganda. *BMC Infectious Diseases, 8*, 110. https://doi.org/10.1186/1471-2334-8-110

Kabuyaya, M., Chimbari, M. J., & Mukaratirwa, S. (2018). Efficacy of praziquantel treatment regimens in pre-school and school aged children infected with schistosomiasis in sub-saharan Africa: A systematic review. *Infect Dis Poverty, 7*(1), 73. https://doi.org/10.1186/s40249-018-0448-x

Keshavamurthy, R., Dixon, S., Pazdernik, K. T., & Charles, L. E. (2022). Predicting infectious disease for biopreparedness and response: A systematic review of machine learning and deep learning approaches. *One Health, 15*, Article 100439. https://doi.org/10.1016/j.onehlt.2022.100439

Kokaliaris, C., Garba, A., Matuska, M., Bronzan, R. N., Colley, D. G., Dorkenoo, A. M., et al. (2022). Effect of preventive chemotherapy with praziquantel on schistosomiasis among school-aged children in sub-saharan Africa: A spatiotemporal modelling study. *The Lancet Infectious Diseases, 22*(1), 136–149. https://doi.org/10.1016/S1473-3099(21)00090-6

Kolodziej, P., Szostakowska, B., Lass, A., Sulima, M., Sikorska, K., Kocki, J., et al. (2023). Chronic intestinal schistosomiasis caused by co-infection with Schistosoma intercalatum and Schistosoma mansoni. *The Lancet Infectious Diseases*. https://doi.org/10.1016/S1473-3099(23)00486-3

Lai, Y. S., Biedermann, P., Ekpo, U. F., Garba, A., Mathieu, E., Midzi, N., et al. (2015). Spatial distribution of schistosomiasis and treatment needs in sub-saharan Africa: A systematic review and geostatistical analysis. *The Lancet Infectious Diseases, 15*(8), 927–940. https://doi.org/10.1016/S1473-3099(15)00066-3

Liu, X., Sun, Y., Yin, Y., Dai, X., Bergquist, R., Gao, F., et al. (2023). Influence of urbanization on schistosomiasis infection risk in anhui province based on sixteen year's longitudinal surveillance data: A spatio-temporal modelling study. *Infect Dis Poverty, 12*(1), 108. https://doi.org/10.1186/s40249-023-01163-3

Lu, Z. H., Yang, M., Pan, C. H., Zheng, P. Y., & Zhang, S. X. (2022). Multi-modal deep learning based on multi-dimensionaland multi-level temporal data can enhance the prognostic predictionfor multi-drug resistant pulmonary tuberculosis patients. *Science in One Health*. , Article 100004. https://doi.org/10.1016/j.soh.2022.100004

Mduluza, T., Jones, C., Osakunor, D. N. M., Lim, R., Kuebel, J. K., Phiri, I., et al. (2020). Six rounds of annual praziquantel treatment during a national helminth control program significantly reduced schistosome infection and morbidity levels in a cohort of schoolchildren in Zimbabwe. *PLoS Neglected Tropical Diseases, 14*(6), Article e0008388. https://doi.org/10.1371/journal.pntd.0008388

Midzi, N., Mduluza, T., Chimbari, M. J., Tshuma, C., Charimari, L., Mhlanga, G., et al. (2014). Distribution of schistosomiasis and soil transmitted helminthiasis in Zimbabwe: Towards a national plan of action for control and elimination. *PLoS Neglected Tropical Diseases, 8*(8), Article e3014. https://doi.org/10.1371/journal.pntd.0003014

Mooney, S. J., & Pejaver, V. (2018). Big data in public health: Terminology, machine learning, and privacy. *Annual Review of Public Health, 39*, 95–112. https://doi.org/10.1146/annurev-publhealth-040617-014208

Mutapi, F., Pfavayi, L., Osakunor, D., Lim, R., Kasambala, M., Mutemeri, A., et al. (2021). Assessing early child development and its association with stunting and schistosome infections in rural Zimbabwean children using the Griffiths Scales of Child Development [Article]. *PLoS Neglected Tropical Diseases, 15*(8). https://doi.org/10.1371/journal.pntd.0009660. Article e0009660.

Mutsaka-Makuvaza, M. J., Matsena-Zingoni, Z., Katsidzira, A., Tshuma, C., Chin'ombe, N., Zhou, X. N., et al. (2019). Urogenital schistosomiasis and risk factors of infection in mothers and preschool children in an endemic district in Zimbabwe. *Parasites & Vectors, 12*(1), 427. https://doi.org/10.1186/s13071-019-3667-5

Mutsaka-Makuvaza, M. J., Matsena-Zingoni, Z., Tshuma, C., Katsidzira, A., Webster, B., Zhou, X. N., et al. (2019). Knowledge, perceptions and practices regarding schistosomiasis among women living in a highly endemic rural district in Zimbabwe: Implications on infections among preschool-aged children. *Parasites & Vectors, 12*(1), 458. https://doi.org/10.1186/s13071-019-3668-4

Mutsaka-Makuvaza, M. J., Matsena-Zingoni, Z., Tshuma, C., Ray, S., Zhou, X. N., Webster, B., et al. (2018). Reinfection of urogenital schistosomiasis in preschool children in a highly endemic district in northern Zimbabwe: A 12 months compliance study. *Infect Dis Poverty, 7*(1), 102. https://doi.org/10.1186/s40249-018-0483-7

Mwai, J., Omogi, J. O., & Abdi, M. H. (2021). Environmental factors influencing prevention and control of schistosomiasis infection in mwea, kirinyaga county Kenya: A cross sectional study. *East Afr Health Res J, 5*(1), 99–105. https://doi.org/10.24248/eahrj.v5i1.657

Nyati-Jokomo, Z., & Chimbari, M. J. (2017). Risk factors for schistosomiasis transmission among school children in Gwanda district, Zimbabwe. *Acta Tropica, 175*, 84–90. https://doi.org/10.1016/j.actatropica.2017.03.033

Pedersen, U. B., Karagiannis-Voules, D. A., Midzi, N., Mduluza, T., Mukaratirwa, S., Fensholt, R., et al. (2017). Comparison of the spatial patterns of schistosomiasis in Zimbabwe at two points in time, spaced twenty-nine years apart: Is climate variability of importance? *Geospat Health, 12*(1), 505. https://doi.org/10.4081/gh.2017.505

Pedersen, U. B., Stendel, M., Midzi, N., Mduluza, T., Soko, W., Stensgaard, A. S., et al. (2014). Modelling climate change impact on the spatial distribution of fresh water snails hosting trematodes in Zimbabwe. *Parasites & Vectors, 7*, 536. https://doi.org/10.1186/s13071-014-0536-0

PopulationPyramid.net. (2023). Demography, population pyramid, age pyramid, aging, retirement, Zimbabwe, 2022. https://www.populationpyramid.net/zimbabwe/2022/.

Qian, M. B., Chen, Y. D., Fang, Y. Y., Xu, L. Q., Zhu, T. J., Tan, T., et al. (2011). Disability weight of clonorchis sinensis infection: Captured from community study and model simulation. *PLoS Neglected Tropical Diseases, 5*(12), Article e1377. https://doi.org/10.1371/journal.pntd.0001377

Rampogu, S. (2023). A review on the use of machine learning techniques in monkeypox disease prediction. *Science in One Health*. , Article 100040. https://doi.org/10.1016/j.soh.2023.100040

Stensgaard, A. S., Booth, M., Nikulin, G., & McCreesh, N. (2016). Combining process-based and correlative models improves predictions of climate change effects on Schistosoma mansoni transmission in eastern Africa. *Geospat Health, 11*(1 Suppl), 406. https://doi.org/10.4081/gh.2016.406

Taylor, P., & Makura, O. (1985). Prevalence and distribution of schistosomiasis in Zimbabwe. *Annals of Tropical Medicine and Parasitology, 79*(3), 287–299. https://doi.org/10.1080/00034983.1985.11811921

Utzinger, J., Raso, G., Brooker, S., De Savigny, D., Tanner, M., Ornbjerg, N., et al. (2009). Schistosomiasis and neglected tropical diseases: Towards integrated and sustainable control and a word of caution. *Parasitology, 136*(13), 1859–1874. https://doi.org/10.1017/S0031182009991600

Van der Werf, M. J., de Vlas, S. J., Brooker, S., Looman, C. W., Nagelkerke, N. J., Habbema, J. D., et al. (2003). Quantification of clinical morbidity associated with schistosome infection in sub-Saharan Africa. *Acta Tropica, 86*(2–3), 125–139. https://doi.org/10.1016/s0001-706x(03)00029-9

Vilne, B., Meistere, I., Grantina-Ievina, L., & Kibilds, J. (2019). Machine learning approaches for epidemiological investigations of food-borne disease outbreaks. *Frontiers in Microbiology, 10*, 1722. https://doi.org/10.3389/fmicb.2019.01722

World Health Organization. (2020). WHO methods and data sources for global burden of disease estimates 2000-2019. *Global Health Estimates*, (3). *Technical Paper WHO/DDI/DNA/GHE/2020*.

World Health Organization. (2021). *Ending the neglect to attain the sustainable development goals: A road map for neglected tropical diseases 2021–2030*.

World Health Organization. (2023). Schistosomiasis (Bilharzia). https://www.who.int/schistosomiasis/en/.

Yang, G. J., & Bergquist, R. (2018). Potential impact of climate change on schistosomiasis: A global assessment attempt. *Travel Medicine and Infectious Disease, 3*(4). https://doi.org/10.3390/tropicalmed3040117

Zheng, J. X., Xia, S., Lv, S., Zhang, Y., Bergquist, R., & Zhou, X. N. (2021). Infestation risk of the intermediate snail host of schistosoma japonicum in the yangtze river basin: Improved results by spatial reassessment and a random forest approach. *Infect Dis Poverty, 10*(1), 74. https://doi.org/10.1186/s40249-021-00852-1

Zhou, X. N., Yang, G. J., Yang, K., Wang, X. H., Hong, Q. B., Sun, L. P., et al. (2008). Potential impact of climate change on schistosomiasis transmission in China. *The American Journal of Tropical Medicine and Hygiene, 78*(2), 188–194. https://www.ncbi.nlm.nih.gov/pubmed/18256410.