IJCRSSSL International Journal of Current Researches in Sciences, Social Sciences and Languages

Volume: 03 | Issue: 04 | 2023 | Open Access | Impact Factor: 5.735

A Data Mining Technology for Controlling Neglected Tropical Diseases (Ntds) in Taraba State, Nigeria

Odesanya Akeem Dele

Department of Computer Science, Federal Polytechnic, Bali. Nigeria

Abstract: The neglected tropical diseases (NTDs) are infectious diseases that principally impact the world's poorest people. They are found in several countries in Africa, Asia, and Latin America. NTDs are especially common in tropical areas where people do not have access to clean water or safe ways to dispose of human waste. They are among the most common conditions globally, affecting approximately one billion people. Many NTDs have long-term consequences, such as visual and physical impairments. As a result, people with NTDs may have difficulties in carrying out activities or participating in society - in other words, NTDs can cause disabilities. This research work will enhance a better understanding of risk factors that could serve as awareness about the disease and a better way of controlling the Neglected Tropical Diseases (NTDs) using a Data Mining Technology. The aim of this research is to develop a system for the controlling schistosomiasis a variant of NTDs in Taraba state Nigeria using the C4.5 decision tree with WEKA 3.6 software. Primary data were collected from Specialist Hospital, Jalingo and Primary healthcare centres across local governments in Taraba State Nigeria.

Keywords: Control, Neglected tropical diseases, Schistosomiasis, Data mining, Taraba state

INTRODUCTION

Schistosomiasis (or liver fluke infection) is caused by trematodes of the genus Schistosoma. These three types of worms' cause schistosomiasis in humans, including two in Nigeria. These are Schistosomes. Haematobium causes urinary schistosomiasis and Schistosoma mansoni causes intestinal schistosomiasis. In Nigeria, schistosomiasis is a serious and increasingly serious disease due to lack of drinking water and activities related to water resources development projects for irrigation, fisheries and hydropower. Overall, the disease mainly affects the rural poor and vulnerable age groups. Students are the main victims of this disease. Schistosomiasis can cause diseases such as bladder cancer, anemia, liver dysfunction, etc. Nigeria is the country most affected by this disease in Africa. In 2006, about 116 million of the estimated 555 million Africans were at risk of the disease (WHO 2008). Schistosomiasis is the most studied neglected tropical disease (NTD) in Nigeria (Oluwole, Bernsah, Dixon, Dean, 2017). Schistosomiasis is a common parasitic disease in tropical developing countries (Gyuse, Ofoezie, & Ogunniyi, 2010). Schistosomiasis (or liver fluke infection) is caused by trematodes of the genus Schistosoma (FMOH, 2015). Schistosomiasis transmission has been reported in 78 countries worldwide, it affects nearly 240 million people worldwide and more than 700 million people live in endemic areas (WHO, 2020). Schistosoma haematobium is endemic to sub-Saharan Africa, the Middle East and some islands in the Indian Ocean. In 2012, 779 million people worldwide were at risk of schistosomiasis (Utzinger, Becker, Knopp, Blum, Neumayr, Keizer, & Hatz, 2012). It is more common in sub-Saharan Africa, where more than 90% of infected people live. About 62% of all cases live in 10 African countries (Nigeria, Ethiopia, Congo, Kenya, Tanzania, Cameroon, Uganda, Malawi, Ghana and Mozambique (WHO, 2010; Bishop, 2017). Nigeria is the country that bears the brunt severity of the disease In Africa, about 116 million of the estimated 555 million Africans were at risk of the disease in 2006.

SYMPTOMS OF SCHISTOSOMIASIS

Symptoms of schistosomiasis are caused by the body's reaction to the worm eggs. Intestinal schistosomiasis can cause abdominal pain, diarrhea, and blood in the coprolite. Liver impingement is

common in advanced cases and is often associated with fluid accumulation in the peritoneum and increased blood pressure of the abdominal vessels. In similar cases, splenic rupture may also occur. A typical sign of urogenital schistosomiasis is hematuria (blood in the urine). Bladder and ureteral fibrosis as well as dislocation lesions are sometimes diagnosed in severe cases.

Bladder cancer is another complication that can occur in the later stages. In women, urinary schistosomiasis can manifest as genital lesions, vaginal bleeding, pain during sex, and lymphadenopathy in the vulva. In men, genitourinary schistosomiasis can cause pathology in the seminal vesicles, prostate, and other organs. This claim can also cause other irreversible long-term consequences, including severity. A child with severe and severe infections will likely experience frequent and irreversible problems later in life, such as cirrhosis of the liver, bladder cancer or treatment failure. Schistosomiasis is characterized in the intestinal or urinary tract, depending on the location of the ulcer in adults (WHO, 2010).

Schistosomiasis transmission is most likely linked to lack of water, sanitation and hygiene. Schistosomiasis transmission is often seasonal, mainly due to changes in temperature and irrigation cycles. approximately 43 million Nigerians require treatment for schistosomiasis (FMOH, 2015).

Research of this nature has become urgent due to the schistosomiasis situation in Taraba state. Various measures have been taken to control, prevent and eliminate schistosomiasis in Nigeria. They include, but are not limited to:

Overall, different means have been used to control and eliminate schistosomiasis in Nigeria, none has implemented or attempted to use data mining to predict or control schistosomiasis in this area. Therefore, this study aimed to control schistosomiasis in Taraba state Nigeria using K-nearest neighbor decision and C4.5 decision tree using WEKA 3.6 software.

The aim of this research is to develop a system that will control the increase of schistosomiasis a variant of Neglected Tropical Diseases (NTDs) using a data mining technology.

LITERATURE REVIEW

Neglected Tropical Diseases (Ntds)

NTD is a chronic disease with clinical characteristics similar to many non-communicable diseases (Tan et al. 2004). Many NTDs and their associated interventions and policies need further evaluation (Lee, Bartsch, & Gorham, 2015).

NTDs are an important public health problem that can be prevented through folic acid supplementation and fortification of essential foods (Zaganjor et.al.2016).

As presented in Lenk et.al. (2016). It is clear that productivity is affected by NTDs, although the actual impact depends on the type and severity of the NTD and the context in which it occurs. About the best way to control and eliminate NTDs:

Banda, Deribe, and Davey (2021) estimate that the data set identifies integrated management of NTDs that should be cost-effective to increase treatment coverage.

Bhaumik et al (2015) expressed the need to increase the priority of systematic reviews of NTDs, especially assessments of the accuracy of diagnostic tests.

However, Hotez and Daar (2008) stated that neglected tropical diseases (NTDs) have chronic, insidious clinical manifestations and cause severe chronic disease syndromes similar to chronic non-communicable diseases. infections (CNCD), such as Chagas disease, urinary schistosomiasis, and trichotillomania.

SCHISTOSOMIASIS

Schistosomiasis is a poverty-related health problem that affects more than 200 million people, and while significant progress has been made in understanding the disease, it is important to strengthen capacity research in countries.). It should be noted that schistosomiasis remains a neglected public health problem with high mortality in endemic and emerging areas (Martins-Melo et. al. 2015).

Studies by Casavechia et.al (2018) have presented epidemiological factors that influence environmental and social conditions on the occurrence of schistosomiasis.

DATA MINING AND HEALTHCARE

Koh and Tan (2005) note that the vast amount of data generated by healthcare transactions is too complex and voluminous to be processed and analyzed using traditional methods. What is driving the use of data mining applications in healthcare is the realization that data mining can generate very useful information for all stakeholders in the healthcare sector. Strong. It automatically detects relevant patterns in the database, using deterministic methods and algorithms to examine current and historical data, which can then be analyzed to predict future trends. hybrid (Ramamohan, Vasantharao, Chakravarti, & Ratnam, 2012).

Data mining is the analysis of large data sets to discover patterns and use those patterns to forecast or predict the likelihood of future events (Crockett et al., 2017). Data mining is a broad field that integrates techniques from many different fields, including machine learning, artificial intelligence, statistics, and pattern recognition to analyze large volumes of data (Archana and Elangovan, 2014).

Many research articles have been published showing the application of data mining algorithms in the medical field such as disease diagnosis and prognosis (Gupta, Kumar, and Sharma 2011), disease prediction (Arafiyah and Hermin, 2018; Lakshmi, Nagesh and VeeraKrishna, 2004; Amosa, Onyeka, Olatunji and Ugwu, 2021; Banu and Gomathy, 2014), as well as survival and disease identification (Delen, et. al, 2005).

METHODOLOGY

The development of this system which employs a Data mining technique for the controls of NEGLECTED TROPICAL DISEASES (NTDs) in Taraba State encompasses the following:

- An environment characterized by the Microsoft Windows Operating System - for the running of a Command Line Interface or Graphical User Interface Program.
- Data and records of patients suffering from various NEGLECTED TROPICAL DISEASES (NTDs) were collected from Specialist Hospital, Jalingo and the Primary healthcare centers across governments in Taraba State Nigeria. The proposed system made use of data collected from 480 records with nine attributes. The flow of the design is in Figure 1. The attributes are Male, Female, Age below 18, Age above 18, Bathing in River or Stream, Washing Clothes in River or stream, Fishing in River or stream, Playing in River or stream, Swimming in River or stream.
- The C4.5 decision tree with WEKA 3.6 software was used for the building and analyzing the System.

The Study Location are:

- Specialist Hospital, Jalingo, and Primary healthcare Centers in Taraba state, (Collection of Medical Records and consultation with domain experts).
- Federal Polytechnic, Bali (Design, Coding, System testing and Documentation).

IJCRSSL
International Journal of Current Researches in Sciences, Social Sciences and Languages
International Journal of Current Researches in Sciences, Social Sciences and Languages
Volume: 03 | Issue: 04 | 2023 | Open Access | Impact Factor: 5.735

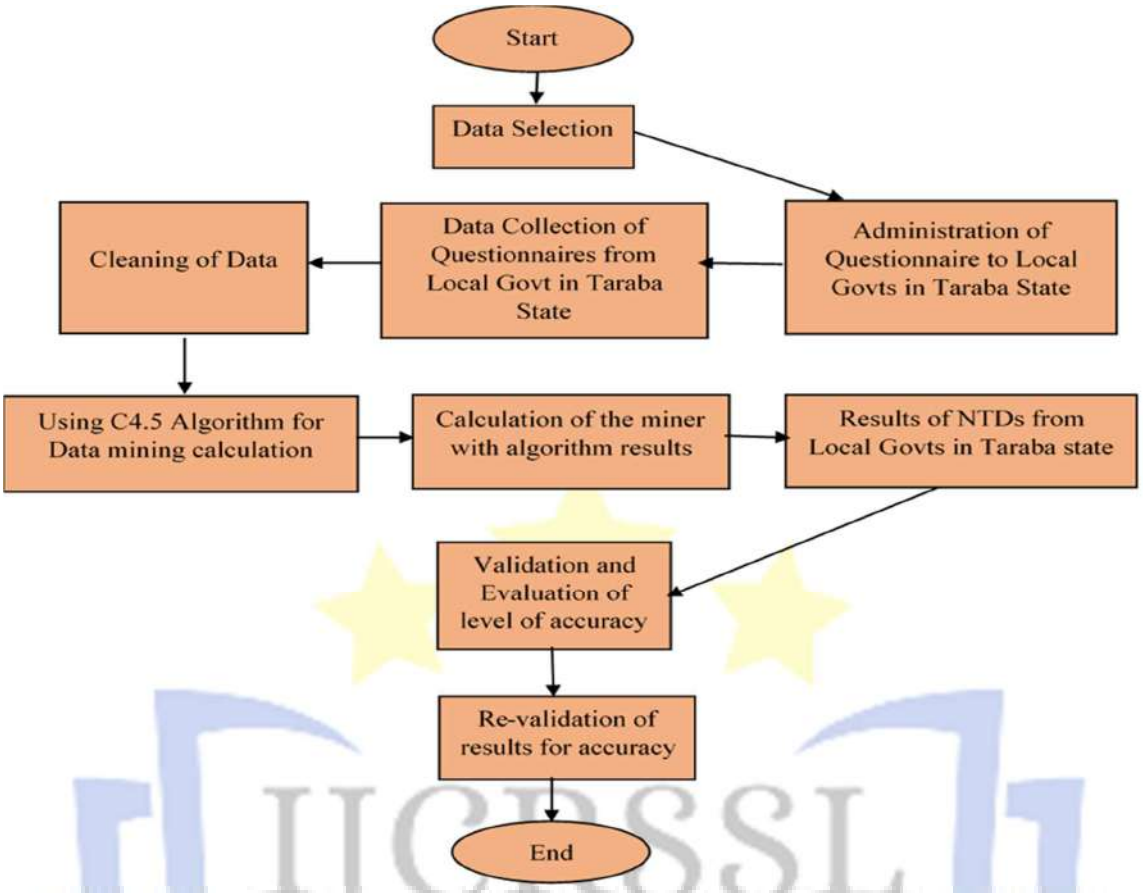

Fig.1 Flow of the Design

Fig. 2 Weka Tool Screen

IJCRSSSL International Journal of Current Researches in Sciences, Social Sciences and Languages

Volume: 03 | Issue: 04 | 2023 | Open Access | Impact Factor: 5.735

WEKA DATA MINING TOOL

WEKA (Figure 2) is one of the favourite tool for data mining in research. It has many powerful features. The steps performed for data mining in WEKA are:

- Data pre-processing and visualization
- Attribute selection
- Classification (Decision trees)
- Prediction (Nearest neighbour)
- Model evaluation
- Clustering (Cobweb, K-means)
- Association rules

FEATURES OF THE TARGET ALGORITHM FOR THE PREDICTION

- Both the discrete and continuous attributes are handled by this algorithm. A threshold value is decided by C4.5 for handling continuous attributes. This value divides the data list into those who have their attribute value below the threshold and those having more than or equal to it.
- This algorithm also handles the missing values in the training data.
- After the tree is fully constructed, this algorithm performs the pruning of the tree. C4.5 after its construction drives back through the tree and challenges to remove branches that are not helping in reaching the leaf nodes.

The C4.5 algorithm is used as the basis for selecting attributes to carry out the testing process in the classification concept.

- Prepare training data. Training data is usually taken from historical data has happened before or is called past data and has been grouped in certain classes;
- Counting the roots of the tree. The root will be taken from the attribute to be selected, with how to calculate the gain value of each attribute, the highest gain value which will be the first root.

Before calculating the gain value of the attribute, calculate first the entropy value;

- Calculate the Gain value;
- Repeat step 2 and step 3 until all records are partitioned;
- The decision tree partitioning process will stop when:
- All records in node N have the same class.
- There are no attributes in the partitioned record anymore and no mores records in the empty branch.

Table 1 Dataset used in WEKA

Attribute	Abbreviation	Value of Attribute	Scale of Measurement
Gender	GD	Male, Female	Nominal
Below 18 years	BE	Number	Numeric
Above 18 years	AE	Number	Numeric
Bathing in River or Stream	BIR	Number	Numeric
Washing Clothes in River or stream	WSR	Number	Numeric
Fishing in River or stream	FIR	Number	Numeric
Playing in River or stream	PIR	Number	Numeric
Swimming in River or stream	SIR	Number	Numeric

IJCRSSL International Journal of Current Researches in Sciences, Social Sciences and Languages

Volume: 03 | Issue: 04 | 2023 | Open Access | Impact Factor: 5.735

PERFORMANCE EVALUATION

The dataset used in WEKA is in Table 1 and the comparison of the different data mining algorithms is presented in Table 2. The formula to calculate accuracy is:

$$\text{Accuracy (TA)} = \frac{TP + TN}{TP + TN + FP + FN}$$

In the equation TA represents Total Accuracy, TP is True Positive, TN is True Negative, FP is False Positive and FN is False Negative.

The results of C4.5 algorithm are quite significant as compared to other algorithms in data mining in Table 1. All the algorithms other than C4.5 algorithm have accuracy rate below than 78.08% but the accuracy of the proposed algorithm is 98.85%. Also, the output of the Accuracy and Error value of the compared Algorithms is in Figure 3.

Algorithm	Accuracy	Error
Naive Bayes	73.4506	26.4506
DT	73.8081	26.1919
KNN	78.0887	21.9113
SVM	68.0991	31.9009
RF	72.7134	27.2866
ANN	74.4829	25.5171
C4.5	98.8530	1.15

Table 2 : Performance comparison between different algorithms

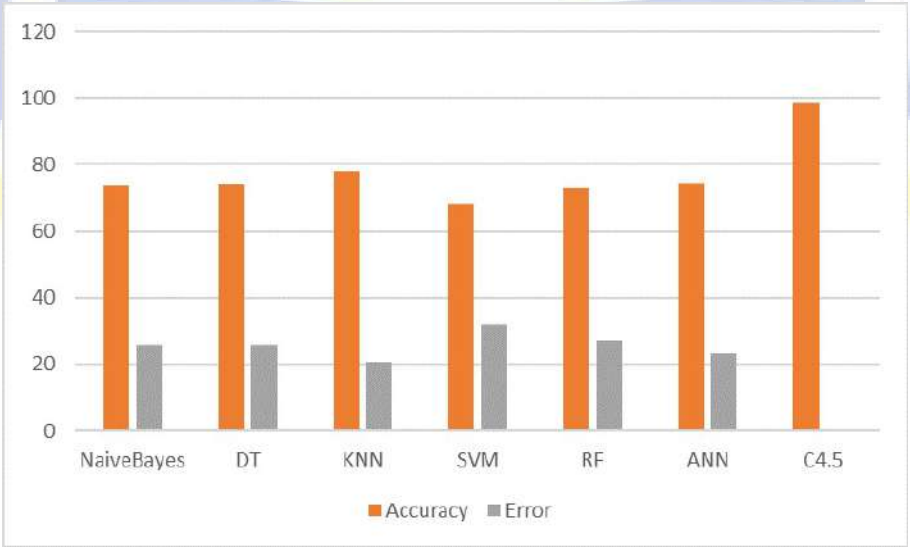

Fig. 3 Output of the Accuracy and Error value of the compared Algorithms

DISCUSSIONS

Preparation and Selection Of Data

Variables that affects the possible infection of humans by the Neglected Tropical Diseases (NTDs) were used and processed in the implementation of various data mining Algorithms. The study identified the prediction and classification accuracy of the algorithms. Naive Bayes (73.45), DT (73.80), KNN (78.08), SVM (68.09), RF (72.71), ANN (74.48) and C4.5 (.85). The performance of Algorithms was based on the presented datasets. It became obvious from the results that C4.5 Algorithm had the best accuracy percentage, hence the best for determining the prediction of prevalence of Neglected Tropical Diseases (NTDs) in Taraba state, Nigeria.

CONCLUSION

In the study, seven algorithms were presented and implemented. Their performance was duly compared. The decision of C4.5 algorithm shows that from the datasets collected from the primary source, health practitioners can actually find the best solution to treatment of Neglected Tropical Diseases (NTDs) in Taraba State. Further research can be made with more datasets and attributes. Also, another sets of algorithms may be used the classifications and comparison.

RECOMMENDATIONS

- Early diagnosis of the Neglected Tropical Diseases (NTDs) may be actualize through the implementation of the data mining technique
- A control plan may be created by using the result of this study
- The technique used may be of assistance to health workers to make effective and accurate decisions on the prevalent of Neglected Tropical Diseases (NTDs).

REFERENCES

[1] Amosa B. M. G., Onyeka N. C., Olatunji F. and Ugwu J., (2021). Using Data Mining to Predict the prevalence of Chronic Diarrhea. International Journal of Advances in Engineering and Management (IJAEM), vol. 3 (1), Jan-Feb, 2021.

[2] Arafiyah, R. and Hermin F. (2018). Data mining for dengue hemorrhagic fever (DHF) prediction with naive Bayes method J. Phys.: Conf. Ser. 948 012077

[3] Archana, S. and Elangovan, K. (2014). Survey of Classification Techniques in Data Mining, International Journal of Computer Science and Mobile Applications, 2 (2) 65-71.

[4] Banda, G., Deribe, K., and Davey, G. (2021). How can we better integrate the prevention, treatment? control and elimination of neglected tropical diseases with other health interventions? A systematic review. BMJ Global Health, 6. https://doi.org/10.1136/bmjgh-2021-006968

[5] Banu M.A and Gomathy B. (2014). Disease Forecasting System Using Data Mining Methods 978-1-4799-3966-4/14 $31.00 © 2014 IEEE, 2014 International Conference on Intelligent Computing Applications

[6] Bhaumik, S., Karimkhani, C., Czaja, C., Williams, H., Rani, M., Nasser, M., Boyers, L., Dmitruk, S., and Dellavalle, R. (2015). Identifying gaps in research prioritization: The global burden of neglected tropical diseases as reflected in the Cochrane database of systematic reviews. Journal of Family Medicine and Primary Care, 4, 507 - 513. https://doi.org/10.4103/2249-4863.174266.

[7] Casavechia, M., Melo, G., Fernandes, A., Castro, K., Pedroso, R., Santos, T., and Teixeira, J. (2018). Systematic review and meta-analysis on Schistosoma mansoni infection prevalence, and associated risk factors in Brazil. Parasitology, 145, 1000 - 1014. https://doi.org/10.1017/S0031182017002268.

[8] Crockett D., Johnson R., and Eliason B. (2017). What is Data Mining in Healthcare? http://www.healthcatalyst.com/what-is-data-mining-in-healthcare retrieved on 29th March, 2023.

[9] Delen, D., Walker, G. and Kadam, A. (2005). Predicting breast cancer survivability: a comparison of three data mining methods Artificial intelligence in medicine 34 (2), 113-127, 2005.

[10] FMOH (2015). Neglected Tropical Diseases Nigeria Multi - Year Master Plan 2015 – 2020 https://www.health.gov.ng/doc/NTD-MASTER-PLAN.

[11] Gupta S., Kumar D., and Sharma A. (2011). Data Mining Classification Techniques Applied for breast cancer diagnosis and prognosis Indian Journal of Computer Science and Engineering (IJCSE) Vol. 2 No. 2

[12] Gyuse, K. I., Ofoezie, I. E., and Ogunniyi, T. A. B. (2010). The effect of urinary schistosomiasis on the health of children in selected rural communities of Osun State, Nigeria. J Trop Med Parasitol, 33, 7-16.

[13] Hotez, P., and Daar, A. (2008). The CNCDs and the NTDs: Blurring the Lines Dividing Noncommunicable and Communicable Chronic Diseases. PLoS Neglected Tropical Diseases, 2. https://doi.org/10.1371/journal.pntd.0000312.

[14] Koh, H. C and Gerald Tan, G. (2005). Data Mining Applications in Healthcare. Journal of Healthcare Information Management - Vol. 19, No. 2

[15] Lakshmi K.R., Nagesh Y. and VeeraKrishna M. (2014). Performance Comparison of three Data Mining Techniques for predicting Kidney Dialysis Survivability. International Journal of Advances in Engineering & Technology, Vol. 7, Issue 1, pp. 242-254

[16] Lee, B., Bartsch, S., and Gorham, K. (2015). Economic and financial evaluation of Neglected Tropical Tiseases. Advances in parasitology, 87, 329-417.

[17] Lenk, E., Redekop, W., Niessen, P., Luyendijk, M., Rijnsburger, A., and Severens, J. (2016). Productivity Loss Related to Neglected Tropical Diseases Eligible for Preventive Chemotherapy: A Systematic Literature Review. PLoS Neglected Tropical Diseases, 10. https://doi.org/10.1371/journal.pntd.0004397.

[18] Martins-Melo, F., Pinheiro, M., Junior, A., Alencar, C., Bezerra, F., and Heukelbach, J. (2015). Spatiotemporal Patterns of Schistosomiasis-Related Deaths, Brazil, 2000–2011. Emerging Infectious Diseases, 21, 1820 - 1823. https://doi.org/10.3201/eid2110.141438.

[19] Oluwole A. S., Bernsah L. D., Dixon R., and Dean L. (2017). A Scoping Review of Nigerian Neglected Tropical Disease Literature: Annotated Bibliography. Countdown: calling time on Neglected Tropical Diseases. www.countdownonntds.org

[20] Ramamohan, Y., K. Vasantharao, K., Chakravarti, C. K. and Ratnam, A. S. (2012), A Study of Data Mining Tools in Knowledge Discovery Process. International Journal of Soft Computing and Engineering (IJSCE) ISSN: 2231-2307, Volume-2, Issue-3

[21] Tan, M., Kusriastuti, R., Savioli, L., and Hotez, P. (2014). Indonesia: An Emerging Market Economy Beset by Neglected Tropical Diseases (NTDs). PLoS Neglected Tropical Diseases, 8. https://doi.org/10.1371/journal.pntd.0002449.

[22] Utzinger U., Becker S., Knopp S., Blum J, Neumayr A., Keiser J., Hatz J (2012). Neglected Tropical Diseases: diagnosis, clinical management, treatment and control Swiss Medical Weekly • Swiss Med Wkly.142: w13727. www.smw.ch

[23] WHO (2010) First WHO report on neglected tropical diseases: Working to overcome the global impact of neglected tropical diseases. https://www.who.int/neglected_diseases/resources/9789241564090/en/

[24] WHO (2020). Ending the neglect to attain the Sustainable Development Goals: A road map for neglected tropical diseases 2021–2030 https://www.who.int/neglected_diseases/WHONTD-roadmap-2030/en retrieved 10th January, 2023

[25] Zaganjor, I., Sekkarie, A., Tsang, B., Williams, J., Razzaghi, H., Mulinare, J., Sniezek, J., Cannon, M., and Rosenthal, J. (2016). Describing the Prevalence of Neural Tube Defects Worldwide: A Systematic Literature Review. PLoS ONE, 11. https://doi.org/10.1371/journal.pone.0151586.