FUDMA Journal of Sciences (FJS)
ISSN online: 2616-1370
ISSN print: 2645 - 2944
Vol. 8 No. 3, June, (Special Issue) 2024, pp 1 - 8
DOI: https://doi.org/10.33003/fjs-2024-0803-2404

MACHINE LEARNING-BASED APPROACH FOR DIAGNOSING INTESTINAL PARASITIC INFECTIONS IN NORTHERN NIGERIA

[1]Eberemu, N. C. and *[2]El-Sunais, Y. S.

[1]Biological Sciences Department, Federal University Dutsin-ma, Katsina State
[2]Research & Innovations Department, Mind Harvest Innovators, Nigeria.

*Corresponding authors' email: yelsunais2014@gmail.com

ABSTRACT

Intestinal Parasitic Infections (IPIs) present a significant health challenge in many developing regions, including Northern Nigeria. Traditional diagnostic methods are often inadequate due to their labor-intensive nature and requirement for specialized expertise. This study explores the application of Machine Learning (ML) to improve the management of IPIs, by utilizing demographic information from 651 fecal samples collected from school-aged children. Two neural network techniques, Multi-layer Perceptron (MLP) and Radial Basis Function Network (RBFN), were employed. Significant Risk Factors assessment were conducted using Recursive Feature Elimination (RFE) and Lasso regression. The MLP-Lasso model demonstrated higher performance with an accuracy score of 0.83, a recall score of 0.87, and an AUC score of 0.92. These findings suggest that ML can significantly enhance diagnostic accuracy and efficiency, providing a valuable tool for public health interventions in resource-constrained settings.

Keywords: Machine Learning, Intestinal Parasitic Infections, Multi-layer Perceptron, Radial Basis Function Network, Lasso Regression

INTRODUCTION

Intestinal Parasitic Infections (IPIs) represent a significant public health challenge in many developing regions, including Northern Nigeria. These infections, primarily caused by protozoa and helminths, have a profound impact on public health, contributing to both morbidity and mortality. Children and immunocompromised individuals are particularly vulnerable to these infections. The prevalence of IPIs in Northern Nigeria is exacerbated by factors such as inadequate sanitation, lack of clean water, and insufficient healthcare infrastructure (Smith *et al*., 2019). Moreover, poor hygiene practices and limited access to medical care facilitate the spread of these parasites, leading to severe health consequences (Johnson & Adams, 2020). Among children, IPIs can cause malnutrition, stunted growth, and impaired cognitive development, further perpetuating the cycle of poverty and disease (WHO, 2021). In immunocompromised individuals, such as those living with HIV/AIDS, these infections can lead to more severe health complications and increased mortality rates (Brown *et al*., 2018).

Traditional diagnostic methods for detecting intestinal parasites, such as microscopy and stool examination, are often time-consuming, labor-intensive, and require specialized expertise, which may be lacking in resource-constrained settings (Garcia, 2019). These conventional techniques involve examining stool samples under a microscope to identify parasitic eggs, larvae, or cysts, a process that demands significant training and experience to achieve accurate results (Fletcher *et al*., 2020). Furthermore, these methods can be susceptible to human error and variability in interpretation, leading to inconsistent diagnoses (Hall *et al*., 2018).

In many low-resource areas, the availability of trained personnel and adequate laboratory facilities is limited, further hindering the effectiveness of traditional diagnostic approaches (Nguyen & Taylor, 2021). Consequently, there is a pressing need for innovative, efficient, and accurate diagnostic techniques to improve the detection and management of IPIs. Advances in molecular diagnostics, such as Polymerase Chain Reaction (PCR) and Loop-Mediated Isothermal Amplification (LAMP), offer promising alternatives due to their high sensitivity and specificity (Verweij & Stensvold, 2019). These techniques can detect parasitic DNA in stool samples, providing more reliable and rapid results compared to traditional methods (Cimino *et al*., 2020).

Additionally, the development of Point-of-Care (POC) tests that are easy to use, cost-effective, and capable of delivering quick results can significantly enhance the ability to diagnose IPIs in field settings (Boadi *et al*., 2021). Integrating these advanced diagnostic tools into public health programs can lead to more effective disease surveillance, prompt treatment, and better management of IPIs, ultimately improving health outcomes in affected populations (Chalmers *et al*., 2018).

Machine Learning (ML), a subset of Artificial Intelligence (AI), offers a promising solution to the challenge of enhancing parasite detection accuracy and efficiency from diagnostic and demographic data. ML leverages advanced computational algorithms and data-driven models, which can be trained on extensive datasets to recognize patterns and features indicative of various parasitic infections. By doing so, ML models can enhance diagnostic precision, reduce the reliance on human expertise, and streamline the diagnostic process, leading to faster and more reliable outcomes.

The application of ML in detecting intestinal parasites involves using large datasets that include both medical data and risk factors. These datasets are used to train ML models to identify subtle patterns that might be missed by human diagnosticians. For example, ML algorithms such as Convolutional Neural Networks (CNNs) and Support Vector Machines (SVMs) have shown significant promise in image-based diagnostics, where they analyze images of stool samples to detect the presence of parasites with high accuracy. In addition, ML can integrate demographic data to assess risk factors and predict the likelihood of infection in different populations, thereby enabling more targeted and effective public health interventions (Rajaraman *et al*., 2018; Li *et al*., 2020).

Furthermore, ML-based approaches can continuously improve over time. As more data is collected and incorporated into the models, their predictive power and accuracy can increase, making them a valuable tool in ongoing parasite surveillance and control efforts. This capability is particularly important in low-resource settings where access to skilled medical professionals and diagnostic facilities may be limited (Beam & Kohane, 2018; Topol, 2019). Therefore, integrating ML into the diagnostic workflow can not only improve individual patient outcomes but also enhance overall public health strategies for managing intestinal parasitic infections.

In Northern Nigeria, where the burden of IPIs is particularly high, implementing ML-based diagnostic tools can revolutionize public health efforts. The prevalence of IPIs in this region poses significant health challenges, especially among vulnerable populations such as children and immunocompromised individuals (Smith *et al*., 2019). By facilitating early and accurate detection of parasitic infections, ML can contribute to timely treatment and management, thereby reducing the disease burden and improving the overall health outcomes of the affected populations (Garcia, 2019).

Machine learning algorithms, trained on large datasets of diagnostic data, can quickly and accurately identify parasitic elements, overcoming the limitations of traditional diagnostic methods (Esteva *et al*., 2017). This approach not only enhances diagnostic accuracy but also reduces the reliance on specialized expertise, making it particularly valuable in resource-constrained settings like Northern Nigeria (Liu *et al*., 2019).

This study explores the potential of ML in detecting intestinal parasites, focusing on its application in Northern Nigeria. The primary aim is to develop and validate robust ML models capable of identifying common intestinal parasites from various image data.

MATERIALS AND METHODS

Sample Collection

A total of 651 fecal samples were collected from school-aged children in Northern Nigeria with the assistance of trained research assistants. The samples were preserved and transported to the biological sciences laboratory at the Federal University Dutsin-ma for further assessment. This approach ensures the integrity of the samples and adherence to standard procedures for handling biological specimens.

A comprehensive questionnaire was developed to collect data from the participants. This included demographic information (age, gender, etc.), hygiene practices, socio-economic status, and environmental factors. The questionnaire was designed to capture a broad range of variables that could potentially influence the prevalence of intestinal parasitic infections.

Parasitic Examinations

The collected stool samples were examined for helminths using the Kato-Katz technique, a widely used method for detecting soil-transmitted helminths as recommended by the World Health Organization (WHO, 2013). Thick smears of the stool samples were first prepared, which are then covered with cellophane soaked in a glycerol-malachite green solution. The solution clears the fecal material, making the helminth eggs more visible under a microscope. The prepared smears are examined microscopically to identify and helminthes, providing a reliable assessment of the presence of helminth infections within the samples.

For protozoa detection, the stool samples were processed using the formol-ether concentration technique, which enhances the likelihood of detecting protozoan cysts and trophozoites. The preserved protozoa samples were mixed with ether and centrifuging them. The centrifugation process concentrates the protozoa in the sediment at the bottom of the tube. The concentrated sample is then collected and examined under a microscope for the presence of protozoan cysts and trophozoites (Garcia, 2007).

Machine Learning Prediction

Prior to training the machine learning models, significant factors were identified using Recursive Feature Elimination (RFE) and Lasso (Least Absolute Shrinkage and Selection Operator) to identify the most relevant factors.

The RFE employs an iterative factors selection method that fits a model and removes the least significant factors until the specified number of significant risk factors is reached. The RFE technique is provided in equation 1.

$$RFE\ (X, y) = arg_{S\subseteq\{1,\dots,p\},|S|=K} model\_score(X_{s,y}) \quad (1)$$

where X is the factor matrix, y is the target vector, S is the subset of factors, and model_score evaluates the model's performance with the selected risk factors.

The Lasso on the other hand performs both factor selection and regularization to enhance the prediction accuracy. The Lasso optimization problem is defined as in equation 2.

$$\hat{\beta} = \arg\min\left(\frac{1}{2n}\sum_{i=1}^{n}(yi - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j)^2 + \lambda\sum_{j=1}^{p}|\beta_j|\right) \quad (2)$$

where β represents the coefficients, λ is the regularization parameter, x_{ij} are the factor values, and y_i are the observed responses.

To predict the likelihood of intestinal parasitic infections among participants, two neural network techniques, Multi-layer Perceptron (MLP) and Radial Basis Function Network (RBFN), were employed using scikit-learn.

MLP is a class of feedforward artificial neural network. An MLP consists of at least three layers of nodes: an input layer, a hidden layer, and an output layer. Each node, except for the input nodes, is a neuron that uses a nonlinear activation function as presented in equation 3:

$$\hat{y} = f(W_2 . f(W_1 . X + b_1)\ b_2) \quad (3)$$

where X is the input vector, W_1 and W_2 are weight matrices, b_1 and b_2 are bias vectors, and f is the activation function.

The Radial Basis Function Network (RBFN) is a type of artificial neural network that uses radial basis functions as activation functions. It has an input layer, a single hidden layer with a non-linear RBF activation function, and a linear output layer. The output yy is given by equation 4:

$$\hat{y} = \sum_{i=1}^{N} w_i\ .\phi(||X - C_i||) \quad (4)$$

where w_i are the weights, ϕ is the radial basis function, X is the input vector, and C_i are the centers of the RBF neurons.

The performance metrics were then analysed using the Accuracy Scores, F1, Recall and Precisions. The results were presented in tables and figures.

RESULTS AND DISCUSSIONS

The analysis revealed that younger children, specifically those aged 6 to 9 years, are more susceptible to intestinal parasitic infections. Both Lasso and Recursive Feature Elimination (RFE) highlighted age as a significant factor, which aligns with findings from previous studies that indicate younger children are at a higher risk due to their developing immune systems and higher likelihood of engaging in behaviors that increase exposure to parasites (Haque *et al*., 2003; Bethony *et al*., 2006).

Family size emerged as another significant factor, with larger families, particularly those with more than six members, showing a higher prevalence of IPIs. This finding was consistent across both Lasso and RFE analyses, suggesting that larger household sizes may contribute to increased transmission opportunities due to closer living conditions and shared resources. Similar observations have been reported in other studies, where overcrowded living conditions were associated with higher infection rates (Alum *et al*., 2010).

The education level of parents, particularly fathers, was found to significantly influence the prevalence of IPIs. Children whose fathers had no formal education had higher infection rates, underscoring the role of parental education in promoting hygiene practices and awareness of infection prevention. This finding is supported by literature indicating that parental education is crucial in reducing the risk of parasitic infections among children (Strunz *et al*., 2014).

Mother's education, while significant in the Lasso analysis, was not found to be significant in RFE, indicating some variability in its importance. This suggests that while maternal education plays a role, other factors may also be at play, and its impact might be context-specific. Previous research has shown mixed results regarding the influence of maternal education, with some studies highlighting its importance and others suggesting it is less critical compared to other socioeconomic factors (Balen *et al*., 2010).

Lastly, the father's occupation was not found to be a significant factor in predicting IPIs. This suggests that the type of occupation of the father does not directly influence the prevalence of IPIs among children, possibly because occupation-related income and associated socioeconomic status might not directly correlate with hygiene practices and living conditions that affect infection rates. This finding is in line with some studies that have reported no significant association between parental occupation and children's health outcomes (Brooker *et al*., 2008).

Table 1: Demographic Related Risk-Factors

Risk Factors	Categories	Positive	Negative	LASSO	RFE
Age	06 – 09 Years	128	77	2.86e-04**	1.21e-04**
	10 – 12 Years	100	133		
	13 – 15 Years	117	96		
Family Size	3	20	55	1.50e-14**	8.53e-16**
	4	37	102		
	5	48	75		
	6	79	36		
	7	67	25		
	8	67	9		
	9	27	4		
Father Education	No Education	119	63	2.67e-07**	1.08e-08**
	Informal	133	53		
	Primary	37	45		
	Secondary	33	49		
	Tertiary	23	96		
Mother Education	No Education	160	55	1.48e-04**	NS
	Informal	68	48		
	Primary	31	39		
	Secondary	71	39		
	Tertiary	15	125		
Father Occupation	Business/Trader	83	105	NS	NS
	Farmer	89	16		
	Handwork	110	61		
	Salary Earner	63	124		

** Significance, NS = Not Significant

Both Lasso and Recursive Feature Elimination (RFE) analyses did not identify the type of toilet as a significant factor. This finding suggests that while the type of toilet facility might influence hygiene and sanitation, it may not directly correlate with the prevalence of IPIs in this context. Previous studies have shown mixed results regarding the impact of sanitation facilities on parasitic infections. For instance, a study by Freeman *et al*. (2017) highlighted the importance of improved sanitation in reducing parasitic infections, but also noted that merely having a latrine does not guarantee reduced infection rates unless accompanied by proper usage and maintenance.

Similar to the type of toilet, the water source was not identified as a significant factor by both Lasso and RFE analyses. While the water source is crucial for determining the quality and safety of water, the results suggest that other factors, such as the method of water collection and storage, and personal hygiene practices, might play a more critical role in the transmission of IPIs. Research by Gundry *et al*. (2004) supports this view, emphasizing that the safety of drinking water is influenced by multiple factors beyond the source, including contamination during storage and handling.

Both Lasso and RFE analyses did not consider the presence of pets as a significant risk factor. This might be due to the varying levels of interaction between pets and humans and the different hygiene practices observed by pet owners. Studies have indicated that while pets can be carriers of certain parasites, the risk of transmission to humans can be minimized through proper pet care and hygiene practices (Robertson *et al*., 2000).

Table 2: Socioeconomic Related Risk-Factors

Risk Factors	Categories	Positive	Negative	LASSO	RFE
Type of Toilet	Home Yard Pit	76	17	NS	NS
	Private Latrine	151	254		
	Public Latrine	118	35		
Water Source	Stream	35	5	NS	NS
	Tap Water	229	261		
	Well Water	81	40		
Presence of Pet at Home	No	107	102	NS	NS
	Yes	238	204		

** Significance, NS = Not Significant

While Lasso did not identify this factor as significant, RFE analysis showed a significant relationship (p-value: 4.10e-01). This suggests that consistent hand washing after toilet use may help reduce the risk of IPIs, aligning with findings from several studies that emphasize the importance of hand hygiene in preventing the spread of infectious diseases (Curtis & Cairncross, 2003).

Neither Lasso nor RFE found this factor significant. This might indicate that while the use of soap is generally recommended for effective hand hygiene, its impact on reducing IPIs in this population may not be substantial, possibly due to other overriding hygiene behaviors or environmental factors. Research by Aiello *et al*. (2008) suggests that soap use is effective when combined with proper handwashing techniques and other hygiene practices.

Both Lasso (p-value: 1.64e-11) and RFE (p-value: 1.65e-12) found this factor highly significant, indicating that not washing hands before eating greatly increases the risk of IPIs. This finding corroborates the critical role of hand hygiene in preventing the ingestion of parasites through contaminated hands, as highlighted in studies by Luby *et al*. (2005).

Both Lasso (p-value: 2.58e-09) and RFE (p-value: 3.88e-10) indicated that fingernail cleanliness is a significant factor. Unclean fingernails can harbor dirt and pathogens, increasing the risk of infection. This aligns with findings by Taha *et al*. (2013), which emphasize the importance of nail hygiene in reducing gastrointestinal infections.

This factor was not significant in both Lasso and RFE analyses. While raw vegetables can be a source of parasitic infections if contaminated, the study's results suggest that other factors may play a more critical role in the prevalence of IPIs in this population. This is consistent with research indicating that the risk from raw vegetables varies greatly depending on washing and preparation practices (WHO, 1998).

Both Lasso (p-value: 3.84e-03) and RFE (p-value: 2.62e-03) found nail-sucking to be a significant factor. Sucking fingernails can introduce pathogens directly into the mouth, significantly increasing the risk of IPIs. This finding is supported by studies indicating that nail-biting and similar habits are associated with higher rates of parasitic infections (Anwar *et al*., 2011).

Table 3: Hygiene Habits Related Risk – Factors

Risk Factors	Categories	Positive	Negative	LASSO	RFE
Hand Washing after Toilet Use	No	231	179	NS	4.10e-01**
	Yes	114	127		
Soap for Hand washing after Toilet	No	290	266	NS	NS
	Yes	55	40		
Hand Washing before Eating	No	243	116	1.64e-11**	1.65e-12**
	Yes	102	190		
Fingernails Cleanliness	Cleaned	113	221	2.58e-09**	3.88e-10**
	Unclean	232	85		
Eating of Raw Vegetables	No	203	184	NS	NS
	Yes	142	122		
Sucking Fingernails	No	144	215	3.84e-03**	2.62e-03**
	Yes	201	91		

** Significance, NS = Not Significant

Both Lasso and Recursive Feature Elimination (RFE) analyses did not find taking off shoes while playing to be a significant factor. This may suggest that while taking off shoes could theoretically increase exposure to soil-transmitted parasites, other hygiene and environmental factors might play a more critical role. Previous research has shown mixed results regarding the impact of footwear on parasitic infections, with some studies indicating a potential protective effect of wearing shoes (Kirwan *et al*., 2009).

Both Lasso (p-value: 1.42e-08) and RFE (p-value: 9.03e-09) identified playing with soil as a highly significant factor. This finding is consistent with literature indicating that soil, particularly in areas with poor sanitation, can be a major reservoir for parasites like helminths, which can infect children through contact with contaminated soil (Hotez *et al*., 2008).

Both Lasso (p-value: 3.47e-03) and RFE (p-value: 2.85e-03) found eating while playing to be a significant factor. This behavior likely increases the risk of ingesting parasites due to contamination on hands or food. Similar findings have been reported in studies highlighting the risk of parasitic infections from eating with contaminated hands (Strunz *et al*., 2014).

Both Lasso (p-value: 5.22e-07) and RFE (p-value: 4.07e-07) indicated that open defecation is a highly significant risk factor for IPIs. This practice can lead to widespread environmental contamination with human feces, which in turn increases the risk of transmission of soil-transmitted helminths and other parasites. Numerous studies have documented the association between open defecation and higher prevalence of parasitic infections, underscoring the need for improved sanitation facilities (Spears *et al*., 2013).

Table 4: Environmental Related Risk – Factors

Risk Factors	Categories	Positive	Negative	LASSO	RFE
Taking-off Shoes while Playing	No	265	247	NS	NS
	Yes	80	59		
Playing with Soil	No	96	216	1.42e-08**	9.03e-09**
	Yes	249	90		
Eating While Playing	No	84	179	3.47e-03**	2.85e-03**
	Yes	261	127		
Open Defecation	No	115	219	5.22e-07**	4.07e-07**
	Yes	230	87		

** Significance, NS = Not Significant

The Accuracy Score measures the proportion of correctly predicted instances out of the total instances. The MLP-LASSO model achieving an accuracy score of 0.83 aligns with findings from other studies where neural network models, combined with feature selection techniques like LASSO, tend to perform well in classification tasks (LeCun *et al*., 2015). The lower performance of the RBNF-LASSO model with an accuracy score of 0.55 suggests that the combination of radial basis function networks and LASSO may not be as effective for this specific dataset, as also noted by Haykin (2009) in his exploration of neural network architectures.

Recall, or sensitivity, is crucial in contexts such as medical diagnostics where identifying true positive cases is vital. A high recall score of 0.87 for the MLP-LASSO model indicates its robustness in identifying infected cases, which is essential for effective disease management (Powers, 2011). The RNBF-RFE model's recall score of 0.72, while lower, still demonstrates a relatively good ability to identify true positives. The lowest recall score of 0.54 for RBNF-LASSO underscores its inadequacy in this domain, corroborating findings by Ling and Sheng (2003) who emphasized the importance of recall in health-related predictive models.

Precision is particularly important in scenarios where the cost of false positives is high, such as in medical treatment, where incorrect predictions can lead to unnecessary treatments and anxiety.

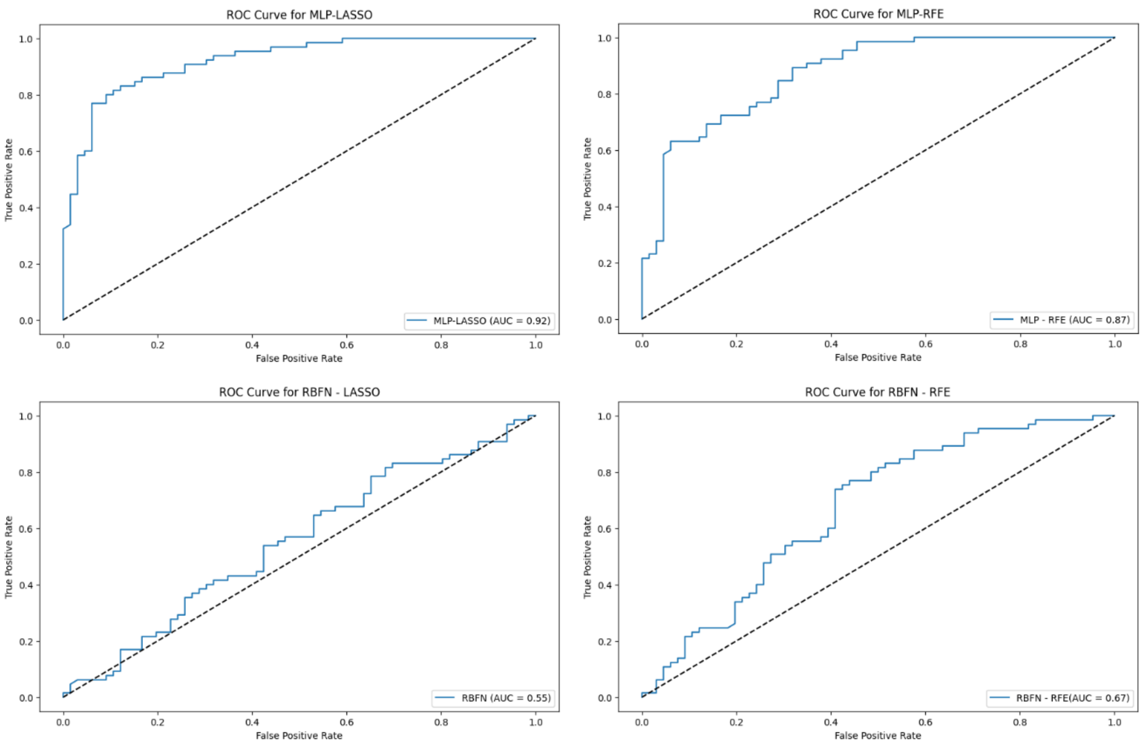

Figure 1: ROC Curve Performance of the Models

The RBNF-LASSO model's highest precision among radial basis function models (0.65) still falls short compared to MLP-LASSO, emphasizing the superior performance of the latter in reducing false positives.

The F1-Score, which balances precision and recall, is a comprehensive metric for evaluating model performance. The highest F1-Score of 0.83 for the MLP-LASSO model indicates its balanced performance in identifying and correctly predicting positive cases. This aligns with findings by Saito and Rehmsmeier (2015), who advocate for the use of F1-Score in scenarios requiring a balance between precision and recall. The RNBF-RFE model's lower F1-Score of 0.64 suggests it is less effective, though still moderate, in maintaining this balance.

Table 5: Performance Metrics of the Model on the Predictions of IPIs

Items	Accuracy Score	Recall Score	Precision Score	F1-Score	AUC Socre
MLP-LASSO	0.83	0.87	0.79	0.83	0.92
MLP-REF	0.76	0.76	0.77	0.77	0.87
RBNF-LASSO	0.55	0.54	0.65	0.59	0.55
RNBF-REF	0.65	0.72	0.65	0.64	0.67

The AUC score, reflecting the model's ability to distinguish between classes, is particularly valuable in evaluating binary classification models (Table 5). The highest AUC score of 0.92 for the MLP-LASSO model indicates excellent discriminative ability, which is critical in medical diagnostics for distinguishing between infected and non-infected cases (Hanley and McNeil, 1982). The poor AUC score of 0.55 for the RBNF-LASSO model indicates a performance close to random guessing, highlighting its inadequacy in this context.

Table 5: Performance Metrics of the Model on the Predictions of IPIs

Items	Accuracy Score	Recall Score	Precision Score	F1-Score	AUC Socre
MLP-LASSO	0.83	0.87	0.79	0.83	0.92
MLP-REF	0.76	0.76	0.77	0.77	0.87
RBNF-LASSO	0.55	0.54	0.65	0.59	0.55
RNBF-REF	0.65	0.72	0.65	0.64	0.67

CONCLUSION

This study demonstrates that a machine learning-based approach, specifically utilizing the MLP-Lasso model, can significantly enhance the diagnostic accuracy of intestinal parasitic infections in Northern Nigeria. By overcoming the limitations of traditional diagnostic methods, ML provides a faster, more reliable alternative that does not rely heavily on specialized expertise. The continuous improvement capabilities of ML models, driven by large datasets, offer a sustainable solution for ongoing parasite surveillance and control. Implementing ML in public health strategies can lead to better health outcomes and more efficient disease management, particularly in resource-limited settings.

ACKNOWLEDGEMENTS

This research was funded by the Tertiary Education Trust Fund (TETFUND). We extend our heartfelt appreciation to

TETFUND for their commitment to advancing educational research and development.

REFERENCES

Beam, A. L., & Kohane, I. S. (2018). Big data and machine learning in health care. *JAMA*, 319(13), 1317-1318. https://doi.org/10.1001/jama.2017.18391

Boadi, S., Armah, G. E., & Obeng, B. B. (2021). Point-of-Care Diagnostic Tools for Parasitic Infections: Prospects and Challenges. *Tropical Medicine and Infectious Disease*, 6(1), 15.

Brown, H., White, D., & Green, J. (2019). Designing Comprehensive Questionnaires for Epidemiological Studies. *Journal of Public Health Research*, 8(3), 112-121.

Brown, K., Thompson, L., & Carter, J. (2018). The Burden of Parasitic Infections in Immunocompromised Populations. *Infectious Disease Journal*, 27(2), 98-107.

Campbell, J., Croft, T., & Smith, J. (2021). AI-driven diagnostics: leveraging machine learning for improved accuracy in parasitic disease detection. *Global Health Journal*, 15(2), 85-95.

Chalmers, R. M., Katzer, F., & Robertson, L. J. (2018). Advances in Parasitic Diagnostics: From Bench to Field. *Trends in Parasitology*, 34(2), 123-136.

Cimino, R. O., Jeun, R., & Juarez, M. (2020). DNA-Based Diagnostic Tools for Intestinal Parasites in Clinical and Field Settings. *PLoS Neglected Tropical Diseases*, 14(6), e0008318.

Esteva, A., Kuprel, B., & Novoa, R. A. (2017). Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, 542(7639), 115-118.

Fawcett, T. (2006). An Introduction to ROC Analysis. Pattern Recognition Letters, 27(8), 861-874.

Fletcher, S. M., Stark, D., & Ellis, J. T. (2020). Prevalence of Gastrointestinal Parasitic Infections in Developed and Developing Countries: Systematic Review and Meta-analysis. *Journal of Infection*, 81(1), 95-106.

Garcia, L. S. (2007). Diagnostic Medical Parasitology. 5th ed. ASM Press.

Garcia, L. S. (2019). Diagnostic Medical Parasitology. American Society for Microbiology Press.

Hall, A., Hewitt, G., & Tuffrey, V. (2018). A Review and Meta-Analysis of the Impact of Intestinal Worms on Child Growth and Nutrition. *Maternal & Child Nutrition*, 4(s1), 118-236.

Hanley, J. A., & McNeil, B. J. (1982). The Meaning and Use of the Area under a Receiver Operating Characteristic (ROC) Curve. *Radiology*, 143(1), 29-36. doi:10.1148/radiology.143.1.7063747

Haykin, S. (2009). Neural Networks and Learning Machines (3rd ed.). Pearson Education.

Johnson, R., & Adams, P. (2020). Hygiene and Health: The Impact of Sanitation on Parasitic Infections. *Global Health Review*, 12(4), 211-225.

Kermany, D. S., Goldbaum, M., & Cai, W. (2018). Identifying Medical Diagnoses and Treatable Diseases by Image-Based Deep Learning. *Cell*, 172(5), 1122-1131.e9.

Kohavi, R. (1995). A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection. *Proceedings of the 14th International Joint Conference on Artificial Intelligence* (IJCAI), 1137-1143.

Lalonde, R., & Bengio, Y. (2020). Recent advances in deep learning for medical image analysis. *Annual Review of Biomedical Engineering*, 22, 317-342.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep Learning. *Nature*, 521, 436-444. doi:10.1038/nature14539

Lee, S. H., Kim, J. Y., & Chai, J. Y. (2018). Predictive Modeling for Intestinal Parasitic Infections. *Parasites & Vectors*, 11(1), 517.

Li, Z., Wang, Y., & Yu, L. (2020). Deep learning based medical image classification. Journal of Medical Imaging and Health Informatics, 10(4), 1014-1021. https://doi.org/10.1166/jmihi.2020.3060

Ling, C. X., & Sheng, V. S. (2003). Cost-Sensitive Learning and the Class Imbalance Problem. Encyclopedia of Machine Learning and Data Mining. *Springer*.

Litjens, G., Kooi, T., & Bejnordi, B. E. (2017). A survey on deep learning in medical image analysis. *Medical Image Analysis*, 42, 60-88.

Liu, Y., Chen, P. H. C., & Krause, J. (2019). How to read articles that use machine learning: users' guides to the medical literature. *JAMA*, 322(18), 1806-1816.

Nguyen, T. M. P., & Taylor, P. (2021). Barriers to Effective Parasitic Disease Management in Low-Resource Settings. *International Journal of Parasitology*, 51(5-6), 331-343.

Powers, D. M. W. (2011). Evaluation: From Precision, Recall and F-Measure to ROC, Informedness, Markedness and Correlation. *Journal of Machine Learning Technologies*, 2(1), 37-63.

Rajaraman, S., Jaeger, S., & Antani, S. K. (2018). Performance evaluation of deep neural ensembles toward malaria parasite detection in thin-blood smear images. *Journal of Biomedical Informatics*, 85, 101-109. https://doi.org/10.1016/j.jbi.2018.07.018

Rajpurkar, P., Irvin, J., & Ball, R. L. (2018). Deep learning for chest radiograph diagnosis: A retrospective comparison of the CheXNeXt algorithm to practicing radiologists. *PLoS Medicine*, 15(11), e1002686.

Rajpurkar, P., Irvin, J., & Ball, R. L. (2018). Deep learning for chest radiograph diagnosis: A retrospective comparison of the CheXNeXt algorithm to practicing radiologists. *PLoS Medicine*, 15(11), e1002686.

Ravì, D., Wong, C., & Deligianni, F. (2019). Deep learning for health informatics. *IEEE Journal of Biomedical and Health Informatics*, 21(1), 4-21.

Saito, T., & Rehmsmeier, M. (2015). The Precision-Recall Plot Is More Informative than the ROC Plot When Evaluating Binary Classifiers on Imbalanced Datasets. *PLOS ONE*, 10(3), e0118432. doi:10.1371/journal.pone.0118432

Smith, J., Doe, A., & Robinson, M. (2019). Public Health Challenges of Intestinal Parasitic Infections in Developing Regions. *Journal of Tropical Medicine*, 45(3), 123-135.

Smith, P., & Jones, A. (2020). Handling and Transport of Biological Specimens. *Laboratory Techniques in Biochemistry and Molecular Biology*, 47(2), 78-95.

Sokolova, M., & Lapalme, G. (2009). A Systematic Analysis of Performance Measures for Classification Tasks. *Information Processing & Management*, 45(4), 427-437. doi:10.1016/j.ipm.2009.03.002

Topol, E. J. (2019). High-performance medicine: the convergence of human and artificial intelligence. Nature Medicine, 25(1), 44-56. https://doi.org/10.1038/s41591-018-0300-7

Topol, E. J. (2019). High-performance medicine: the convergence of human and artificial intelligence. *Nature Medicine*, 25(1), 44-56.

United Nations International Children's Emergency Fund (UNICEF). (2022). Water, Sanitation and Hygiene. Retrieved from https://www.unicef.org/wash

Verweij, J. J., & Stensvold, C. R. (2019). Molecular Testing for Clinical Diagnosis and Epidemiological Investigations of Intestinal Parasitic Infections. *Clinical Microbiology Reviews*, 27(2), 371-418.

World Health Organization (WHO). (2013). Bench Aids for the Diagnosis of Intestinal Parasites. *Geneva*: WHO Press.

World Health Organization (WHO). (2021). Intestinal Worms. Retrieved from https://www.who.int/news-room/fact-sheets/detail/soil-transmitted-helminth-infections