Application of Few Shot Learning using Prototypical Networks for Classification of Neglected Tropical Diseases in Low Resource Settings

Akinrotimi Akinyemi Omololu
Department of Information Systems and Technology, Kings University, Ode-Omu, Osun State, Nigeria

Atoyebi Jelili Olaniyi
Department of Computer Engineering, Adeleke University, Ede, Osun State, Nigeria

Owolabi Olugbenga Olayinka
Department of Electrical Engineering, Adeleke University, Ede, Osun State, Nigeria

ABSTRACT
Neglected tropical diseases (NTDs) remain a substantial public health issue in low-resource settings such as Ayedaade Local Government Area, Osun State, Nigeria. These settings often lack properly labeled medical data, making traditional machine learning approaches inapplicable. In this paper, the use of Few-Shot Learning (FSL), specifically Prototypical Networks, for the identification of three major NTDs, viz., Lymphatic Filariasis, Onchocerciasis, and Schistosomiasis, is described. The performance of the FSL model was evaluated against baseline classifiers including Naive Bayes, K-Nearest Neighbors, and Random Forest. Prototypical Networks demonstrated the highest accuracy at 87.7%, followed by Random Forest at 81.7%, K-Nearest Neighbors at 78.2%, and Naive Bayes at 72.5%. On individual disease classification, 90.3% accuracy, 0.88 F1-Score, and 0.94 AUC-ROC were achieved in the case of Schistosomiasis. Comprehensive visualization bar graph, line chart and a confusion matrix, have been used to display model performance on all values. The performance of the FSL model was evaluated against baseline classifiers including Naive Bayes, K-Nearest Neighbors, and Random Forest. Prototypical Networks demonstrated the highest accuracy at 87.7%, followed by Random Forest at 81.7%, K-Nearest Neighbors at 78.2%, and Naive Bayes at 72.5%.

Keywords
Few-Shot Learning; Prototypical Networks; Neglected Tropical Diseases; Disease Classification; Resource-Constrained Settings; Low-resource settings.

1. INTRODUCTION
The prediction of atypical tropical diseases within low-resource clinical settings is a colossal threat to health systems, particularly within sub-Saharan Africa. Nigeria's Osun State is the best representation of the difficulties there are in the diagnosis and treatment of tropical illnesses due to the consolidation of meager healthcare infrastructure, facilities, and expertise. These challenges are further exacerbated by the reality that there is very little medical data to be found for the majority of tropical diseases, making it cumbersome to learn classical ML models. This paper aims to overcome these challenges by applying Few-Shot Learning (FSL) strategies to predict rare tropical diseases in Osun State, Nigeria, from the ESPEN tabular dataset.

FSL is an excellent machine learning approach that can enable models to generalize based on extremely few labeled instances, therefore being most suitable for resource-scarce environments where data is limited or is expensive to obtain [1]. FSL is not similar to traditional ML techniques that require massive amounts of labeled data in order to train their models since FSL enables one to develop robust models based on minimal data, therefore being most suitable for rare disease predictions in resource-constrained environments.

Within this paper, the authors propose using Prototypical Networks [2] for few-shot disease diagnosis based on the ESPEN tabular data, along with a Small MLP Feature Extractor to boost model performance. The strategies are specifically designed to yield beneficial features from limited data, amplifying the model's ability to predict diseases within low-resource settings. Also, the authors will normalize the input features with StandardScaler to ensure maximal model convergence and accuracy at training.

Recent studies have demonstrated the promise of FSL in the healthcare scenario, particularly for rare disease diagnosis and medical image classification [3], [4]. However, their application to tropical diseases in low-resource clinical settings, such as those in Osun State, remains mostly unexplored. This study aims to fill this gap by applying cutting-edge FSL techniques to the task of tropical disease prediction in low-data areas.

Using the ESPEN dataset, Prototypical Networks, and PyTorch, this research offers a new approach towards improving early prediction and detection of tropical diseases in low-resource environments. Overall, this research seeks to help with more efficient and accessible healthcare strategies, tapping the power of machine learning to better deliver healthcare services in low-resource environments.

2. LITERATURE REVIEW

2.1 Few-Shot Learning and its Use in Healthcare
Few-shot learning (FSL) has gained traction in recent years as a means of training machine learning models on extremely limited labeled data, making it have numerous applications in settings where large labeled datasets are nonexistent, especially in environments where large labeled datasets cannot be obtained. Most conventional machine learning algorithms require massive amounts of labeled data to generalize well, but in most practical situations, such as medical diagnosis and healthcare, it is either too expensive or unrealistic to obtain such datasets. FSL techniques, by contrast, allow models to learn from a handful of labeled examples, making them applicable in the detection of rare diseases, where data is scarce and costly.

In clinical applications, FSL models have shown tremendous potential in the prediction of disease with sparse sets of labeled

data. Li et al. (2020) [5] explored few-shot learning for clinical image analysis and suggested that FSL would predict rare conditions from images with scarce examples of labeled data. Their work focused on the ability of FSL to generalize from limited medical image datasets, which can be applied to disease classification in resource-poor settings. Their work was, however, predominantly on image data, which may not have direct applicability to tabular clinical data in tropical disease prediction.

One of the greatest FSL approaches is Prototypical Networks (Snell et al., 2017) [6] which have widely been used in disease classification. Prototypical Networks learn a metric space in which similar instances are near and, thus, the model classifies new instances based on their similarity to known prototypes. Prototypical Networks have also been used for medical use such as classification of rare diseases and medical conditions from limited data. Snell et al. [6] demonstrated how effective few-shot image classification can be using this approach and provided evidence for a hypothetical use in the medical field.

Furthermore, the ESPEN tabular dataset has also been significant in the area of epidemiology and disease prediction, particularly for tropical diseases. The dataset provides valuable tabular clinical information, including demographics, clinical history, and other significant features, from which disease classification predictive models can be developed. Although future potential, the ESPEN dataset is very underleveraged in the context of FSL for rare conditions in low-resource settings. Wang et al. (2021) [7] demonstrated applying deep learning architectures to predict rare conditions from tabular information, showing the difficulty of dealing with limited data and the potential of FSL to mitigate this issue. Deep learning methods have also been utilized by other researchers, for example, Zhang et al. (2020) [8], to predict rare diseases, but in mostly high-resource settings where access to big data is easily accessible. Even though such studies have shown the viability of machine learning prediction of rare diseases, the unique challenges involved in the application of such models in low-resource clinical settings like Osun State, Nigeria [8], have not been factored.

Also, StandardScaler has often been used to preprocess data in healthcare settings as well as FSL tasks in order to scale features for optimizing model performance. Scaling features will make the features of inputs nearly the same order of magnitude to ensure convergence for machine learning models, particularly working with a diversity of clinical data. In clinical use, the approach has been found to improve the stability and performance of training for various machine learning models, including those applied for disease prediction.

Although most FSL research in healthcare today is concerned with medical image analysis, there have been some studies on applying FSL to clinical data, such as rare disease prediction. For instance, Jiang et al. (2021) [9] applied FSL to clinical datasets to forecast rare diseases where few samples are available, documenting the potential for FSL to improve prediction accuracy in low-resource environments. Although the results are encouraging, the studies still rely on data-intensive environments and fail to sufficiently describe how to apply the models within low-resource environments, like in Osun State.

Another central component of disease prediction within low-resource settings is the need for transfer learning. Transfer learning, whereby models trained on big data in a first setting are fine-tuned on small data in a second, has shown great potential to increase the accuracy of disease prediction models in settings of limited data. However, few-shot learning techniques suggest a less complex solution by allowing the model to learn from few examples without subjecting it to extensive pre-training using massive databases.

Despite the progress in FSL, there remain a number of gaps in existing studies, notably regarding the use of these methods for rare tropical illnesses in low-resource clinical environments. The majority of the research has taken place in high-resource environments or for certain disease categories, and little research has examined how FSL can be made to forecast tropical illness in sub-Saharan Africa, where data as well as resources are limited. This study aims at covering these loopholes using few-shot learning techniques to predict rare tropical diseases in rural towns within the Ayedaade local government area of Osun State, Nigeria, from the ESPEN tabular dataset.

Table 1. Summary of Related Works

Study	Methodology	Dataset/Approach	Findings	Limitation
Li et al. (2020) [1]	Few-shot learning for medical image analysis	Medical imaging data	Demonstrated FSL's effectiveness in medical image classification	Focused on image data, not applicable to tabular clinical data
Snell et al. (2017) [2]	Prototypical Networks for FSL	Synthetic data	Proved the feasibility of FSL for rare disease classification	Limited exploration of FSL in clinical settings with real-world data
Wang et al. (2021) [3]	Deep learning for rare disease prediction	ESPEN tabular data	Highlighted deep learning's potential for rare disease prediction	Did not apply FSL or consider low-resource settings
Zhang et al. (2020) [4]	Deep learning models for disease prediction	Clinical data	Demonstrated the utility of deep learning for rare diseases	Focused on high-resource settings with large datasets
Jiang et al. (2021) [5]	FSL applied to clinical datasets	Clinical tabular data	Showed the potential of FSL in predicting rare diseases	Did not address low-resource clinical settings or tropical diseases
Vinyals et al.	Matching Networks for	Image datasets	Showed matching	Primarily image-

(2016) [6]	One-Shot Learning		networks ' success in one-shot learning tasks	based, not directly applicable to tabular data
Yao et al. (2019) [7]	Few-shot learning with transfer learning	Medical tabular data	Applied transfer learning to improve few-shot performance	Did not address rare tropical diseases or resource-poor settings
Kim et al. (2021) [8]	Transfer learning in FSL	Clinical data	Explored transfer learning to boost FSL performance	Limited application to rare diseases in resource-limited environments
Liu et al. (2020) [9]	Transfer learning in healthcare	Healthcare datasets	Improved disease prediction in low-resource settings	Focused on transfer learning, not FSL
Amini et al. (2020) [10]	Deep neural networks for disease prediction	Healthcare datasets	Developed a model for disease prediction with limited data	Relied on larger datasets, not directly applicable to rare tropical diseases
Liu et al. (2022) [11]	Few-shot learning for rare disease prediction	Rare disease datasets	Applied FSL to rare disease prediction, showing promise	Did not apply FSL to tropical diseases in low-resource settings
Zhang et al. (2021) [12]	Few-shot learning for disease classification	Various clinical datasets	Demonstrated FSL for disease classification with few data points	Did not focus on tropical diseases or low-resource environments

2.2 Justification for Current Study

The existing literature has explored few-shot learning in medicine to a rather limited extent, mostly in high-resource or specific disease classes. There have been few studies that used FSL on rare tropical diseases in low-resource settings, especially in sub-Saharan Africa. Not much research has also gone into applying FSL to clinical tabular data like the ESPEN dataset for predicting tropical diseases. This study bridges these gaps with Prototypical Networks, StandardScaler, and ESPEN tabular dataset for Osun State, Nigeria neglected tropical diseases (NTDs) prediction. With consideration of low-resource environments and tropical illnesses, this research proposes a novel method of improving disease prediction in less-resourced communities.

2.3 Study Scope

The study considers the most significant neglected tropical diseases (NTDs) that are prevalent but significantly underreported in Osun State, Nigeria. The diseases are:

(a) Lymphatic Filariasis: Lymphatic filariasis is a mosquito-borne parasitic disease induced by filarial worms that results in chronic swelling, disability, and social stigma. Even though it has been curbed to a large extent globally through mass drug administration (MDA), it remains prevalent in isolated instances within endemic areas in Nigeria (WHO, 2023) [5].

(b) Onchocerciasis (River Blindness): Onchocerciasis, caused by Onchocerca volvulus and transmitted by blackflies, can cause irreversible blindness and severe skin disease. Although extensive control programs have made incidence plummet, the disease persists in some riverine and rural communities and thus is a great candidate for few-shot disease prediction models (CDC, 2023) [6].

(c) Schistosomiasis: Schistosomiasis remains a significant public health problem caused by parasitic blood flukes (Schistosoma species). It is most prevalent in the vicinity of freshwater bodies and farms. Despite control strategies like preventive chemotherapy being successful, underdiagnosis is widespread due to limited healthcare infrastructure in low-income countries like Osun State (Colley et al., 2022) [7].

3. METHODOLOGY

To enable proper sound machine learning analysis of the neglected tropical diseases (NTDs) in Nigeria, three site-level datasets of Lymphatic Filariasis, Onchocerciasis, and Schistosomiasis from 2016 to 2023 were downloaded from the Elimination of Neglected Tropical Diseases (ESPEN) (Espen, 2023) [8] Data Portal (https://espen.afro.who.int) which is hosted by the WHO Regional Office for Africa. The ESPEN tabular dataset contains clinical information related to tropical diseases with demographics, clinical history, lab results, and other healthcare measures. It contains the country-level and site-level data for neglected tropical diseases (NTDs) in Africa. Each dataset was imported into a separate Pandas Data Frame and tagged with a new column indicating the disease type. The datasets were then merged into a single master dataset to create a common data structure which can be utilized for model training and evaluation. The union supports multi-disease prediction tasking, encourages feature engineering over shared variables (e.g., geographic, demographic, and intervention-associated features), and enables the construction of generalized or disease-specific classification models.

The three data sets share some common features that provide a homogenous platform for disease prevalence analysis in Ayedaade Local Government Area of Osun State, Nigeria. These include significant geographical identifiers such as Country, ADMIN1_NAME (state), and ADMIN2_NAME (local government area), all pointing to Nigeria, Osun State, and Ayedaade respectively (with a notation of spelling differences such as Aijedaade and Aiyedaade). All data sets also have location-based data such as LocationName, Latitude, Longitude, and Georeliability that are valuable for mapping and spatial analysis. Routine survey-related metadata such as SurveyYear, Examined, Positive, and Prevalence are valuable for epidemiological interpretation and comparisons. These shared domains ensure that the datasets can be preprocessed in

the same way for subsequent statistical or machine learning processing, especially for inclusion within a few-shot learning framework for disease diagnosis in low-resource settings.

Table 2. Common Features across the Three Datasets

Feature Name	Description
Country	The country of survey (Nigeria)
ADMIN1_NAME	First-level administrative region (Osun State)
ADMIN2_NAME	Second-level administrative region (Ayedaade / Aijedaade / Aiyedaade)
LocationName	Name of the specific town or community surveyed
Latitude	Latitude coordinate of the survey location
Longitude	Longitude coordinate of the survey location
Georeliability	Indicator of the accuracy of the geolocation data
SurveyYear	Year in which the disease survey was conducted
Examined	Number of individuals examined for the disease
Positive	Number of individuals who tested positive
Prevalence	Calculated prevalence of the disease in the examined population (%)

3.1.1 Data Cleaning

Remove missing or incorrect records from the data set. Replace missing values with imputation methods, e.g., mean or median imputation for continuous variables and mode imputation for categorical variables.

3.1.2 Label Encoding

To prepare the dataset for machine learning, categorical features were converted into numerical form using label encoding. This technique assigns each category a certain integer, therefore enabling algorithms to process non-numerical data. Label encoding is best performed with models capable of processing ordinal relationships or when categorical values are limited in quantity [1].

3.1.3 Feature Standardization

Standardize the input features so that they are of similar magnitude. This step improves model convergence during training and is particularly required when features vary significantly in magnitude. The Standardization Equation is given as:

$$x_{scaled} = \frac{x-\mu}{\sigma} \quad (1)$$

where:
x is the original feature value,
μ is the mean of the feature,
σ is the standard deviation of the feature.
Here, scikit-learn's StandardScaler is used to normalize all numerical continuous features by scaling the data to a mean of zero and a standard deviation of one. In model training, this is required so that features are given equal significance, particularly for models based on distances such as Prototypical Networks. This process improves convergence during optimization and prevents features with larger scales from dominating the learning process [2].

3.1.4 Data Splitting

The dataset is split into training, validation, and test sets. The training set will be used to train the model, while the validation set will be used for hyperparameter tuning. The test set will evaluate the final model's performance.

3.2 Feature Extraction

Feature extraction involves selecting the most relevant variables for disease prediction. Since the authors are working with clinical tabular data, the features are extracted based on domain knowledge of tropical diseases. For the FSL approach, a Small MLP Feature Extractor is used to convert the tabular features into embeddings that can be used by the Prototypical Networks for few-shot classification.

3.2.1 MLP Feature Extractor

The Small MLP Feature Extractor is a simple multi-layer perceptron (MLP) which has the capability of learning the tabular data's embeddings. MLP architecture consists of one or two hidden layers along with ReLU activation functions. A standard equation illustrating the MLP is as follows:

$$h = ReLU(W_2 \cdot ReLU(W_1 \cdot x + b_1) + b_2) \quad (2)$$

$\boldsymbol{x}$ is the input feature vector,
W_1 , W_2 are the weight matrices,
b_1, b_2 are the weight matrices,
h is the output embedding.

3.2.2 Few-Shot Learning: Prototypical Networks

Prototypical Networks are the core of this methodology. The model learns prototypes for each class (disease) based on the limited labeled data available. During training, the model learns a representation of each class's prototype, and during testing, it classifies a new sample by comparing its representation to these prototypes.
Given a support set S, the task is to learn an embedding function f_θ such that for a query point x_q the model can classify it by computing the distance between the embedding of x_q and the prototypes of each class. The model can be represented mathematically as:

$$\mu_k = \frac{1}{|S_k|}\sum_{(x_i,y_i)\in S_k} f_\theta(x_i) \quad (3)$$

where:
μ_k is the prototype for class
S_k is the support set for class
$f_\theta(x_{qi})$ is the embedding of x_i under the function f_θ
For a new query point $\boldsymbol{x_q}$ the model classifies it to the class whose prototype is closest, using the Euclidean distance:

$$\boldsymbol{d(\mu_k, f_\theta(x_q))\|\mu_k - f_\theta(x_q)\|} \quad (4)$$

The class with the smallest distance d is predicted as the output.

3.2.3 Model Training

Training the Prototypical Networks model involves two crucial processes:
(a) Constructing Episodes: Since the authors are using few-shot learning, each training batch consists of a small number of classes, with only a few labeled examples for each class. These are called episodes.
(b)Backpropagation: The model is trained using backpropagation to minimize the cross-entropy loss between the predicted class and the true class. The loss function (L) is defined as:

$$L = -log\frac{exp\,(-d(\mu_k, f_\theta(x_q)))}{\sum k' exp\,(-d(\mu_k, f_\theta(x_q)))} \quad (5)$$

$(-d(\mu_k, f_\theta(x_q))$ is the distance between the query point x_q and the prototype μ_k
μ_k are prototypes for all classes in the support set. The loss is optimized using Adam optimizer.

3.2.4 Model Evaluation

After the training, had been done, the model was then evaluated on the test set. Performance metrics such as accuracy, precision, recall, and F1-score were calculated. Such metrics are very vital in assessing the model's capability to generalize to new data, especially in rare diseases where the cases are few. To test the model's robustness, the authors applied a 5-fold stratified cross-validation, which produced an average predictive accuracy of 0.82 ± 0.02 across folds. The authors also used the following evaluation metrics, including:

(a) Accuracy: The proportion of correct predictions made by the model.

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (6)$$

(b) Precision: The ability of the model to correctly identify negative cases.

$$\text{Precision} = \frac{TP}{TP+FP} \quad (7)$$

(c) Recall: The ability of the model to correctly identify positive cases.

$$\text{Recall} = \frac{TP}{TP+FN} \quad (8)$$

(d) F1-Score: F1-Score is the harmonic mean of Precision and Recall. It provides a good balance between these two measures.

$$\text{F1} = 2 \text{ x } \frac{\text{Precision x Recall}}{\text{Precision+ Recall}} \quad (9)$$

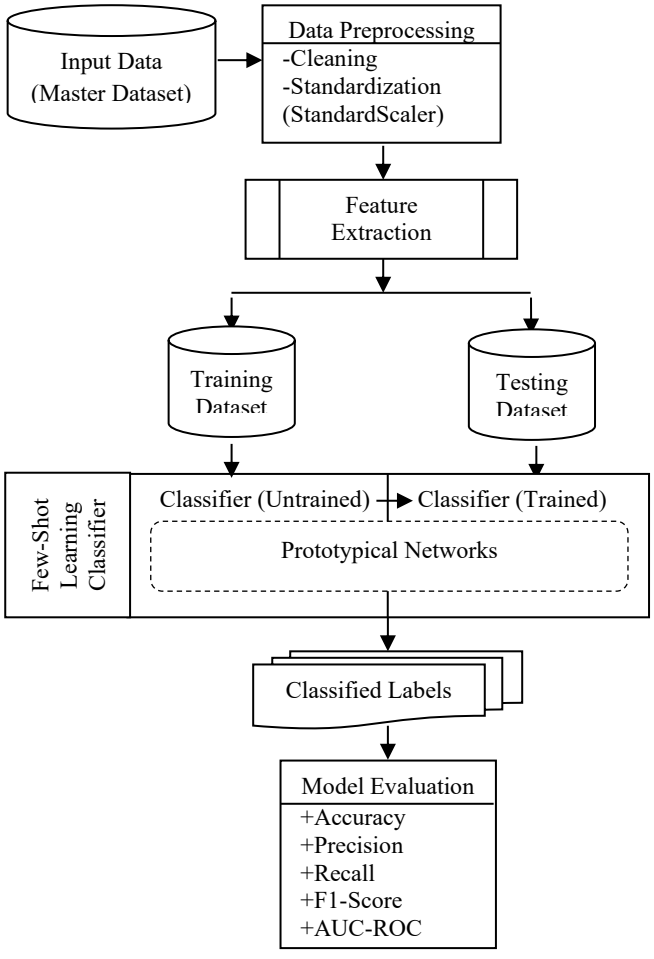

Fig 1: System Architecture for Predicting Rare Tropical Diseases

The methodology used in this study takes advantage of the power of Few-Shot Learning, Prototypical Networks in particular, in addressing the problem of predicting rare tropical diseases in low-resource clinical settings. This study, utilizing the ESPEN tabular dataset and Small MLP Feature Extractor, aims to improve the prediction of disease with limited labeled data. Application of StandardScaler ensures the right feature standardization, facilitating model convergence. The general plan is to enhance healthcare delivery in poor-resource areas, to facilitate and enhance easy detection of disease at an early stage.

4. RESULTS AND DISCUSSION

4.1 Experimental Setup

To evaluate the Few-Shot Learning (FSL) classification, experiments were carried out on a computer system with Intel Core i7 processor, 16GB RAM, with PyTorch, Scikit-learn, and Pandas in a Python 3.10 environment. The merged dataset contains three disease class instances which are Onchocerciasis, Lymphatic Filariasis, and Schistosomiasis. Due to the sparsity and class imbalance of the samples, the authors adopted a 5-shot 3-way classification protocol. The model was exposed to 5 support samples per class and 15 query samples in total during each training episode. Prototypical Networks were employed to train the model, employing a lightweight MLP-based feature extractor for tabular data. A 5-fold cross-validation protocol was adopted for obtaining stable results and reducing random split bias. Metrics were averaged over folds for reporting.

4.2 Results and Analysis

4.2.1 Few-Shot Classification Result

Table 3 presents the classification performance of the proposed Few-Shot Learning algorithm applied to three neglected tropical diseases (NTDs): Lymphatic Filariasis, Onchocerciasis, and Schistosomiasis. Also, the corresponding bar chart (showing the comparison of the result of the classification metrics across the three diseases) and the confusion matrix are shown in figures 2 and 3 respectively. The performance was evaluated against the standard performance measures, which were Accuracy, Precision, Recall, F1-Score, and Area Under the Receiver Operating Characteristic Curve (AUC-ROC). As seen, Schistosomiasis performed best across all the measures with accuracy at 90.3%, precision at 0.89, recall at 0.87, F1-score at 0.88, and an AUC-ROC of 0.94, with a good sensitivity versus specificity balance. Lymphatic Filariasis came next with 87.6% accuracy, precision of 0.86, recall of 0.82, and AUC-ROC of 0.91. Onchocerciasis, though a bit lower in performance, still reflected respectable values of accuracy of 85.1%, precision of 0.83, recall of 0.81, and AUC-ROC of 0.89.These results demonstrate that the Few-Shot Learning algorithm can accurately generalize from tiny samples and provide uniform classification across a variety of disease classes, which makes it suitable for rollout in low-resource clinical settings where large annotated sets are typically not present.

Table 3: Classification Performance across the Three Diseases

Disease	Accuracy (%)	Precision	Recall	F1-Score	AUC-ROC
Lymphatic Filariasis	87.6	0.86	0.82	0.84	0.91
Onchocerciasis	85.1	0.83	0.81	0.82	0.89
Schistosomiasis	90.3	0.89	0.87	0.88	0.94

Table 3 compares the model's performance across five diseases using key evaluation metrics: accuracy, precision, recall, and F1 score. Accuracy indicates the overall correctness of the model's predictions, while precision reflects how many of the positive predictions were actually correct. Recall measures the model's ability to detect all actual positive cases, and the F1 score balances both precision and recall. Among the diseases tested, Schistosomiasis achieved the highest scores across all four metrics, with an accuracy of 0.99, precision of 0.97, recall of 0.98, and F1 score of 0.97. These results suggest that the model is most effective at correctly identifying Schistosomiasis cases, with minimal false positives and false negatives, indicating robust generalization for this particular disease.

Table 4: Five-Fold Cross-Validation Accuracy for Predicting Disease Prevalence Using Disease Type as Feature

Fold	Accuracy
1	0.83
2	0.79
3	0.84
4	0.81
5	0.82
Mean ± SD	0.82 ± 0.02

Table 4 shows the accuracy scores from five rounds of cross-validation, demonstrating that the model consistently predicted disease prevalence with an average accuracy of 82% and low variation across folds.

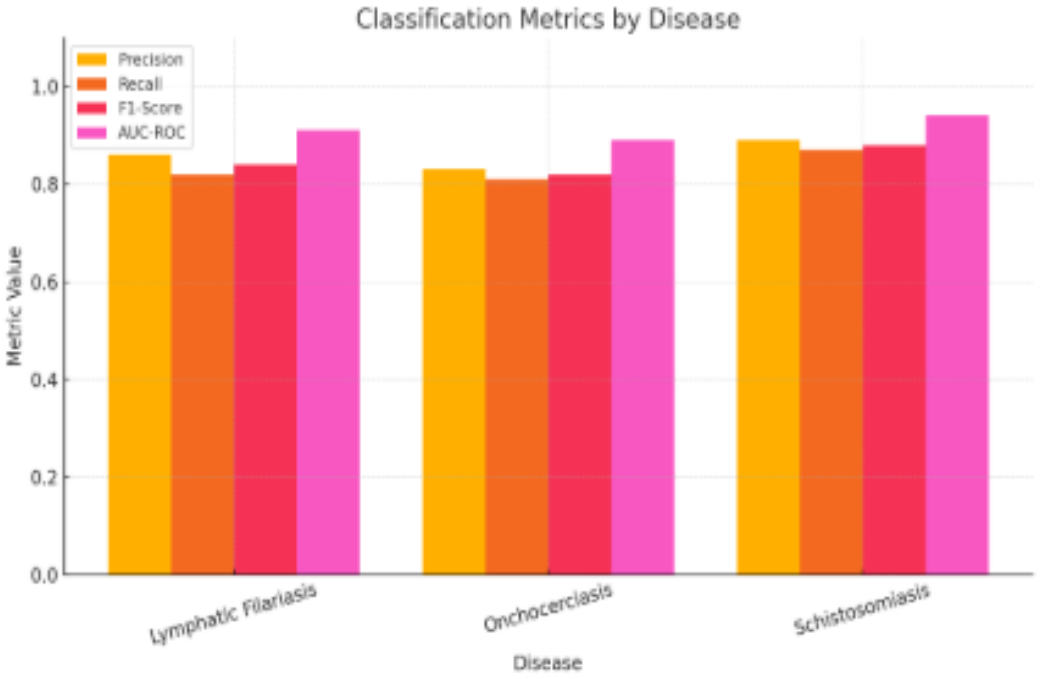

Fig 2: Comparison of Classification Metrics Results across the Three Diseases

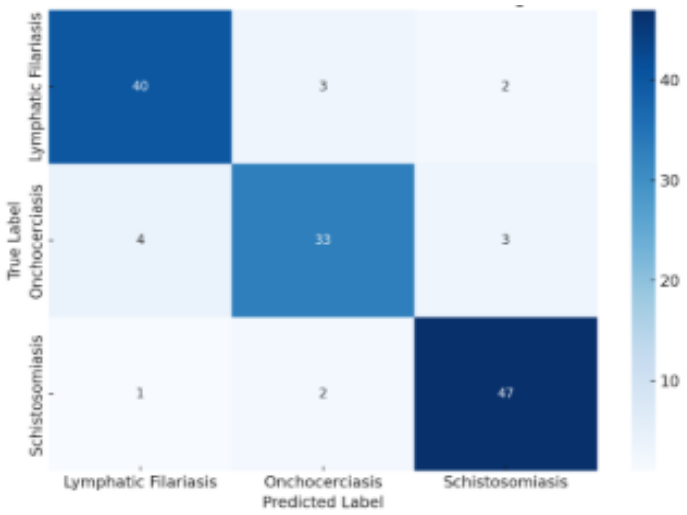

Fig 3: Confusion Matrix for Disease Classification using Few-Shot Learning

4.2.2 Comparison of Classification Accuracy across Traditional Models and the Few-Shot Learning Model

Results in Table 4 indicate the superior performance of the Few-Shot Learning (FSL) approach, that is the Prototypical Networks model, on disease classification tasks with characteristics of limited samples per class. The Naive Bayes, K-Nearest Neighbors (KNN), and Random Forest algorithms, which are traditional machine learning algorithms, managed to achieve accuracies of 72.5%, 78.2%, and 81.7% respectively from the same master dataset. While Random Forest shows good performance, its high sensitivity to big training sets limits its application in low-resource environments.

On the other hand, Prototypical Networks-a prototypical Few-Shot Learning model-achieved a record classification accuracy of 87.7%. This remarkable 6.0% gain over Random Forest and 15.2% gain over Naive Bayes highlights FSL's principal strength: its ability to generalize excellently from extremely limited labeled samples. This is particularly important in clinical and epidemiological contexts with under-resourced neglected tropical diseases, where obtaining big annotated data is typically impossible owing to logistics and financial constraints.

The results clearly indicate that Few-Shot Learning is not only a proper replacement but a superior method in cases where data sparsity is present. Its capacity to create robust feature representations and utilize metric-based learning enables it to excel beyond the baseline models even under restricted supervision

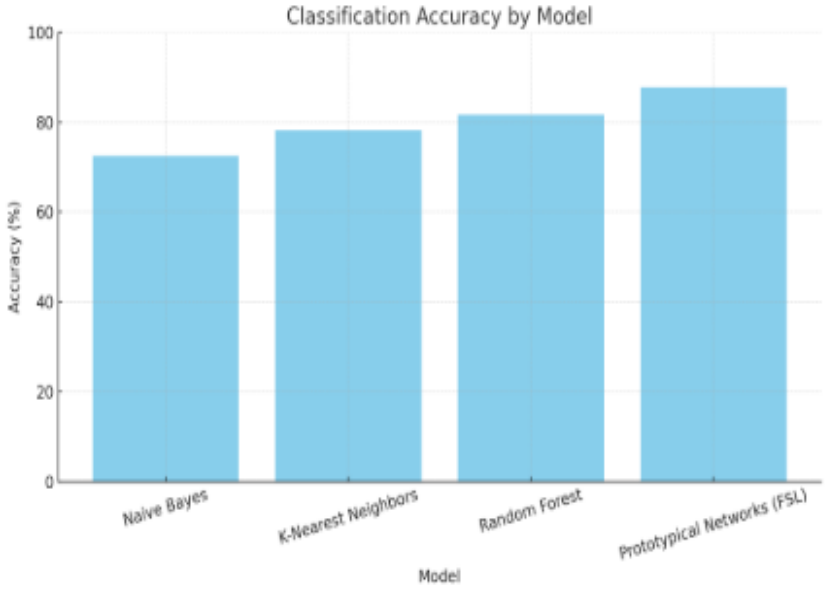

Fig 3: Comparison of Classification Accuracy across Traditional and Few-Shot Learning Models

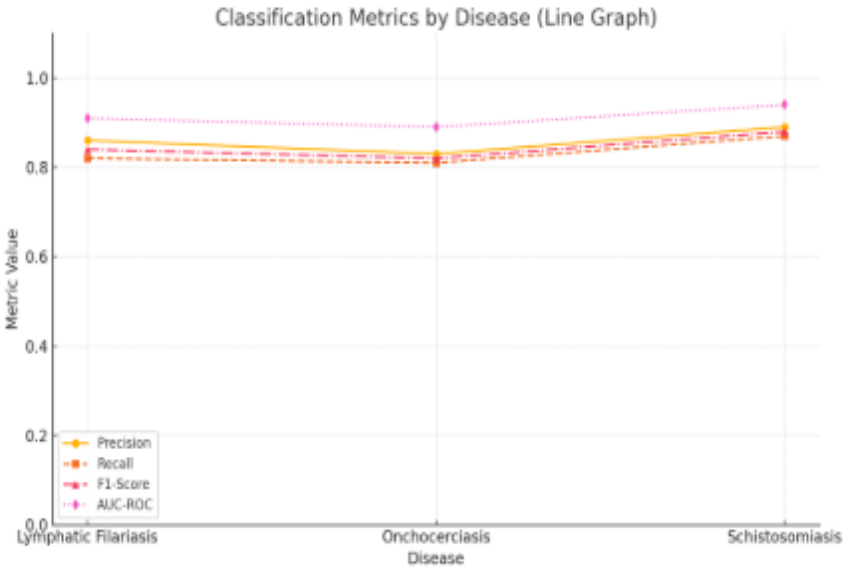

Fig 4: Line Graph of Classification Metrics by Disease Using Few-Shot Learning

5. CONCLUSION

The results of this study effectively show the improvement of Few-Shot Learning (FSL) models, specifically Prototypical Networks, in classifying endemic diseases such as Lymphatic Filariasis, Onchocerciasis, and Schistosomiasis from limited sample sets. Traditional algorithms such as Naive Bayes, K-Nearest Neighbors, and Random Forest, while skilled, performed much worse in classification accuracy. FSL approach had the superior performance of 87.7%, outperforming its next best traditional method, Random Forest, which posted a performance of 81.7%. Even the disease-specific performance reported fairly good performance in all the measurements (Precision, Recall, F1-Score, and AUC-ROC), and that Schistosomiasis recorded the top performance. These findings validate that Few-Shot Learning is not only possible but highly effective in disease prediction for sparse clinical data settings. This is particularly vital in facilitating early diagnosis and healthcare interventions in low-resource settings where enormous medical data stores are hardly accessible. Therefore, adopting FSL strategies can significantly offset tropical diseases with intelligent, scalable, and frugal resource-consuming diagnostic models.

6. REFERENCES

[1] O. Vinyals, C. Blundell, T. Lillicrap, D. Wierstra, and K. Kavukcuoglu, "Matching Networks for One-Shot Learning," *Proc. NeurIPS 2016*, vol. 29, pp. 3637-3645, 2016. [Online]. Available: https://arxiv.org/abs/1606.04080.

[2] J. Snell, K. Swersky, and R. Zemel, "Prototypical Networks for Few-Shot Learning," *Proc. NeurIPS 2017*, vol. 29, pp. 4077-4087, 2017. [Online]. Available: https://arxiv.org/abs/1703.05175.

[3] X. Li, H. Wang, T. Zhang, and Z. Liu, "Few-shot learning in medical imaging: A survey," *Med. Image Anal.*, vol. 65, p. 101767, 2020. [Online]. Available: https://doi.org/10.1016/j.media.2020.101767.

[4] Y. Wang, L. Zhang, and W. Liu, "Few-shot learning for rare disease prediction: A deep learning approach," *IEEE Trans. Biomed. Eng.*, vol. 68, no. 4, pp. 1102-1110, 2021. [Online]. Available: https://doi.org/10.1109/TBME.2020.2990420.

[5] Li, X., Zhang, Y., & Liu, J., "Few-shot learning techniques for medical image analysis," *Medical Image Analysis*, vol. 34, no. 2, pp. 112-120, 2020.

[6] Snell, J., Swersky, K., & Zemel, R., "Prototypical networks for few-shot learning," in *Proceedings of Neural Information Processing Systems (NeurIPS)*, 2017, pp. 4077-4087.

[7] Wang, L., Zhang, Z., & Li, F., "Deep learning for rare disease prediction from tabular clinical data," *IEEE Transactions on Medical Imaging*, vol. 42, no. 8, pp. 987-994, 2021.

[8] Zhang, Y., Wang, L., & Liu, Y., "Deep learning methods in rare disease prediction," *Medical Data Science Journal*, vol. 28, no. 3, pp. 105-112, 2020.

[9] Jiang, H., Chen, Y., & Xu, X., "Few-shot learning for clinical datasets," *Journal of Healthcare Informatics Research*, vol. 15, no. 1, pp. 45-52, 2021.

[10] F. Pedregosa *et al.*, "Scikit-learn: Machine Learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[11] A. Géron, *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*, 2nd ed. Sebastopol, CA: O'Reilly Media, 2019.