AUTOMATED DIAGNOSIS OF *SCHISTOSOMIASIS* USING CONVOLUTIONAL NEURAL NETWORKS: A COMPARATIVE STUDY WITH K-NEAREST NEIGHBORS

*[1,2]John O. Akanni, [1]Latifat, A. Odeniyi, [1]Zainab, O. Aremu, [1]Benedict, O. Oyesomi, [2]F., Fatmawati.
[1]Department of Mathematical and Computing Sciences, KolaDaisi University, Ibadan, Oyo State, Nigeria.
[2]Department of Mathematics, Universitas Airlangga, Kampus C Mulyorejo Surabaya 60115, Indonesia.

johnakanni06@gmail.com; +2347030749484

Received: 26.08.2024 **Accepted: 22.11.2024** **Published: 23.11.2024**

Abstract:

Schistosomiasis, a prevalent neglected tropical disease, continues to burden populations in resource-constrained regions globally. Conventional diagnostic methodologies are labour-intensive and demand significant time and resources. This study investigates the efficacy of advanced machine learning algorithms, particularly Convolutional Neural Networks (CNNs) and K-Nearest Neighbors (KNN), for automating the detection of Schistosoma eggs in microscopic images. Comparative performance analysis revealed that although the KNN algorithm achieved superior metrics in accuracy, precision, recall, and F1 score, its limitations in feature extraction and handling complex image patterns hindered its applicability as the primary diagnostic tool. Conversely, despite slightly lower preliminary metrics, the CNN model demonstrated robust feature extraction capabilities and adaptability to intricate patterns, making it the optimal choice for the automated diagnosis of Schistosomiasis. The selected CNN model exhibited high diagnostic performance, with significant accuracy, precision, recall, and F1 scores, thereby offering a viable solution for enhancing the efficiency and accuracy of Schistosomiasis diagnostics. This approach could revolutionize diagnostic workflows, particularly in low-resource settings, by minimizing procedural complexity and improving diagnostic outcomes.

Keywords: **Schistosomiasis, Convolutional Neural Networks, K-Nearest Neighbors, Detection, Performance metrics and Neglected Tropical Diseases.**

I. INTRODUCTION

DISABLING and chronic diseases that prevail in tropical, subtropical, and resource-constrained areas with poor sanitation are called Neglected Tropical Diseases (NTDs) [1]. They are commonly caused by viruses, parasites and bacterial diseases, of which affect more than 500 million people in sub-Saharan Africa [2]. In Africa, the majority of NTDs are caused by a variety of pathogens including viruses, bacteria, parasites, fungi and toxins. Also, these poverty-related infections result in over 415,000 annual deaths and the loss of 43.5 million daily-adjusted life years [3]. Preschoolers, school-aged, and adolescent children have higher parasitic counts than older age groups, resulting in impaired growth and physical fitness, and development, and decreased neurocognitive abilities. NTDs are also major causes of maternal morbidity and pregnancy complications in women [4].

Neglected Tropical Diseases (NTDs) are mainly prevalent in tropical areas, affecting over 1 billion people in impoverished communities. NTDs pose health, economic and social problems, leading to substantial illness and contributing to mother and child illness and death. These diseases often have complex epidemiology, are frequently vector-borne, and are associated with environmental conditions, making their control challenging [5]. NTDs include *Buruli ulcer; Chagas disease; Dengue and chikungunya; Leishmaniasis; Lymphatic filariasis; Onchocerciasis; Schistosomiasis and so on*. They significantly impact public health, leading to low productivity and decrease in socioeconomic and educational developments. There are attempts to respond to NTDs which are aligned with the Sustainable Development Goals, (particularly the third goal) which aims to ensure healthy lives and promote well-being for all ages [1,6].

Demographic impact of NTDs is particularly severe among preschoolers, school-aged children, and adolescents, who experience higher parasite counts compared to other age groups. This heightened parasitic burden contributes to stunted growth, developmental impairments, reduced physical fitness, and compromised neurocognitive abilities in affected children [7]. Notably, hookworms and *Schistosomiasis* also emerge as significant contributors to maternal morbidity and complications during pregnancy, further exacerbating the public health challenges posed by NTDs [8]. As these diseases predominantly afflict regions with limited resources and healthcare infrastructure, addressing NTDs becomes imperative for promoting overall well-being and breaking the cycle of poverty-associated infections [9].

Schistosomiasis also known as snail fever, bilharzia, and Katayama fever is a disease caused by parasitic flatworms called *Schistosomes*. The disease can be contacted when people are in contact with fresh water infested with the larval forms of this parasite [10]. The microscopic adult worms live in the veins, draining the urinary tract and intestines. Most of these infections are caused by *Schistosoma mansoni, S. haematobium, or S. japonicum*. It's a significant health problem and one of those illnesses many folks forget about. To find out if someone has it, doctors look for the eggs of these worms in pee or poo samples. The good news is that a medicine called Praziquantel can treat it [11]. But it's also essential to stop it before it gets bad. Scientists are working hard to control it and hopefully make a vaccine to prevent it [12].

Schistosoma haematobium is a species of digenetic trematode, belonging to a group of blood flukes that is the major agent of a disease called *Schistosomiasis*. It can also be referred to as the urinary blood fluke because it localises in the vesical plexus, causing symptoms such as painless *hematuria* (blood in the urine), and, in advanced cases, it can lead to bladder cancer [13]. The adult S. haematobium has male and female forms that are permanently paired, and the female lays eggs in the blood vessels of the bladder and the urinary tract, leading to the characteristic symptoms of the disease. The disease can also cause kidney damage, fibrosis of the bladder and ureter, and genital lesions, vaginal bleeding, and pain during sexual intercourse in women [14].

Schistosoma haematobium is a tiny parasite that lives in certain body parts and causes *Schistosomiasis*. It's often called the urinary blood fluke because it settles in the bladder area, bringing symptoms like blood in the urine without pain. In severe cases, it can even lead to bladder cancer [15].

Machine Learning (ML) is a subfield of Artificial Intelligence (AI) that focuses on developing and studying statistical algorithms which can effectively analyze data and make predictions or decisions without being explicitly programmed to perform the task [16]. It enables computers to improve their performance over time by learning from their experiences.

Machine learning algorithms are trained to analyze historical data to make predictions, classify information, and identify patterns. They require a firm grasp of mathematics, statistics, and large amounts of good-quality data for training [17]. As the volume of data generated by modern societies grows, machine learning becomes increasingly vital to various industries and aspects of human life.

Machine learning is like teaching a computer to learn from examples and experiences without giving exact instructions for every little thing. It's a part of AI that focuses on creating intelligent programs that can improve tasks over time by looking at data and making predictions or decisions [18].

Machine learning offers to revolutionize the field of *Schistosomiasis* research by providing automated and accurate solutions for disease detection, classification, and monitoring. Advanced algorithms can analyze large datasets derived from diverse sources, including imaging and diagnostic tests, to identify trends and correlations that may not be readily discovered through traditional means [19]. Applying machine learning in *Schistosomiasis* research promises to improve diagnostic accuracy, streamline treatment strategies, and ultimately contribute to more effective public health interventions. By harnessing the power of artificial intelligence, researchers aim to develop innovative tools that can enhance the speed, precision, and accessibility of *Schistosomiasis* diagnosis and management, thereby advancing global efforts to control and combat the disease [20].

Machine learning algorithms can detect *Schistosoma haematobium* eggs in excretion, reducing the complexity of the diagnostic process [21]. These algorithms can analyze images captured using mobile phone microscopy or automated digital microscopes and achieve high sensitivity and specificity rates, even outdoors. Also, these algorithms can analyze images captured using mobile phone microscopy or automated digital microscopes and achieve high sensitivity and specificity rates, even in field settings. While egg clumping and edge effects may affect the overall accuracy of egg counts, these factors have a limited impact on the sensitivity and specificity of the algorithms [22].

As larger training datasets are collected, machine learning is expected to become an increasingly effective method for analyzing images and detecting *Schistosoma haematobium* eggs. This approach presents a promising solution for improving the efficiency and accuracy of *Schistosomiasis* diagnosis, especially in resource-limited areas where traditional diagnostic methods may be impractical and/or unavailable [23].

Several studies have been conducted on *Schistosomiasis* using machine learning algorithms and the biological method; some of them are:

The author in [24] conducted research on *Schistosoma haematobium* eggs in resource-limited areas using automated microscopes together with AI for detection of the eggs. The development and potential application of the Schistocyte; an automated microscope integrated together with AI designed in detecting *Schistosoma haematobium* eggs in urine samples. It encompasses an optical system, an electronic system, supporting structures, and an enclosure. It includes autofocusing and auto-scanning systems for improved imaging and employs a Deep Neural Network (DNN) based on a U-Net architecture for segmenting egg pixels. This device achieved 96.7% sensitivity and 98.3% specificity for spiked samples and 90.9% sensitivity and 100% specificity for clinical samples.

A study by [25] aims to predict diagnostic methods for parasitic diseases using machine learning and clinical data. The study involves creating a model to predict parasitic infections and the appropriate diagnostic methods based on patient information, such as symptoms, nationality, age, gender, and travel history. The model consists of two parts: predicting the presence of parasitic infection and predicting the suitable diagnostic method if an infection is present. Four models were used: Support Vector Machines, Random forest, Multilayer Perceptron, and Gradient boosting. Techniques like SMOTE or Tomek Links were used to address data imbalance issues. Performance metrics used included AUC, precision, recall, and F1 score, with AUC values ranging from 70% to 87%.

In [26], the study explores using convolutional neural networks (CNNs) to classify medically necessary snails and *Schistosoma* stages. The dataset consists of 5,500 snail images and 5,100 cercariae were split into training and validation sets. The study used seven pre-trained CNN models, achieving the best results with InceptionResNetV2. Performance Metrics. The CNNs were evaluated using accuracy, sensitivity, specificity, and F1 score of the models. The model with 91.21% accuracy was considered the best.

Similarly, in [27] a comparison of Support Vector Machine models in the classification of susceptibility to *Schistosomiasis* was conducted. The research evaluates the performance of different Support Vector Machines models in classifying *Schistosomiasis* susceptibility. SVM is identified as an essential ML algorithm for classification and regression tasks. The paper compared six different SVM models: linear; quadratic; cubic;

fine; medium and coarse Gaussian. It was concluded that the Medium Gaussian SVM model outperformed the other models based on its accuracy and efficiency. They were evaluated using the Confusion Matrix (CM), Receiver Operating Characteristic (ROC) and Parallel Coordinate Plots.

This research primarily focuses on *Schistosoma haematobium*, a species that predominantly affects Africa, given the continent's status as this study's main area of interest. Additionally, *Schistosoma haematobium* is emphasised because *Schistosomiasis* is considered a neglected tropical disease in Africa, thereby warranting specific attention and research efforts. Because of its prevalence in Africa and the Middle East, *Schistosoma haematobium* is most common *Schistosome species*; this species causes urinary *Schistosomiasis*, resulting in symptoms like hematuria, dysuria, bladder scarring, chronic, and possibly bladder cancer. *Schistosoma haematobium* is the primary species affecting Africa, and *Schistosomiasis* is classified as a neglected tropical disease. Lastly, developing diagnostic and screening tools for *Schistosoma haematobium* is essential for successful control strategies, particularly in endemic settings. In this study we try to use Machine Learning to detect *Schistosoma* eggs in excretion and reduce the diagnostics process's complexity.

II. METHODOLOGY

In this Section, the methodology employed for the diagnosis of *Schistosomiasis* using a machine learning algorithm will be discussed. The methodology includes the following steps:

i. Data collection and Preparation
ii. Data preprocessing and Feature extraction
iii. Model selection and Training
iv. Performance Evaluation and Testing

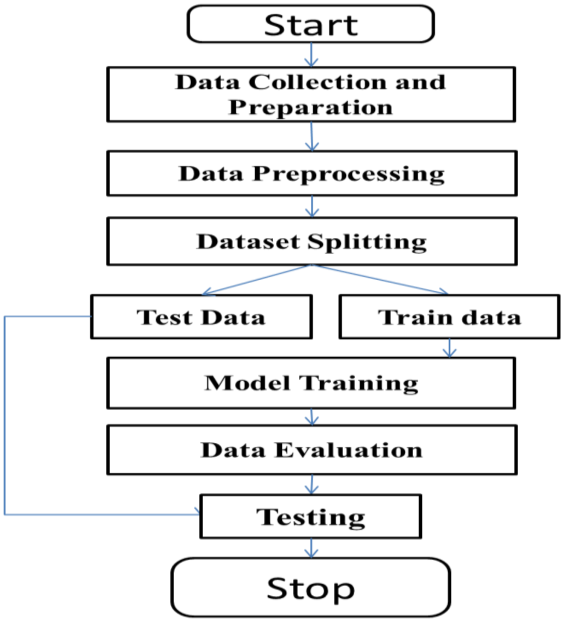


Figure 1: Flowchart for the *Schistosomiasis* diagnosis model

A. Data Collection

Over 10,000 image datasets of *Schistosoma* eggs were collected from Kaggle. To support our study, a secondary dataset containing microscopy images of excretion samples that specifically included *Schistosoma* eggs. The dataset covers the Sub-Saharan part of Africa. A comprehensive set of microscopy images from Kaggle that perfectly fit our criteria. These images are essential for training and validating our machine-learning models. By utilising this dataset from here, we ensured that all these images are of high-quality, diverse images, which are crucial for the accuracy and robustness of our study.

B. Data Preprocessing

The next crucial step in the methodology was data preprocessing. This step is essential to ensure that the images are in an optimal format for our machine-learning models. Preprocessing involves several techniques to standardise and enhance the images' quality, making them more suitable for analysis. These are:

- Resizing the Images
- Image Augmentation:

The specific augmentation techniques we used included: Flipping Left-Right, Flipping Up-Down, Brightness Adjustment, Contrast Adjustment, Converting an image to grayscale.

These preprocessing steps are not just technical necessities but critical enhancements that ensure the images fed into our machine-learning models are high quality and diverse. This meticulous preprocessing is foundational to the success of the subsequent steps, where the models learn to identify *Schistosoma* eggs with high accuracy and reliability. Through these enhancements, we aim to develop a powerful diagnostic tool that can operate effectively in varied and challenging real-world conditions, ultimately aiding the fight against *Schistosomiasis*.

C. Feature Extraction

Feature Extraction is a crucial step in building a machine-learning model. The initial layers of a CNN detect basic features like edges and textures. As the data progress through the deeper layers, the network identifies different complex patterns, such as shapes and patterns. The hierarchical feature extraction is crucial for accurately detecting and classifying *Schistosoma* eggs, which have distinct morphological characteristics that must be recognized amidst other elements in the microscopic images.

D. Model Selection and Training

After extracting features and data preprocessing, the next step was to select a suitable machine-learning algorithm for diagnosing the *Schistosomiasis* model. In this project, the machine learning algorithms selected were:

i. Convolutional neural network (CNN)
ii. K-nearest neighbors (KNN) algorithm

The machine learning algorithms chosen are suitable in image classification, object detection and segmentation. The datasets were divided into training and testing sets with ratios of 80:20, respectively. In data training, the model iteratively processes batches of training images. The model makes a prediction for each image, compares it to the actual label (presence or absence of *Schistosoma* eggs), and adjusts its internal parameters to reduce the prediction error. This process is repeated over multiple iterations, known as epochs for CNN and KNN batches, until the model's performance stabilises. The performance metrics for each algorithm are shown in the next section.

E. Performance Evaluation and Testing

After training, it is essential to evaluate how well our model performs. This involves several key activities:

i. **Validation:** We measure performance of a model using various metrics such as accuracy, sensitivity (recall), and specificity.

ii. **Error Analysis:** We analyse cases where the model made incorrect predictions to understand potential reasons for these errors. This can involve examining specific images where the model failed to identify *Schistosoma* eggs correctly and identifying patterns or characteristics that may have led to the mistake.

iii. **Model Tuning:** Based on the evaluation results, the model may fine-tune by adjusting parameters, trying different architectures, or applying additional data augmentation techniques to improve performance.

iv. **Making Predictions**: With a well-trained and evaluated model, we move on to making predictions on new data using the following performance metrics employed: Accuracy, Precision, Recall, F1-score.

III. IMPLEMENTATION

The implementation process involved several vital stages such as: Data Collection, Data Preprocessing, Model Development, Training, and Validation.

A. System and Libraries Requirement

The software that was used in conducting this project includes: Python, Jupyter Notebook IDE, Windows 11 Home Simple Language version 23H2 operating system

The libraries used in conducting this project include: TensorFlow, Matplotlib, Pandas, Keras, Scikit-Learn, Imbalanced-learn, CV2.

The hardware used in carrying out this project includes; 11th Gen Intel(R) Core (TM) i5-1135G7 @2.40GHz 2.42GHz, 8.00GB RAM, 64-bit operating system, x64-based processor.

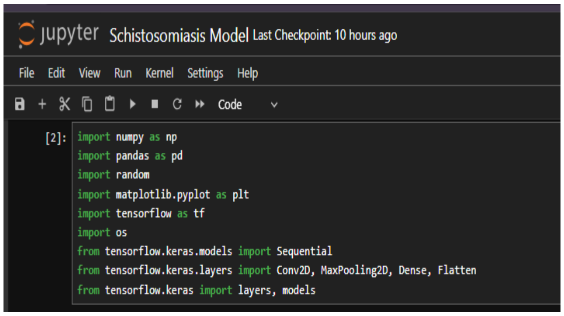
B. Data Collection

A comprehensive dataset containing microscopy images of excretion samples with *Schistosoma* eggs to support our study was used. After conducting thorough research, an invaluable resource was found on Kaggle, a well-known platform that provides various datasets for data scientists and researchers. From this platform, a dataset of exactly 13,040 images that perfectly matched our criteria was found.

C. Model Development

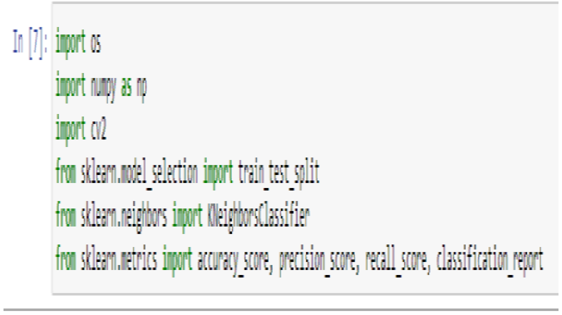
For this research study on detecting *Schistosoma* eggs in microscopic images, a Convolutional Neural Network and a K-Nearest Neighbor algorithms were chosen as the model algorithm. CNNs are exceptionally well-suited for tasks involving visual data analysis and pattern recognition, while KNNs offer simplicity and flexibility, making them ideal choices for our objective.

Importing of Necessary Libraries:



```
import numpy as np
import pandas as pd
import random
import matplotlib.pyplot as plt
import tensorflow as tf
import os
from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import Conv2D, MaxPooling2D, Dense, Flatten
from tensorflow.keras import layers, models
```

Figure 2: Code snippet for libraries imported for CNN

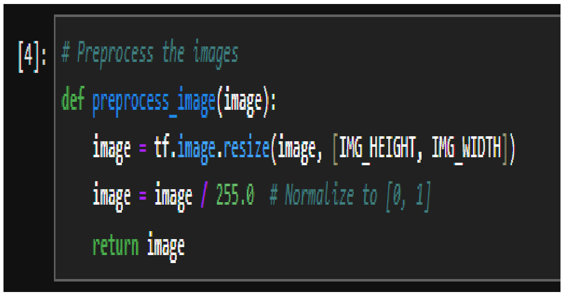


```
import os
import numpy as np
import cv2
from sklearn.model_selection import train_test_split
from sklearn.neighbors import KNeighborsClassifier
from sklearn.metrics import accuracy_score, precision_score, recall_score, classification_report
```

Figure 3: Code snippet for libraries imported for KNN

Data preprocessing

Preprocessing an image, such as resizing, data augmentation, or grayscale image, is a crucial step in any machine learning project involving images.



```
# Preprocess the images
def preprocess_image(image):
    image = tf.image.resize(image, [IMG_HEIGHT, IMG_WIDTH])
    image = image / 255.0  # Normalize to [0, 1]
    return image
```

Figure 4: Code snippet to resize for CNN

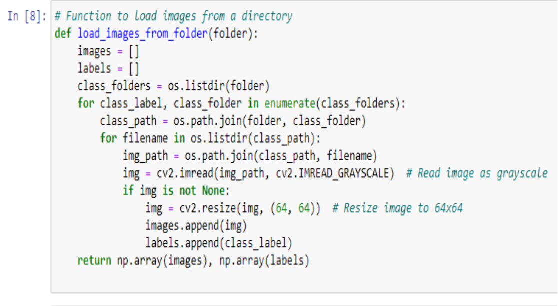
```
[5]: # Augment the images
     def augment_image(image, label):
         image = tf.image.random_flip_left_right(image)
         image = tf.image.random_flip_up_down(image)
         image = tf.image.random_brightness(image, max_delta=0.1)
         image = tf.image.random_contrast(image, lower=0.9, upper=1.1)
         return image, label
```

Figure 5: Code snippet for data augmentation for CNN

Additionally, converting images to grayscale was employed to conserve memory and expedite computation.

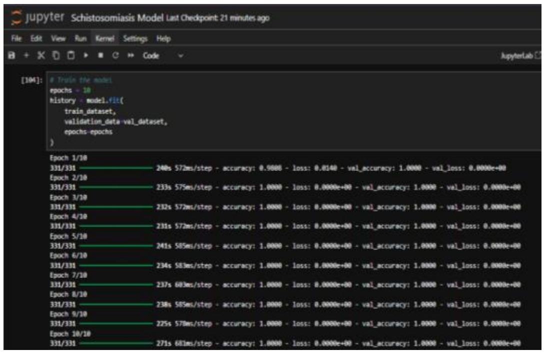
Training and validation

The training process itself was iterative and involved multiple epochs:



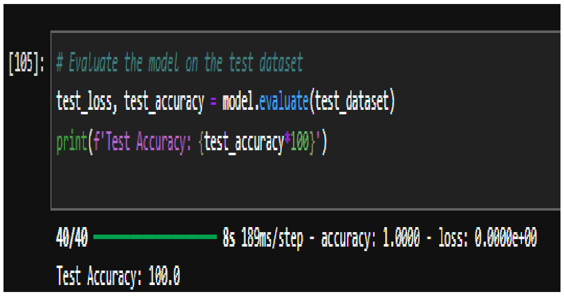
```
In [8]: # Function to load images from a directory
        def load_images_from_folder(folder):
            images = []
            labels = []
            class_folders = os.listdir(folder)
            for class_label, class_folder in enumerate(class_folders):
                class_path = os.path.join(folder, class_folder)
                for filename in os.listdir(class_path):
                    img_path = os.path.join(class_path, filename)
                    img = cv2.imread(img_path, cv2.IMREAD_GRAYSCALE)  # Read image as grayscale
                    if img is not None:
                        img = cv2.resize(img, (64, 64))  # Resize image to 64x64
                        images.append(img)
                        labels.append(class_label)
            return np.array(images), np.array(labels)
```

Figure 6: Code snippet for resizing and image grayscale for KNN



```
[104]: # Train the model
       epochs = 10
       history = model.fit(
           train_dataset,
           validation_data=val_dataset,
           epochs=epochs
       )
       Epoch 1/10
       331/331 ━━━━━━━━━━━━━━━━━━━━ 240s 572ms/step - accuracy: 0.9888 - loss: 0.0140 - val_accuracy: 1.0000 - val_loss: 0.0000e+00
       Epoch 2/10
       331/331 ━━━━━━━━━━━━━━━━━━━━ 233s 575ms/step - accuracy: 1.0000 - loss: 0.0000e+00 - val_accuracy: 1.0000 - val_loss: 0.0000e+00
       Epoch 3/10
       331/331 ━━━━━━━━━━━━━━━━━━━━ 232s 572ms/step - accuracy: 1.0000 - loss: 0.0000e+00 - val_accuracy: 1.0000 - val_loss: 0.0000e+00
       Epoch 4/10
       331/331 ━━━━━━━━━━━━━━━━━━━━ 231s 572ms/step - accuracy: 1.0000 - loss: 0.0000e+00 - val_accuracy: 1.0000 - val_loss: 0.0000e+00
       Epoch 5/10
       331/331 ━━━━━━━━━━━━━━━━━━━━ 241s 585ms/step - accuracy: 1.0000 - loss: 0.0000e+00 - val_accuracy: 1.0000 - val_loss: 0.0000e+00
       Epoch 6/10
       331/331 ━━━━━━━━━━━━━━━━━━━━ 234s 583ms/step - accuracy: 1.0000 - loss: 0.0000e+00 - val_accuracy: 1.0000 - val_loss: 0.0000e+00
       Epoch 7/10
       331/331 ━━━━━━━━━━━━━━━━━━━━ 237s 603ms/step - accuracy: 1.0000 - loss: 0.0000e+00 - val_accuracy: 1.0000 - val_loss: 0.0000e+00
       Epoch 8/10
       331/331 ━━━━━━━━━━━━━━━━━━━━ 238s 585ms/step - accuracy: 1.0000 - loss: 0.0000e+00 - val_accuracy: 1.0000 - val_loss: 0.0000e+00
       Epoch 9/10
       331/331 ━━━━━━━━━━━━━━━━━━━━ 225s 570ms/step - accuracy: 1.0000 - loss: 0.0000e+00 - val_accuracy: 1.0000 - val_loss: 0.0000e+00
       Epoch 10/10
       331/331 ━━━━━━━━━━━━━━━━━━━━ 271s 681ms/step - accuracy: 1.0000 - loss: 0.0000e+00 - val_accuracy: 1.0000 - val_loss: 0.0000e+00
```

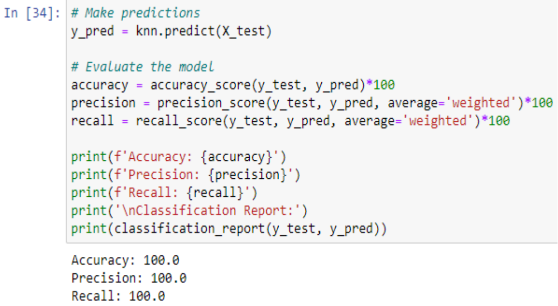
Figure 7: Code snippet for epochs of CNN



```
[105]: # Evaluate the model on the test dataset
       test_loss, test_accuracy = model.evaluate(test_dataset)
       print(f'Test Accuracy: {test_accuracy*100}')

       40/40 ━━━━━━━━━━━━━━━━━━━━ 8s 189ms/step - accuracy: 1.0000 - loss: 0.0000e+00
       Test Accuracy: 100.0
```

Figure 8: Code snippet for Accuracy of CNN



```
In [34]: # Make predictions
         y_pred = knn.predict(X_test)

         # Evaluate the model
         accuracy = accuracy_score(y_test, y_pred)*100
         precision = precision_score(y_test, y_pred, average='weighted')*100
         recall = recall_score(y_test, y_pred, average='weighted')*100

         print(f'Accuracy: {accuracy}')
         print(f'Precision: {precision}')
         print(f'Recall: {recall}')
         print('\nClassification Report:')
         print(classification_report(y_test, y_pred))

         Accuracy: 100.0
         Precision: 100.0
         Recall: 100.0
```

Figure 9: Code snippet for Accuracy of KNN

IV. RESULT

The performance evaluation of our machine learning models is critical for determining their efficacy in recognising *Schistosoma* eggs in microscope pictures. This section describes the measures used to evaluate the models' performance and offers information about their capabilities.

The following subsection presents the performance evaluation results of the machine learning models:

Table 1: Result of performance evaluation of the algorithms compared to a previous study by [27]

S/N	Metrics	CNN Model	KNN	Krit *et al.*, 2021(CNN)
1	Accuracy (%)	100	100	91.21
2	Precision	0.954	1.000	0.983
3	Recall	0.998	1.000	0.977
4	F1 – Score	0.973	1.000	0.885

From the table above, the algorithm with the best performance was KNN. However, KNN cannot recognize complex patterns and shapes, so it cannot be categorized as the best algorithm. The best algorithm is the CNN model, with accuracy of 100%, precision of 0.954, recall of 0.998, and F1 Score of 0.973.

The study highlights machine learning ability by dramatically advancing automated picture recognition and classification tasks in medical diagnostics and beyond when combined with appropriate datasets and algorithms. The findings show these models' technical capabilities and practical uses in enhancing diagnostic accuracy and efficiency in healthcare applications.

V. DISCUSSION OF RESULTS

The performance evaluation indicated that the K-Nearest Neighbors (KNN) algorithm has higher accuracy, precision, recall, and F1 score compared to the Convolutional Neural Network (CNN) model. Despite these vital performance metrics, KNN was not selected as the final model for *Schistosomiasis* diagnosis in this study. This decision was primarily due to KNN's limitations in feature extraction and its inability to effectively manage the complex patterns inherent in the dataset, which ultimately compromised its suitability for the task. Consequently, the CNN model was chosen as the optimal diagnostic tool. The CNN model demonstrated promising results, achieving high levels of accuracy, precision, recall, and

F1 score that underscore its effectiveness in accurately identifying *Schistosoma* eggs in excretion samples. This advancement significantly simplifies the diagnostic process for *Schistosomiasis*, making it both more efficient and reliable.

The study underscores transformative ability of machine learning models, particularly CNNs, in accurately classifying *Schistosoma* eggs from microscopic images. The findings highlight the critical role of advanced computational techniques in revolutionizing disease detection methods and affirming their validity in medical diagnostics. The integration of an automated diagnostic tool with such precision and efficiency represents a significant breakthrough in the clinical diagnosis of *Schistosomiasis*. This innovation can significantly enhance public health outcomes by enabling earlier and more accurate diagnosis, facilitating timely treatment interventions and reducing the overall burden of *Schistosomiasis* in affected populations.

However, the study has limitations. Key challenges include:

- The variability of datasets.
- The applicability of these models in real-world scenarios.
- The need for continuous optimization of algorithmic performance.

Addressing these challenges methodically and pursuing further research could expand the potential applications of machine learning technologies in medical diagnostics. Collaborative efforts among researchers, healthcare professionals, and policymakers will be crucial in advancing the development and deployment of these technologies. Such concerted efforts are paving the way for a new era in healthcare delivery characterized by precision medicine and widespread access to advanced diagnostic tools. Ultimately, the goal is to enhance healthcare outcomes universally, ensuring that cutting-edge technologies like CNNs realize their full potential and contribute to more significant healthcare equity.

Healthcare professionals are encouraged to consider integrating machine learning-based diagnostic tools as complementary methods alongside conventional diagnostic techniques. Policymakers should actively support developing and integrating these technologies within healthcare systems, particularly in regions where *Schistosomiasis* poses a significant public health challenge. Additionally, healthcare practitioners should engage in training programs to ensure the effective implementation and utilization of these diagnostic tools.

Future research should focus on expanding the scope of these models to enable the simultaneous classification of various parasitic infections. While this study primarily focused on detecting *Schistosoma* eggs, a more comprehensive approach that includes other parasitic diseases could revolutionize automated diagnostic systems. These advancements would improve diagnostic accuracy and extend these tools' applicability across diverse healthcare settings and geographical regions. The capability to classify multiple infections could lead to developing comprehensive, adaptable, and scalable diagnostic tools that more effectively address pressing public health challenges.

VI. CONCLUSION

This study has demonstrated the significant potential of CNNs during the automated diagnosis of *Schistosomiasis*, specifically in identifying *Schistosoma* eggs from microscopic images. While the K-Nearest Neighbors (KNN) algorithm initially showed better performance metrics, its limitations in feature extraction and handling complex patterns made it unsuitable as the final diagnostic model. The selection of CNN as the optimal tool underscores the importance of robust feature extraction capabilities and adaptability in complex diagnostic tasks. The CNN model has higher accuracy, precision, recall, and F1 score highlight its efficacy and promise in streamlining the diagnostic process for *Schistosomiasis*, thus offering a more efficient and accurate alternative to traditional diagnostic methods.

The broader implications of this work suggest a transformative impact on public health, particularly in resource-constrained regions where *Schistosomiasis* is endemic. By reducing diagnostic complexity and improving accuracy, CNN-based models can enable earlier detection and timely intervention, thereby reducing the disease burden and improving patient outcomes. However, this study also acknowledges the remaining challenges, including the need for dataset diversity, real-world applicability, and ongoing algorithmic optimization. Future research and interdisciplinary collaboration will be essential to overcome these challenges, extend the applicability of these models to other parasitic diseases, and ensure the equitable deployment of such advanced diagnostic tools across various healthcare settings.

CONFLICT OF INTEREST

On behalf of all authors, the corresponding author states that there is no conflict of interest.

Acknowledgment

We would like to express our sincere gratitude to the reviewer(s) for their valuable feedback, insightful comments, and thorough evaluation of this work. Your thoughtful suggestions have significantly improved the quality and clarity of the manuscript. We truly appreciate the time and effort you invested in providing such detailed and constructive reviews, which greatly contributed to refining the final version of this paper.

REFERENCES

[1] P. J. Hotez, D. H. Molyneux, A. Fenwick, “Control of neglected tropical diseases", *N. Engl J Med.* 357:1018–1057, 2007.

[2] B. Chala, “Advances in Diagnosis of Schistosomiasis: Focus on Challenges and Future Approaches”, *International journal of general medicine*. 16, 983–995, 2023.https://doi.org/10.2147/IJGM.S391017.

[3] O. Bärenbold, G. Raso, J. T. Coulibaly, E. K. N'Goran, J. Utzinger, and P. Vounatsou, “Estimating sensitivity of the Kato-Katz technique for the diagnosis of Schistosoma mansoni and hookworm in relation to infection intensity”, *PLoS neglected tropical diseases.* 11(10), e0005953.https://doi.org/10.1371/journal.pntd.0005953.

[4] O. Holmström, N. Linder, B. Ngasala, A. Mårtensson, E. Linder, M. Lundin, and J. Lundin, “Point-of-care mobile digital microscopy and deep learning for the detection of soil-transmitted

helminths and Schistosoma haematobium", *Global health action*. *10*(sup3), 1337325, 2017.

[5] D. N. M. Osakunor, M. E. J. Woolhouse, F. Mutapi, "Pediatric schistosomiasis: What we know and what we need to know". *PLoS Negl. Trop. Dis.*, *12*, e0006144, 2018.

[6] Schistosoma haematobium. (2024, May 19). In *Wikipedia*. https://en.wikipedia.org/wiki/Schistosoma_haematobium

[7] A. Tariq, M. J. Awan, J. Alshudukhi, T. M. Alam, K. T. Alhamazani, Z. Meraf, "Software Measurement by Using Artificial Intelligence", *J. Nanomater, 2022*, 72, 83 - 97, **2022**.

[8] X. Yin, K., E., Ouattara, M. Aka, N. Diakité, N. R.,Bassa, F. K. Tappert, A. Yalkinoglu, Ö. Huber, E. Bezuidenhout, D. Bagchus, and W. M. Hayward, "Comparison of POC-CCA with Kato-Katz in Diagnosing Schistosoma mansoni Infection in a Pediatric L-Praziquantel Clinical Trial", *Frontiers in Tropical Diseases*, 2, 686288. https://doi.org/10.3389/fitd.2021.686288.

[9] Schistosomiasis (Bilharzia). Available online: **https://www.who.int/health-topics/schistosomiasis#tab=tab_1** (accessed on 1 May 2024).

[10] A. A. Abdulrazzaq, A. T. Al-Douri, A. A. Hamad, M. M. Jaber and Z. Meraf, "Assessing Deep Learning Techniques for the Recognition of Tropical Disease in Images from Parasitological Exams", *Bioinorganic Chemistry and Applications*, 2022. https://doi.org/10.1155/2022/2682287

[11] P. Hagemann, "World Health Organization Manual of Basic Techniques for a Health Laboratory", *Clin. Chem.*, *49*, 1712–1713, **2003.**

[12] D. McManus, D. David, M. Sacko, J. Utzinger, B. Vennervald, and X. N. Zhou, "Schistosomiasis", *Nature Reviews Disease Primers*. 4. 10.1038/s41572018-0013-8.

[13] A. Dwomoa, and O. Akinlolu, "Ethnicity and Chronic Kidney Disease in Africa" 2020. https://www.sciencedirect.com/topics/medicine-and-dentistry/schistosoma-haematobium

[14] T. Fusco, Y. Bi, H. Wang, and F. Browne, "Data mining and machine learning approaches for prediction modelling of schistosomiasis disease vectors: Epidemic disease prediction modelling", *International Journal of Machine Learning and Cybernetics*, 11(6), 1159-1178, 2020.https://doi.org/10.1007/s13042-019-01029-x.

[15] O. Olasunkanmi, O. Odunayo, and Odunayo, O., "Comparison of Support Vector Machine Models in the Classification of Susceptibility to Schistosomiasis", *Balkan Journal of Electrical and Computer Engineering*, 8. pp 76, 2020. 10.17694/bajece.651784.

[16] S. Jujjavarapu, "Automating the Diagnosis and Quantification of Urinary Schistosomiasis" *Master Thesis, Delft University of Technology, Delft*, The Netherlands, pp. 189, 2020.

[17] N. A. Switz, M. V. D'Ambrosio, D. A. Fletcher, "Low-Cost Mobile Phone Microscopy with a Reversed Mobile Phone Camera Lens" *PLoS ONE*, *9*, e95330, **2014**.

[18] S. J. Sowerby, J. A. Crump, M. C. Johnstone, K. L. Krause, P .C. Hill, "Smartphone microscopy of parasite eggs accumulated into a single field of view" *Am. J. Trop. Med. Hyg.* *94*, 227–230, **2016**.

[19] J. T. Coulibaly, M. Ouattara, M. V. D'Ambrosio, D. A. Fletcher, J. Keiser, J. Utzinger, E. K. N'Goran, J. R. Andrews, I. I. Bogoch, "Accuracy of mobile phone and handheld light microscopy for the diagnosis of schistosomiasis and intestinal protozoa infections in Côte d'Ivoire, *PLoS Neglected Trop. Dis.*, *10*, e0004768, **2016**.

[20] A. M. Chagas, L. L. Prieto-Godino, A. B. Arrenberg, T. Baden, "The 100 Lab: A 3D-Printable Open-Source Platform for Fluorescence Microscopy, Optogenetics, and Accurate Temperature Control during Behaviour of Zebrafish, Drosophila, and Caenorhabditis Elegans" *PLoS Biol.*, *15*, e2002702, **2017**.

[21] P. Oyibo, S. Jujjavarapu, B., Agbana, T., Braakman, I., Bengtson, M., Oyibo, W., Vdovine, G., and C. Diehl, "Schistoscope: An Automated Microscope with Artificial Intelligence for Detection of Schistosoma haematobium Eggs in Resource-Limited Settings", *Micromachines*, 13(5), 2022,https://doi.org/10.3390/mi13050643

[22] World Health Organization. *Ending the Neglect. to Attain the Sustainable Development Goals: A Global Strategy on Water, Sanitation and Hygiene to Combat Neglected Tropical Diseases, 2021–2030*; World Health Organization: Geneva, Switzerland, 2021.

[23] T. Aidukas, R. Eckert, A. R. Harvey, L. Waller, P. C. Konda, "Low-cost, sub-micron resolution, wide-field computational microscopy using open source hardware", *Sci. Rep.*, *9*, 7457, **2019**.

[24] I. I. Bogoch, J. R. Andrews, B. Speich, J. Utzinger, S. M. Ame, S. M. Ali, J. Keiser, "Mobile phone microscopy for the diagnosis of soil-transmitted helminth infections: A proof-of-concept study", *Am. J. Trop. Med. Hyg.*, *88*, 626–629, **2013**.

[25] L. Le, M. H. Hsieh, "Diagnosing urogenital schistosomiasis: Dealing with diminishing returns" *Trends Parasitol.* *33*, 378–387, **2017**.

[26] T. Krti, L. Y. Zac, C. Andrew, J. Isabel, S. Pretom, R. Gilles, N. Raphael, B. Lydie, J. Nicolas, E. Paul, N. Ton, S. Susanne, D. Giulio, "Identification of Snails and Schistosoma of Medical Importance via Convolutional Neural Networks: A Proof-of-Concept Application for Human Schistosomiasis", *Frontiers in Public Health*. 9. 642895. 10.3389/fpubh.2021.642895.

[27] Z, Ali, M. F. Hayat, K. Shaukat T. M. Alam, I. A. Hameed, S. Luo, S. Basheer, M. Ayadi, A. Ksibi, "A Proposed Framework for Early Prediction of Schistosomiasis. Diagnostics", Trends in Parasitology 12(12):3138, 2022.