International Journal of Interdisciplinary Innovative Research & Development (IJIIRD)
ISSN: 2456-236X
Vol. **11** Special Issue **02** | **2026**

Integrating Collective Intelligence and Deep Learning for Scalable Diagnosis of Global Health Diseases Using Crowdsourced Medical Images

Prof. Neha Vilas Patil[1]
[1] *Lecturer, Computer Science & Engineering, Padm. Dr. V. B. Kolte College Of Engg., M.H, India*

DOI: 10.5281/zenodo.19254051

ABSTRACT

Microscopy-based diagnosis remains a cornerstone for detecting a wide range of global health diseases, particularly in low- and middle-income countries where advanced diagnostic infrastructure is limited. Despite its diagnostic reliability, manual microscopic examination is highly dependent on trained experts, making it labor-intensive, time-consuming, and difficult to scale. Artificial intelligence has emerged as a powerful alternative to automate such diagnostic workflows; however, the development of accurate deep learning models requires large volumes of annotated medical images. Expert-driven annotation is costly and often unavailable at scale, creating a critical bottleneck for AI deployment in healthcare. This study proposes an integrated framework that combines collective intelligence and deep learning to address this challenge. Crowdsourced annotations were obtained using a gamified digital platform that enabled non-expert participants, including school-age children and adults, to contribute to the labeling of microscopy images. These annotations were aggregated through majority voting and used to train convolutional neural networks based on a MobileNetV2 architecture employing transfer learning. Comprehensive experiments demonstrate that models trained on crowdsourced data achieve diagnostic performance comparable to models trained on expert annotations, with macro-average AUC values exceeding 0.92. Incremental training experiments further highlight the positive correlation between dataset size and model accuracy. The proposed framework illustrates how human collective intelligence can effectively complement artificial intelligence, reducing annotation costs while maintaining clinical relevance. This approach is generalizable to other medical imaging domains and supports the development of scalable, accessible diagnostic solutions for global health challenges.

Keywords:- Collective Intelligence, Deep Learning, Medical Image Analysis, Crowdsourcing, Microscopy Diagnosis, Global Health

1. INTRODUCTION

The achievement of universal health coverage depends heavily on the availability of accurate and timely diagnostic services. However, a significant proportion of the global population continues to lack access to essential diagnostic resources, particularly in regions affected by neglected tropical diseases. Many parasitic and infectious diseases rely on visual inspection of biological samples using microscopy, which remains the gold standard diagnostic technique. Although effective, microscopy-based diagnosis is constrained by the need for trained personnel, consistent laboratory conditions, and substantial time investment. These limitations hinder large-scale disease surveillance and timely treatment. Recent advances in artificial intelligence have demonstrated strong potential in automating image-based diagnostic tasks. Deep learning models, particularly convolutional neural networks, have achieved remarkable success in medical image analysis. Nevertheless, the performance of such models is fundamentally dependent on the availability of large, high-quality annotated datasets. Generating expert annotations at scale is both expensive and time-consuming, motivating alternative strategies such as crowdsourcing. Crowdsourcing leverages the collective intelligence of non-expert contributors and has shown promise in medical image annotation. This work explores the integration of crowdsourced annotations with deep learning to develop scalable diagnostic systems for global health applications.

2. THEORETICAL BACKGROUND

Collective intelligence refers to the shared or group intelligence that emerges from the collaboration and competition of many individuals. In the context of medical image annotation, collective intelligence enables large-scale labeling efforts by distributing tasks among numerous non-expert participants. Prior research has

demonstrated that aggregated annotations obtained through majority voting or probabilistic fusion can approximate expert-level accuracy. Gamification further enhances participation by increasing engagement and motivation, making it particularly suitable for data-intensive annotation tasks.

Deep learning has revolutionized medical image analysis by enabling automated feature extraction and classification. Convolutional neural networks are particularly well-suited for visual recognition tasks due to their hierarchical feature-learning capabilities. Transfer learning allows models pretrained on large natural image datasets to be adapted to medical imaging tasks with limited data, reducing training time and improving generalization. However, noisy labels introduced through crowdsourcing present additional challenges that must be addressed through robust loss functions and regularization strategies.

3. METHODOLOGY

The proposed methodology consists of three primary components: crowdsourced data collection, annotation aggregation, and deep learning-based classification. Microscopy images were digitized and segmented into fixed-size patches, which were then introduced into a gamified annotation platform. Participants classified each image patch into predefined categories. Annotations from multiple contributors were aggregated using a majority voting mechanism to generate training labels. A MobileNetV2-based convolutional neural network was trained using transfer learning, data augmentation, and noise-robust loss functions to mitigate the impact of incorrect labels. Early stopping and validation-based tuning were employed to prevent overfitting.

3.1 Crowdsourced Medical Image Data Collection

Microscopy images used in this study were obtained from digitized stool samples acquired using standardized optical microscopy settings. To facilitate computational processing, each high-resolution image was segmented into fixed-size, non-overlapping patches. This patch-based strategy allows localized feature learning, reduces computational complexity, and increases the effective size of the training dataset.

The extracted image patches were incorporated into a custom-designed, gamified annotation platform. Gamification was intentionally employed to improve user engagement, motivation, and sustained participation, particularly among non-expert contributors. Participants, including both school-age children and adults, were presented with individual image patches and asked to classify them into predefined categories representing different helminth species or non-infected samples. Clear visual examples and intuitive interface elements were provided to ensure accessibility and reduce cognitive load for non-specialist users.

This approach enables large-scale data annotation without reliance on continuous expert involvement, thereby significantly reducing annotation time and cost while promoting inclusivity and scalability.

3.2 Annotation Aggregation and Label Generation

Given the inherent variability and potential noise in crowdsourced annotations, individual labels were not directly used as ground truth. Instead, annotations collected from multiple contributors per image patch were aggregated using a majority voting mechanism. This strategy assumes that while individual annotations may contain errors, the collective consensus converges toward the correct label as the number of contributors increases.

Majority voting provides a simple yet effective method for noise reduction and has been shown in prior studies to approximate expert-level annotation quality when sufficient quorum sizes are used. The aggregated labels obtained through this process were treated as pseudo-ground truth and used for training the deep learning models. This aggregation step is critical in ensuring robustness and reliability when leveraging non-expert annotations.

3.3 Deep Learning Architecture and Training Strategy

For automated image classification, a convolutional neural network based on the MobileNetV2 architecture was employed. MobileNetV2 was selected due to its lightweight design, reduced parameter count, and suitability for deployment in resource-constrained environments, which aligns with the needs of global health applications.

To overcome the limitations imposed by relatively small training datasets, transfer learning was utilized. The network was initialized with weights pre-trained on a large-scale natural image dataset, allowing the model to leverage generic low-level visual features such as edges, textures, and shapes. Higher-level layers were then fine-tuned using the crowdsourced medical image dataset to adapt the network to domain-specific features associated with helminth morphology.

To enhance generalization and mitigate overfitting, extensive data augmentation techniques were applied during training. These included random rotations, flips, scaling, and translation operations, which simulate variations commonly encountered in real-world microscopy imaging conditions.

Furthermore, to address label noise introduced through crowdsourcing, a noise-robust loss function was employed during optimization. This approach reduces the influence of incorrectly labeled samples by dynamically adjusting the contribution of each training instance to the loss computation. Model training was

monitored using a validation dataset, and an early stopping strategy was applied to terminate training when performance ceased to improve, thereby preventing overfitting and ensuring stable convergence.

4. EXPERIMENTAL RESULTS AND DISCUSSION

Experimental evaluation focused on annotation quality, model performance, and the influence of training data size. Crowdsourced annotations exhibited high agreement with expert labels when sufficient quorum sizes were used. Models trained on annotations from both children and adults achieved performance comparable to expert-trained models, demonstrating robustness to label noise. Incremental training experiments revealed substantial performance gains as the number of training samples increased. Visualization techniques confirmed that the models learned clinically relevant features corresponding to parasitic structures within the images.

4.1 Evaluation Metrics

To objectively assess annotation quality and model performance, standard evaluation metrics commonly used in medical image analysis were employed. Annotation quality was measured using classification accuracy, defined as the ratio of correctly labeled samples to the total number of samples. This metric enables direct comparison between crowdsourced labels and expert-generated annotations.

For evaluating the deep learning models, the Area Under the Receiver Operating Characteristic Curve (AUC) was selected as the primary performance indicator. AUC provides a threshold-independent measure of a model's ability to distinguish between different classes and is particularly suitable for imbalanced datasets. To ensure fairness across all categories, macro-averaged AUC was computed by averaging class-wise AUC values.

To reduce the effect of randomness during model training, each experiment was repeated multiple times, and the mean performance and standard deviation were reported. This approach provides a more robust estimation of model stability and generalization capability.

4.2 Quality of Crowdsourced Annotations

The reliability of crowdsourced annotations was first examined by comparing aggregated labels with expert annotations. A majority voting mechanism was applied using varying quorum sizes to determine the optimal number of contributors required for reliable label generation.

The results indicate that annotation accuracy improves consistently as the number of contributors increases. Annotations aggregated from larger groups showed higher agreement with expert labels, confirming the effectiveness of collective intelligence in reducing individual annotation errors. While adult participants generally produced higher annotation accuracy than school-age children, annotations obtained from children still achieved high reliability when sufficient quorum sizes were used.

These findings demonstrate that non-expert participants, including younger contributors, can generate medically meaningful annotations when their inputs are properly aggregated. This has significant implications for scaling annotation efforts in data-intensive medical imaging tasks.

4.3 Deep Learning Model Performance

Deep learning models trained using crowdsourced annotations were evaluated on an independent test dataset annotated by experts. Models trained on data provided by adults, school-age children, and experts were compared to assess the impact of annotation source on classification performance.

Across all experiments, models trained on crowdsourced data achieved performance comparable to expert-trained models. Macro-average AUC values exceeded 0.92 for all annotation groups, demonstrating that the proposed framework is robust to moderate levels of label noise. Although models trained using adult annotations showed marginally higher performance, the difference between annotator groups was minimal, highlighting the noise-tolerant nature of the learning approach.

These results confirm that deep learning models can effectively learn discriminative features from crowdsourced labels, provided that appropriate training strategies and loss functions are employed.

4.4 Impact of Training Sample Size

To examine the relationship between dataset size and model performace, incremental training experiments were conducted using progressively larger subsets of the training data. The results reveal a strong positive correlation between the number of training samples and classification performance.

Models trained on smaller datasets exhibited reduced accuracy and higher variance, whereas performance improved steadily as additional annotated samples were incorporated. This trend emphasizes the importance of continuous data acquisition and supports the feasibility of iterative model refinement as more crowdsourced annotations become available.

The observed performance gains highlight the scalability of the proposed framework and its suitability for real-world deployment, where datasets can be expanded incrementally over time.

4.5 Model Interpretability and Visual Analysis

To enhance interpretability and clinical relevance, visualization techniques were applied to examine the regions of microscopy images that contributed most strongly to model predictions. Activation maps revealed that the model consistently focused on diagnostically relevant structures, such as helminth eggs, while exhibiting minimal activation in healthy samples.

These visualizations provide qualitative evidence that the model learns meaningful biological features rather than relying on spurious correlations. Such interpretability is critical for building trust in AI-assisted diagnostic systems, particularly in medical and public health contexts.

4.6 Discussion

The experimental findings validate the central hypothesis of this study: collective intelligence can effectively complement artificial intelligence in medical image analysis. By leveraging crowdsourced annotations and robust deep learning techniques, the proposed approach addresses one of the most significant bottlenecks in AI-driven healthcare—data annotation scalability.

The minimal performance gap between models trained on crowdsourced and expert annotations underscores the practicality of this approach for global health applications. Furthermore, the ability to continuously improve model performance through incremental data collection aligns well with real-world diagnostic workflows.

Overall, the results demonstrate that integrating human participation with machine intelligence offers a powerful and sustainable pathway for developing accessible, scalable, and cost-effective diagnostic solutions.

5. CONCLUSIONS

This study demonstrates the practical feasibility and effectiveness of integrating collective intelligence with deep learning to enable scalable and accurate medical image–based diagnosis. The experimental results confirm that crowdsourced annotations, when systematically aggregated using robust consensus mechanisms, can serve as a reliable alternative to expert-generated labels for training deep learning models. Despite the inherent variability associated with non-expert annotations, the proposed framework successfully mitigates label noise and achieves diagnostic performance comparable to models trained exclusively on expert data.

By leveraging a gamified crowdsourcing platform, this work addresses one of the most critical challenges in medical artificial intelligence—the scarcity and high cost of annotated training data. The findings indicate that large-scale participation from non-expert contributors can significantly reduce the annotation burden on medical professionals while maintaining high levels of classification accuracy. Moreover, the robustness of the deep learning models to annotation noise highlights the suitability of combining collective intelligence with modern learning strategies such as transfer learning, data augmentation, and noise-tolerant loss functions.

An important contribution of this study is the demonstration of continuous and incremental model improvement as additional annotated data becomes available. This characteristic supports real-world deployment scenarios in which datasets grow over time and diagnostic systems must adapt accordingly. The ability to refine model performance through iterative training further enhances the scalability and sustainability of the proposed approach.

Overall, the presented framework offers a cost-effective, scalable, and inclusive solution for developing AI-assisted diagnostic systems in global health contexts. Beyond soil-transmitted helminth diagnosis, the methodology is broadly applicable to other medical imaging domains that rely on visual inspection, including parasitology, pathology, and point-of-care diagnostics. Future research may focus on expanding the framework to additional diseases, optimizing annotation aggregation strategies, and integrating real-time feedback mechanisms to further strengthen human–AI collaboration in healthcare.

6. REFERENCES

[1] United Nations, *Transforming Our World: The 2030 Agenda for Sustainable Development*, Department of Economic and Social Affairs, United Nations, 2015.

[2] U. J. Muehlematter, R. Daniore, and K. Vokinger, “Approval of artificial intelligence and machine learning–based medical devices in the USA and Europe (2015–2020): A comparative analysis,” *The Lancet Digital Health*, vol. 3, no. 3, pp. e195–e203, 2021.

[3] F. Yang, M. Poostchi, H. Yu, Z. Zhou, K. Silamut, J. Yu, and S. Antani, “Deep learning for smartphone-based malaria parasite detection in thick blood smears,” *IEEE Journal of Biomedical and Health Informatics*, vol. 24, no. 5, pp. 1427–1438, 2020.

[4] A. Vijayalakshmi and B. Rajesh Kanna, “Deep learning approach to detect malaria from microscopic images,” *Multimedia Tools and Applications*, vol. 79, no. 21–22, pp. 15297–15317, 2020.

[5] O. Holmström, A. Linder, and M. Lundin, “Point-of-care mobile digital microscopy and deep learning for the detection of soil-transmitted helminths,” *Global Health Action*, vol. 10, no. 3, 2017.

[6] A. Yang, A. Bakhtari, J. Langdon-Embry, et al., "KankaNet: An artificial neural network–based smartphone application for soil-transmitted helminth detection," *PLoS Neglected Tropical Diseases*, vol. 13, no. 8, e0007577, 2019.
[7] B. A. Mathison, K. Pritt, and B. J. Couturier, "Detection of intestinal protozoa in trichrome-stained stool specimens using deep convolutional neural networks," *Journal of Clinical Microbiology*, vol. 58, no. 6, pp. 1–13, 2020.
[8] M. A. Luengo-Oroz, A. Arranz, and J. Frean, "Crowdsourcing malaria parasite quantification: An online game for analyzing images of infected thick blood smears," *Journal of Medical Internet Research*, vol. 14, no. 6, 2012.
[9] M. Linares, D. Capellán-Martín, R. Tomé, and M. Luengo-Oroz, "Collaborative intelligence and gamification for online malaria species differentiation," *Malaria Journal*, vol. 18, no. 1, pp. 1–10, 2019.
[10] A. Keshavan, J. Yeatman, and A. Rokem, "Combining citizen science and deep learning to amplify expertise in neuroimaging," *Frontiers in Neuroinformatics*, vol. 13, article 29, 2019.
[11] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L. Chen, "MobileNetV2: Inverted residuals and linear bottlenecks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018, pp. 4510–4520.
[12] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *International Conference on Learning Representations (ICLR)*, 2015.
[13] R. R. Selvaraju, M. Cogswell, A. Das, et al., "Grad-CAM: Visual explanations from deep networks via gradient-based localization," *International Journal of Computer Vision*, vol. 128, no. 2, pp. 336–359, 2020.
[14] S. E. Reed, H. Lee, D. Anguelov, et al., "Training deep neural networks on noisy labels with bootstrapping," in *International Conference on Learning Representations (ICLR)*, 2015.
[15] O. Russakovsky, J. Deng, H. Su, et al., "ImageNet large scale visual recognition challenge," *International Journal of Computer Vision*, vol. 115, no. 3, pp. 211–252, 2015.
[16] World Health Organization, *Ending the Neglect to Attain the Sustainable Development Goals: A Roadmap for Neglected Tropical Diseases 2021–2030*, Geneva, Switzerland, 2020.
[17] A. Lacoste, A. Luccioni, V. Schmidt, and T. Dandres, "Quantifying the carbon emissions of machine learning," *arXiv preprint arXiv:1910.09700*, 2019.