e-ISSN: 3027-0650
Vol. 2, Issue 1, pp246 -253, November 6 -9, 2024

A Semantic Ontology-Driven Explainable Classifier for Identifying *Plasmodium* Species and Stages in Thin Smear Images

Kamal Bakari Jillahi[1], Gabriel Terna Ayem[2], Ijandir Isaac Samuel[3]

{kamal.bakari@aun.edu.ng, gabriel.ayem@aun.edu.ng, isaac.samuel@aun.edu.ng}

Department of Computer Science, American University of Nigeria, Yola[1], Department of Data Science, American University of Nigeria, Yola[2], Department of Information Systems, American University of Nigeria, Yola[3]

Abstract
This study presents an explainable classifier for identifying *Plasmodium* species and life stages from thin smear images by integrating **Convolutional Neural Networks (CNNs)** with **Pathogen Ontology**. Using the **CDC Thin Smear dataset**, the model applies **SegNet** for pixel-wise semantic segmentation, identifying features like infected red blood cells, ring forms, and gametocytes. Ontological reasoning maps visual features to structured biological concepts, producing interpretable outputs (e.g., "Parasitized: hasCleft AND hasDots"). This approach enhances diagnostic transparency, enabling clinicians to understand the AI's decisions. Additionally, **Grad-CAM** visualizations support explainability by highlighting relevant image regions, fostering trust in the system. The combination of deep learning and ontology ensures real-time, reliable malaria diagnostics, reducing human error while maintaining clinical relevance.

Keywords: Semantic segmentation, Ontology-based reasoning, Malaria diagnosis automation, XAI, Interpretable Classifiers.

1 Introduction

The detection of *Plasmodium falciparum* (Pf), the parasite responsible for severe cases of malaria, is a critical task in healthcare, particularly in malaria-endemic regions [1]. Traditionally, identification of *Pf* is performed through manual microscopic examination of blood smear images, which is labor-intensive, time-consuming, and subject to human error [2]. Automation of this process has gained momentum, with advances in machine learning, particularly Convolutional Neural Networks (CNNs), driving research in this domain [2] [3] [5]. A recent trend within this context is the use of ontology-based frameworks combined with CNNs, which aim to enhance the explainability and transparency of model decisions, crucial for healthcare applications where interpretability is essential [2].

Furthermore, integrating explainability into AI-driven malaria diagnosis is essential to build trust, ensure transparency, and enhance clinical decision-making [6]. While AI models can achieve high accuracy in detecting and classifying *Plasmodium* parasites, understanding the reasoning behind their predictions is crucial, especially in medical contexts where lives are at stake [6][7]. Explainable AI provides insights into which features, such as specific morphological or color patterns, influenced the diagnosis, allowing clinicians to verify the system's output [2][8]. This interpretability helps ensure that AI models are not making decisions based on irrelevant data or hidden biases, reducing the risk of misdiagnosis. Additionally, by offering clear, interpretable explanations, AI systems can facilitate collaboration between healthcare providers and AI tools, enabling clinicians to trust and confidently act on AI-generated insights [2].

To this end, current challenges faced by explainability techniques in AI-driven malaria diagnosis stem from the complexity of medical data and the intricacies of AI models [9]. Many deep learning models, particularly those used for image analysis, operate as "black boxes," making it difficult to interpret how they arrive at their decisions [8]. This lack of transparency becomes critical when diagnosing diseases like malaria, where a wrong diagnosis can have serious consequences[9]. Another challenge is ensuring that these explanations are relevant and reliable across diverse patient populations and environments, particularly in low-resource settings where training data might be limited[9][10]. Balancing model complexity with interpretability, while avoiding oversimplifications that could compromise diagnostic accuracy, remains a significant hurdle in integrating explainability into AI-driven healthcare solutions. This work is an effort towards providing models which can provide reliable and human understandable explanations as basis for their decisions.

2 Review of Related Literature

The identification of malaria parasites from thin blood smear images has traditionally relied on manual microscopic analysis. However, due to the inherent variability in image quality and the expertise required, several automated systems have been proposed [2]. For instance, [9] explored deep learning models to classify different stages of malaria parasites, using CNNs to achieve high accuracy. Their work demonstrated the utility of CNNs in handling complex biological images, enabling automatic detection of parasites and their life cycle stages. Similarly, [10] leveraged a CNN-based model for *Pf* identification, achieving high accuracy with deep feature extraction. These methods, though effective, often function as "black-box" models, providing little insight into the decision-making process.

CNNs have shown remarkable success in medical image classification due to their ability to automatically learn hierarchical features from input images [11]. In particular, CNN architectures such as VGG, ResNet, and Inception have been widely applied in medical diagnosis tasks, including malaria detection from blood smears [6]. One of the key benefits of CNNs is their capacity to reduce the reliance on handcrafted features, a common challenge in traditional image analysis methods. Instead, CNNs automatically learn features at different levels of abstraction, allowing them to capture intricate patterns and structures within biological images [12].

Despite their efficacy, the use of CNNs in medical diagnostics faces a significant challenge in terms of interpretability. Medical practitioners demand transparent models that explain why certain decisions are made, which is particularly important when such decisions influence critical healthcare outcomes [13]. In this regard, ontology-based frameworks are becoming an important tool for augmenting CNN models with interpretability[14].

Ontology-based systems provide a formal representation of knowledge in a specific domain, making them ideal for tasks requiring structured data and reasoning [14]. In medical diagnostics, ontologies are used to organize and represent complex biological knowledge, enhancing both the accuracy and transparency of decision-making systems. According to [15], ontologies in healthcare serve as a structured knowledge base, enabling the explanation of system outputs by linking decision-making processes to domain-specific knowledge. When applied to CNN models for malaria detection, ontology-based systems help enhance explainability by mapping learned features to known biological entities and processes. For example, [15] integrated an ontology-based framework with CNNs to explain model predictions by associating visual features with established clinical knowledge. Their study highlights that ontology-based systems can mitigate the black-box nature of CNNs, offering more transparent and interpretable models. Additionally, ontologies provide a mechanism for validating model predictions against domain knowledge, further enhancing trust in the automated system.

The combination of ontology and CNNs addresses this gap. By embedding domain knowledge in the form of ontologies, models can provide not only visual explanations but also logical reasoning aligned with the biological characteristics of *Plasmodium falciparum*. For instance, [16] proposed a hybrid ontology-based CNN system for malaria detection, wherein ontologies provided the underlying reasoning that complemented the model's visual output, leading to a more transparent and reliable system.

While ontology-based CNN models show promise in improving the identification of *Plasmodium falciparum*, several challenges remain. One key challenge is the development of comprehensive ontologies that accurately represent the complex biological structures and life cycles of malaria parasites [15]. Another challenge is the integration of these ontologies with deep learning models without compromising model performance. Furthermore, there is a need for robust evaluation frameworks to assess the interpretability and explainability of ontology-based CNN models, especially in high-stakes medical scenarios [17][18].

3 Methodology

Using the Kaggle Malaria dataset, which contains labeled images of red blood cells infected with *Plasmodium falciparum* and healthy cells of around 27,558 images across two classes: parasitized and uninfected, a Convolutional Neural Network (CNN) with an input layer for 64x64x3 images, followed by 3 convolutional layers with 32, 64, and 128 filters (each using 3x3 kernels), ReLU activation function, and 2x2 max-pooling after each convolution, two fully connected layers (e.g., 256 and 128 neurons), followed by a softmax output layer for binary classification (infected vs. healthy cells), with dropout (e.g., 0.5 rate) and batch normalization to prevent overfitting and stabilize learning is trained on this dataset. To ensure proper training, we pre-processed the images by normalizing pixel values and augmenting the dataset to improve generalization.

Next, we incorporate the Pathogen Ontology, a structured framework of biological knowledge that categorizes pathogens, including *Plasmodium falciparum*. This ontology helps us align our model with relevant biological concepts, integrating domain knowledge into the AI model's learning process. By associating model outputs with the corresponding pathogen class in the ontology, we can map the prediction of the CNN to the biological meaning, aiding in interpretability.
To enhance explainability further, we employ Grad-CAM (Gradient-weighted Class Activation Mapping), a technique that visualizes the areas of input images that the CNN uses to make its predictions. By applying Grad-CAM to the trained classifier, we can produce heatmaps over the images, highlighting the regions of red blood cells where the model "looks" to identify *Plasmodium falciparum* infection[22]. These visualizations allow medical professionals to verify whether the model focuses on biologically relevant regions, such as certain parts of infected cells, making the system more transparent and trustworthy in real-world diagnostic scenarios.

4 Proposed Method

a) *Define* $C = \{ c_1, c_2, \ldots . c_n \}$, *the classes from the ontology which serve as the output of the classification.*
b) *Define* $D = \{d \mid \exists\, c \in C,\ d \equiv c$ *is a valid axiom of the ontology} i.e. definitions of concepts in the ontology*
   Let P: be a set of concepts (features) in the ontology involved in D.
   Let R: be a set of relationships such as $R = \{r \mid r = \text{relationship}(p),\ p \in P\}$
   Let F: be a set of ontological features $f \in F$, *that is the triplets (c, p, r) in D which will be used to explaining the classification.*
c) *Build a set* $FI \subseteq F$ *of ontological features that match the features of the data point input to the classifier such that* $FI = \{fi \in F \mid fi \equiv \exists p.r\}$
d) *Build an ontological reasoning from D and FI such that* $CI \subseteq C$, *for the classified data point* c_i.
e) *apply* $DI \subseteq D$ *such that* $DI = \{di \equiv ci\}$ *and FI to generate an explanation for CI.*

It is important to note that step (c) refines step (b), as constructing **F** with ontological features that cannot be extracted from the classified data would be ineffective. Additionally, the abstraction level of the explanations is inherently connected to the abstraction level of the ontological features used. Indeed, the refinement of features ensures that only **ontological elements that can be reliably extracted from thin smear images** are included in the model's explanations. It would be inefficient to incorporate features that the classifier cannot detect, such as cellular structures beyond what the microscope captures. The **abstraction level of explanations corresponds to the granularity of the ontological features**—for example, the model can explain that a region contains **P. falciparum** because it detects a **ring form or gametocyte**, but it cannot further explain the molecular composition of these forms. This reflects a broader principle in explainable AI: **at some level of abstraction, explanations are deemed sufficient** without needing deeper detail. Similar to other models combining CNNs and ontologies, the malaria classifier connects image-based outputs (e.g., identifying parasites) with **semantic labels** from the pathogen ontology, ensuring explanations are understandable and clinically useful without attempting to define underlying biological processes [20].

5 Explanation Pipeline

This section outlines the implementation of our approach, which consists of two primary modules. The first is a **semantic segmentation (DL) module**, responsible for extracting ontological features from the input image. The second is an **ontological reasoning module**, referred to as **OntoClassifier**, which computes the set of classes (**CI**) that can be inferred from the identified features (**FI**) while generating corresponding explanations [19]. The design of these modules is detailed as follows:

Semantic Segmentation

The **semantic segmentation process** plays a crucial role in identifying and classifying specific regions within microscopic blood smear images. The objective is to analyze thin smear images to detect **Plasmodium species** (e.g., *P. falciparum*, *P. vivax*) and distinguish between infected and uninfected red blood cells (RBCs). Each pixel in the image is labeled to correspond to a meaningful biological class, such as parasite types, healthy cells, or background. To achieve this, the segmentation process uses the **SegNet architecture**, which is well-suited for pixel-wise labeling, ensuring that every part of

the image is accounted for. This segmentation is not merely visual but also integrates semantic meaning through **pathogen ontology**, making the results explainable to clinicians.

The **SegNet architecture** is employed as the core model for the semantic segmentation task. It consists of an **encoder-decoder network** that reduces the input image's dimensions through max-pooling operations (in the encoder) to extract key features and then upsamples them (in the decoder) to reconstruct the segmentation mask. This structure ensures that important spatial information is retained while achieving efficient pixel-level classification [19]. The **CDC Thin Smear dataset** is used to train the SegNet model, with annotated images containing examples of both **infected and uninfected RBCs**. The dataset provides detailed masks indicating the presence of parasites, specific *Plasmodium* species, and background regions, which are essential for training the model to distinguish subtle visual features. By correctly segmenting the microscopic regions, the model provides granular insights into the health status of each RBC in the sample [21].

The integration of **Pathogen Ontology** ensures that the segmentation results carry **semantic meaning** and are aligned with established biomedical knowledge. Once the SegNet model segments the regions corresponding to different parasite species or uninfected cells, these segmented areas are **mapped to ontological terms** that represent the biological identity of each structure (e.g., NCBITaxon:5833 for *P. falciparum*). This mapping ensures that the predictions are not only accurate but also meaningful within a clinical and scientific context . The use of pathogen ontology provides **explainability**, making the model's output interpretable to clinicians by linking segmented areas to standardized biological identifiers. This semantic segmentation process supports **diagnostic transparency** by clarifying which regions contributed to the classification, thereby enhancing trust and enabling more informed decision-making in malaria diagnosis. This is presented in figure 1

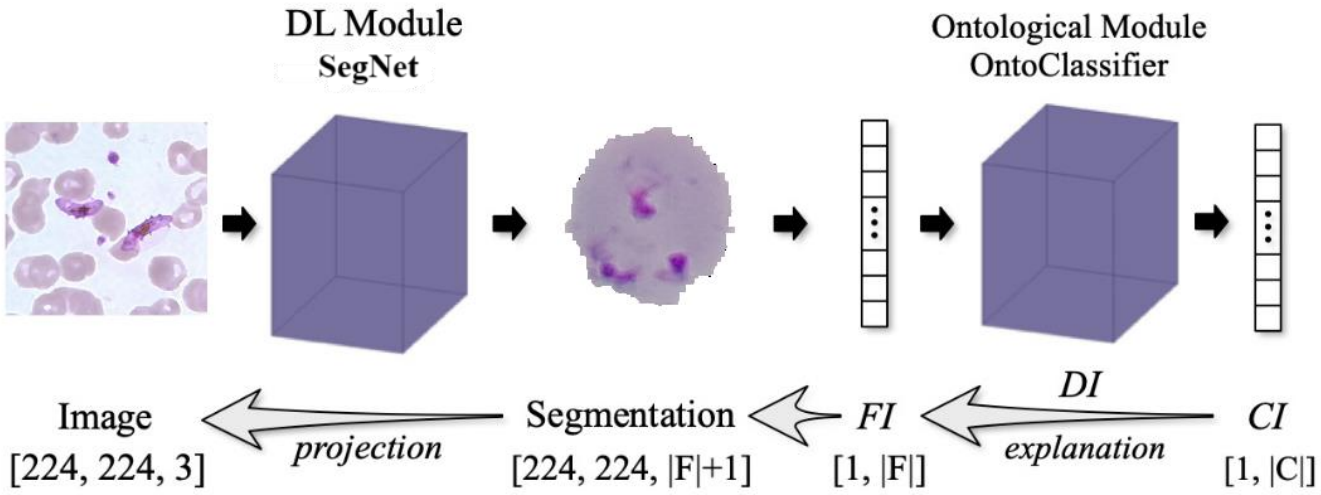

Figure 1: Architecture of the Ontologically explainable classifier [19].

OntoClassifier

The presence of pixels in a segmentation layer signifies the existence of a corresponding ontological feature. That is why the semantic segmentation process utilizes a Convolutional Neural Network (CNN) to analyze thin smear images, identifying ontological features such as **ring forms, schizonts, and gametocytes** as individual pixels in the segmentation mask. Each detected pixel is associated with an ontological assertion (e.g., **(∃ hasFeature . parasite)**), allowing the classifier to compile a set of satisfied assertions (**FI**) for each image. The model then employs a reasoning mechanism using a defined set of relationships from the **Pathogen Ontology** to deduce the corresponding classes (**CI**) that accurately label the image, such as identifying **P. falciparum** based on the observed features. While traditional methods for ontological reasoning can be slow

and resource-intensive, this classifier effectively integrates deep learning capabilities with ontological mapping to ensure efficient, real-time classifications while maintaining explainability in identifying specific malaria parasites.

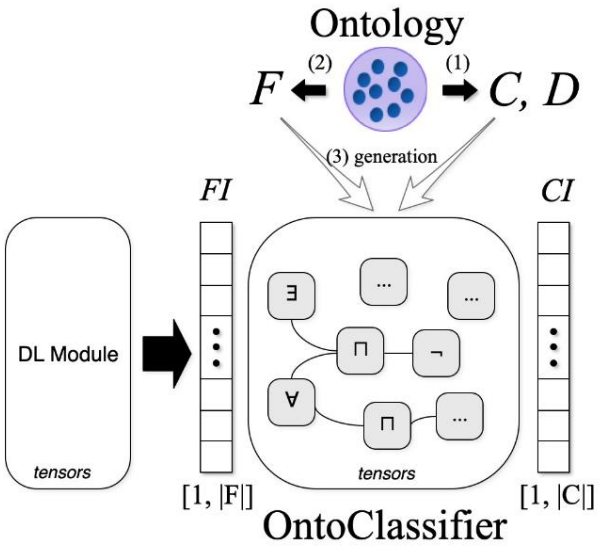

Figure 3: Structure for Generating an OntoClassifier [19]

6 Result

The result presented here are three instances from CDC dataset as outlined above. Two of the test cases are parasitized while one is non-parasitized. The explanations generated are based on the structured of the pathogen ontology and the result of semantic segmentation.

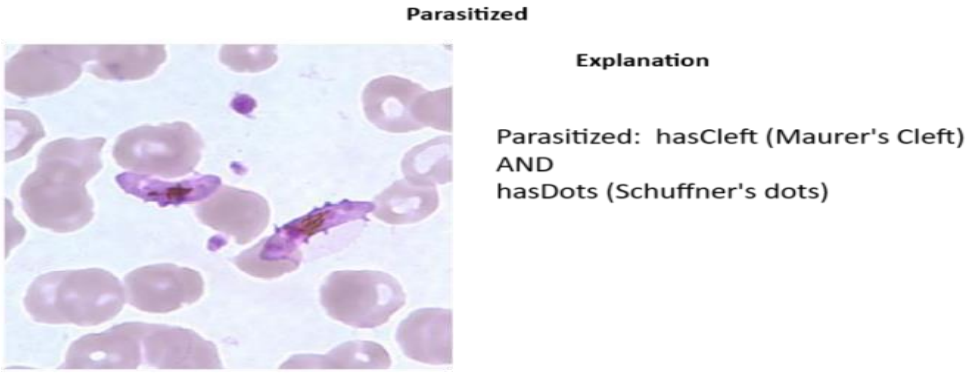

Figure 4: Example of Parasitized Input Image and the corresponding Output Explanation

From figure 3, the explanation "Parasitized: hasCleft (Maurer's Cleft) AND hasDots (Schuffner's Dots)" signifies that the ontology-based explainable model has analyzed the thin smear image and identified specific morphological features

indicative of Pf infection. During the segmentation process, the model utilizes deep learning techniques to accurately detect regions within the image where Maurer's clefts, which are associated with the developmental stage of the parasite, and Schuffner's dots, which appear in infected red blood cells, are present. By referencing a defined ontology that links these features to their corresponding biological terms, the model classifies the image as "Parasitized." The assertion "hasCleft AND hasDots" indicates that both features are observed, suggesting a more detailed characterization of the infection and allowing clinicians to understand the specific pathological changes associated with the malaria parasite in the examined sample. This reasoning integrates visual analysis with semantic knowledge, providing a meaningful and clinically relevant output.

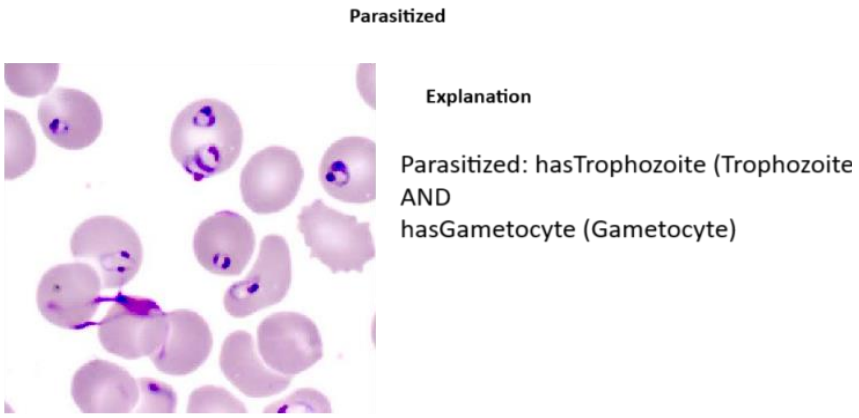

Figure 5: Another example of a Parasitized Image Input and a Corresponding Explanation

From Figure 4 the explanation "Parasitized: hasTrophozoite (Trophozoite) AND hasGametocyte (Gametocyte)" indicates that the ontology-based explainable model has identified specific features within the thin smear image corresponding to two distinct stages of the Plasmodium species lifecycle. The model first segments the image into various regions using a semantic segmentation approach, detecting individual pixels associated with infected red blood cells. It recognizes the presence of trophozoites, which are the active feeding stage of the parasite, as well as gametocytes, which are the sexual forms that can be taken up by mosquitoes. By leveraging a predefined ontology that defines these relationships, the model classifies the image as "Parasitized" based on the detection of both features. The logical assertion "hasTrophozoite AND hasGametocyte" reflects that both stages are present in the analyzed image, allowing for a comprehensive understanding of the infection status and the specific life cycle stages of the parasites present. This reasoning process combines visual data with semantic knowledge from the ontology, ensuring an interpretable and clinically relevant classification.

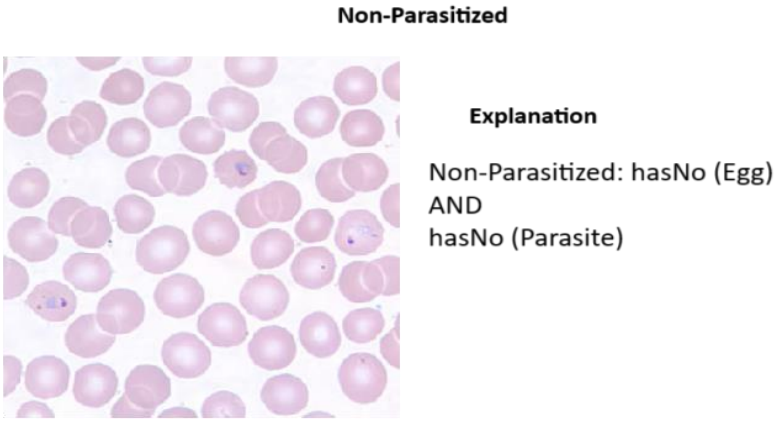

Figure 6: Example of Non-Parasitized Input Image with Corresponding Explanation

From Figure 5, the output "Non-parasitized: hasNo (Egg) AND hasNo (Parasite)" indicates that the ontology-based explainable model has analyzed the thin smear image and determined that it does not contain any signs of malaria infection. During the semantic segmentation process, the model examines the image for specific features that would indicate the presence of parasites, such as Plasmodium species or their developmental stages (like eggs). In this case, the model successfully identifies that there are no egg structures and no parasites present in the segmented regions of the image. By utilizing a predefined ontology, the model classifies the image as "Non-parasitized" based on the logical assertions "hasNo (Egg) AND hasNo (Parasite)," confirming the absence of any parasitic infection. This output provides a clear and interpretable result for clinicians, indicating that the thin smear does not show evidence of malaria, which is essential for accurate diagnosis and treatment decisions.

7 Conclusion

The integration of ontology-based systems with CNN models offers a promising avenue for the development of explainable AI in medical diagnostics, particularly for malaria detection. By providing structured knowledge and reasoning frameworks, ontologies enhance the transparency and interpretability of CNN predictions, addressing one of the key challenges in the application of deep learning in healthcare. As research in this area advances, ontology-based explainable models have the potential to significantly improve the accuracy, trust, and reliability of automated malaria detection systems.

Future research should focus on refining ontology frameworks, enhancing their alignment with CNN-based feature extraction, and developing standardized evaluation metrics for explainability in medical AI systems. Additionally, efforts should be made to ensure that ontology-based systems can scale across different diagnostic tasks and medical conditions, thereby broadening their applicability.

References

[1] Shalini S. *Comparision of the Peripheral Smear Examination by Leishman Stain, Fluorescent Stain and Immunochromatographic Test in Diagnosis of Malaria* (Doctoral dissertation, Rajiv Gandhi University of Health Sciences (India)).

[2] Acherar A, Tannier X, Tantaoui I, Brossas JY, Thellier M, Piarroux R. *Evaluating Plasmodium falciparum automatic detection and parasitemia estimation: A comparative study on thin blood smear images.* Plos one. 2024 Jun 3;19(6):e0304789.

[3] Rashmi R. *A Comparative Study of Blood Smear, Quantitative Buffy Coat and Antigen Detection for Diagnosis of Malaria* (Doctoral dissertation, Rajiv Gandhi University of Health Sciences (India)).

[4] Fox T. *Interrogation of Blood Smears by Digital Microscopy and Machine Learning* (Master's thesis, The University of Utah).

[5] Ogbaga I. *Artificial intelligence (AI)-based solution to malaria fatalities in Africa: An exploratory review.*

[6] Pan WD, Dong Y, Wu D. *Classification of malaria-infected cells using deep convolutional neural networks. Machine learning: advanced techniques and emerging applications.* 2018 Sep 19;159.

[7] Attai K, Ekpenyong M, Amannah C, Asuquo D, Ajuga P, Obot O, Johnson E, John A, Maduka O, Akwaowo C, Uzoka FM. *Enhancing the Interpretability of Malaria and Typhoid Diagnosis with Explainable AI and Large Language Models*. Tropical Medicine and Infectious Disease. 2024 Sep 16;9(9):216.

[8] Jillahi KB, Thandekkattu SG. *A Framework for IoT Data Collection and Fusion in Infectious Diseases Surveillance. InDesigning Sustainable Internet of Things Solutions for Smart Industries* 2025 (pp. 229-278). IGI Global.

[9] Tek FB, Dempster AG, Kale I. *Malaria parasite detection in peripheral blood images.* BMVA

[10] Yang, F., Ganaie, M. A., Zhang, M., & Fang, X. (2020). *Deep convolutional neural networks with a hierarchical classification approach for malaria parasite detection in blood smears. IEEE Access*, 8, 85069-85078.

[11] Litjens G, Kooi T, Bejnordi BE, Setio AA, Ciompi F, Ghafoorian M, Van Der Laak JA, Van Ginneken B, Sánchez CI. *A survey on deep learning in medical image analysis.* Medical image analysis. 2017 Dec 1;42:60-88.

[12] Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, Thrun S. *Dermatologist-level classification of skin cancer with deep neural networks.* Nature. 2017 Feb;542(7639):115-8.

[13] Caruana R, Lou Y, Gehrke J, Koch P, Sturm M, Elhadad N. *Intelligible models for healthcare: Predicting pneumonia risk and hospital 30-day readmission.* InProceedings of the 21th ACM SIGKDD international conference on knowledge discovery and data mining 2015 Aug 10 (pp. 1721-1730).

[14] Rector AL, Qamar R, Marley T. *Binding ontologies and coding systems to electronic health records and messages.* Applied Ontology. 2009 Jan 1;4(1):51-69.

[15] Smith B, Ashburner M, Rosse C, Bard J, Bug W, Ceusters W, Goldberg LJ, Eilbeck K, Ireland A, Mungall CJ, *Obi Consortium. The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration.* Nature biotechnology. 2007 Nov;25(11):1251-5.

[16] Pereira, M., Carvalho, T., & Gattass, M. (2020). *Ontology-based deep learning system for malaria diagnosis. Journal of Medical Systems*, 44(5), 90.

[17] Doshi-Velez F, Kim B. *Towards a rigorous science of interpretable machine learning.* arXiv preprint arXiv:1702.08608. 2017 Feb 28.

[18] Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. *Grad-CAM: visual explanations from deep networks via gradient-based localization.* International journal of computer vision. 2020 Feb;128:336-59.

[19] Bourguin G, Lewandowski A, Bouneffa M, Ahmad A. *Towards ontologically explainable classifiers. InArtificial Neural Networks and Machine Learning*–ICANN 2021: 30th International Conference on Artificial Neural Networks, Bratislava, Slovakia, September 14–17, 2021, Proceedings, Part II 30 2021 (pp. 472-484). Springer International Publishing.

[20] Lipton ZC. *The mythos of model interpretability: In machine learning, the concept of interpretability is both important and slippery.* Queue. 2018 Jun 1;16(3):31-57.

[21] Ribeiro MT, Singh S, Guestrin C. " *Why should i trust you?" Explaining the predictions of any classifier.* InProceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining 2016 Aug 13 (pp. 1135-1144).

[22] Ronneberger O, Fischer P, Brox T. U-net: *Convolutional networks for biomedical image segmentation.* InMedical image computing and computer-assisted intervention–*MICCAI* 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18 2015 (pp. 234-241). Springer International Publishing.

[23] Sandler M, Howard A, Zhu M, Zhmoginov A, Chen LC. *Mobilenetv2: Inverted residuals and linear bottlenecks.* InProceedings of the IEEE conference on computer vision and pattern recognition 2018 (pp. 4510-4520).

[24] Zhitomirsky-Geffet M, Erez ES, Judit BI. *Toward multiviewpoint ontology construction by collaboration of non-experts and crowdsourcing: The case of the effect of diet on health.* Journal of the Association for Information Science and Technology. 2017 Mar;68(3):681-94.