Malaria Parasite Detection using Efficient Neural Ensembles

Saurav Mishra

Masters Graduate - Liverpool John Moores University

Corresponding author: Saurav Mishra (e-mail: Saurav.Mishra@live.com).

ABSTRACT Caused by the bite of the Anopheles mosquito infected with the parasite of genus Plasmodium, malaria has remained a major burden towards healthcare for years, with approximate 400,000 deaths reported globally every year. The traditional diagnosis process for malaria involves an examination of the blood smear slide under the microscope. This process is not only time-consuming but also requires pathologists to be highly skilled in their work. Timely diagnosis and availability of robust diagnostic facilities and skilled laboratory technicians are very much vital to reduce the mortality rate. This study aims to build a system that help in timely and accurate diagnosis of malaria which would help in reducing the mortality rate and eventually help is attaining a malaria free environment. Applying deep learning techniques such as transfer learning and snapshot ensemble to automate the detection of the parasite in the thin blood smear images. Snapshot ensemble, a technique to create better performing ensembles with a limited training budget. Instead of training multiple models, snapshots are recorded during the training phase, which are later ensembled to create one strong model. All the models were evaluated against the following metrics - F1 score, Accuracy, Precision, Recall, Mathews Correlation Coefficient (MCC), Area Under the Receiver Operating Characteristics (AUC-ROC) and the Area under the Precision Recall curve (AUC-PR). The snapshot ensemble model created by combining the snapshots of the EfficientNet-B0 pre-trained model outperformed every other model achieving a f1 score - 99.37%, precision - 99.52% and recall - 99.23%. The results show the potential of model ensembles which combine the predictive power of multiple weal models to create a single efficient model that is better equipped to handle the real-world data. The GradCAM experiment displayed the gradient activation maps of the last convolution layer to visually explicate where and what a model sees in an image to classify them into a particular class. The models in this study correctly activate the stained parasitic region of interest in the thin blood smear images. Such visuals make the model more transparent, explainable, and trustworthy which are very much essential for deploying AI based models in the healthcare network.

INDEX TERMS Deep Learning, Biomedical Imaging, Malaria Infected Cells, Thin Blood Smears, Digital Pathology.

I. INTRODUCTION

Malaria, a common disease which poses life threatening risks is mostly prevalent in the tropical and sub-tropical regions of the world. One of the oldest diseases known to mankind is caused by the bite of the *Anopheles* mosquito that acts as the vector for parasite transmission. The blood sucking event injects the parasite present in its saliva into the person's blood stream. Studies show that the parasites trick the liver to support their growth and replication [1]. The Journal homepage: jeeemi.org

parasites first multiply in the liver and later conquer the blood stream. Once the parasites are transmitted to the blood stream they multiply and invade the red blood cells, eventually killing them. The transmission of parasites between the vectors and the humans is a cyclic process in which the blood sucking event from an already infected person transmits parasites back into the mosquito where they multiply and grow again. Thus, a cyclic process is

established where the mosquito acts as a vector passing on the parasite from one person to another [2]. Malaria is usually caused by the *P. falciparum, P. malariae, P. ovale, P. vivax* and *P. knowlesi* species of the plasmodium in humans [3]. The *P. falciparum* (~75%) and *P. vivax* (~25%) species are known to have the major share on the global burden of malaria. Though *P. falciparum* accounts for most deaths, recent evidence suggests that *P. vivax* could also cause potential life-threatening conditions. Severe levels of the disease can even lead to complications of the respiratory system, infection spreading to the brain, acute renal failure, severe anemia, and bleeding complications. About 10% of patients with severe malaria die, primarily due to multiorgan dysfunction [4].

According to the World Malaria Reports 2020 released by the World Health Organization an estimate of 229 million cases were reported globally in the year 2019. About 409000 deaths were reported globally in the year 2019. India accounted for about 86% of deaths in the South-East Asia Region [5]. Infants, children under 5 years, pregnant women are at a higher risk of getting infected by the parasite. Malaria during pregnancy can cause severe consequences leading to miscarriage, pre-mature delivery, lower than normal birth weight, congenital infection, and can even lead to perinatal death [6]. With such high risk and severity seen especially in case of pregnant women and children (below age of 5 years), it becomes essential to have a more robust system to be in place for an early and prompt diagnosis A timely and accurate diagnosis remains the utmost priority for a smoother recovery path.

The detection of the parasite in the blood stream of the infected person is done using microscopic examination with infected person's blood specimen spread on a slide either as a thick or thin smear, stained with the staining agent to segment out the blood cells for examination under the microscope by highly qualified pathologists. Thick blood smears aid in detecting the presence of malarial parasites while the thin smears aid in parasite species identification and quantification [7]. However, this is a very time consuming and exhaustive process which requires pathologists to be highly skilled in their job. Also, in rural and remote regions that do not have good medical facilities, a timely diagnosis is a challenge which could affect the patient's health condition and may even prove fatal at times. Regions where malaria is still categorized as an endemic is either due to the absence of proper health care facilities or due to the shortage of skilled pathologists [8]. United Nation and the Bill & Melinda Gates Foundation target to eradicate malaria globally by 2040 [9] while India is eyeing to attain the status of malaria free nation by the year 2030 as part of the Government of India's initiative - National Framework for Malaria Elimination In India [10]. To achieve the target, there needs to be robust systems available and accessible globally which aim for quick and prompt diagnosis. Such systems will help in reducing the mortality rates and consequently contribute towards eradicating malaria.

There have been numerous research activities towards detection of malarial parasites using computer vision,

machine learning and deep learning techniques. The computer vision techniques involve segmentation of erythrocytes and feature extraction which are then fed to a classification algorithm. However, extracting good features is a challenging task and requires expert level domain knowledge. Recent development in the deep learning systems have shown to be able to extract and learn the relevant features automatically. Convolutional Neural Networks, Transfer Learning and Model Ensembles are some application of deep learning systems which stand out in image classification tasks performing at par with the human level intelligence.

The objective of this study is to build a deep learning system that help in timely and accurate diagnosis of malaria using a discounted ensemble building technique known as Snapshot Ensemble that assist in building strong learners at the cost of training only a single deep learning model.

The rest of the paper is organized into the sections of literature review which discuss the methods applied in previous studies related to malaria classification, material and methods section which discusses the methods applied in this study, results and discussion section which analyzes the results obtained in the experiments of this study and compares them with the previous studies and finally the *conclusion* section which summarizes the study and mentions the scope for future work.

II. LITERATURE REVIEW

Image processing, and machine learning based algorithms have been used extensively for detection of malarial parasites. Machine learning algorithms require hand crafted features to be fed as input. Feature extraction can be a complicated task that requires an expert level domain knowledge and is also error prone. Various image processing methods are applied to remove unwanted information, enhance the quality, and extract relevant pixel-based features from the images. However, deep learning removes the burden of manual feature extraction, and has turned out to be the preferred choice for malaria classification in the recent times.

Image processing techniques like filtering, contrast stretching, segmentation, thresholding, and morphological operations have been applied to enhance the image quality and extract features. Filtering techniques have been widely used to remove any noise present in the whole slide images. Di Ruberto et al. [11] apply a (5x5) median filter for smoothening and noise removal. Díaz et al. [12] utilized the low-pass filter to get rid of the noisy components from the slide images. Anggraini et al. [13] applied median filter for noise removal. May et al. [14] implemented median filtering technique to get rid of any salt-and-pepper noise in the image and also suggested the use of Weiner filter to get rid of any gaussian noise in the image. Savkare and Narote [15] recommend application of Laplacian filtering technique for image smoothening and edge enhancement. Gaussian low-pass filter has been implemented by Arco et al. [16]. Dong et al. [17] apply filtering techniques as part of the image pre-processing phase to avoid learning of irrelevant information

by the neural network. Loddo et al. [18] discuss the application of morphological filtering techniques for noise removal.

Anggraini et al. [13] apply contrast stretching to boost the properties of the parasite objects. May et al. [14] applied histogram stretching to enhance the contrast of the images. Histogram Equalization is applied by Savkare and Narote [15] to enhance contrast along with Partial Contrast Stretching, a linear mapping which works on the range of the average of the maximum and minimum red, green, blue intensity values. Adaptive Histogram Equalization is applied by Arco et al. [16] for localized contrast corrections to enhance the region of interest. A wide range of contrast correction techniques have been implemented to enhance the images but the partial contrast stretching, and adaptive histogram equalization prove to perform better on the pathological images.

Di Ruberto et al. apply simple thresholding technique to segment the region of interest and apply morphological closing to separate the white blood cells from the red blood cells [11]. Morphological difference is analyzed for pattern recognition and classification of parasites. Anggraini et al. [13] and May et al. [14] applied Otsu's algorithm to determine the objects of interest. A two-stage color segmentation technique to identify the region of interest and remove the white blood cells is applied by Prasad et al. [19]. Khan et al. [20] take the unsupervised learning approach and apply K-means clustering multiple times to segment the infected erythrocytes from other blood components. Savkare and Narote [15] and Bairagi and Charpe [21] apply Otsu's and Watershed algorithm with morphological opening to separate the overlapping red cells. Arco et al. [16] apply the adaptive threshold mechanism along with morphological closing to improve the segmentation of objects of interest. Dong et al. [17] applied thresholding with Hough Circle transformation to separate overlapping erythrocytes. GI et al. [22] applied a two-staged segmentation process involving adaptive thresholding followed by watershed algorithm to detach the overlapping red cells. Pan et al. [23] apply Otsu's algorithm followed by morphological closing to remove the irrelevant information. Nag et al. [24] apply watershed algorithm to separate the overlapping objects.

The degree of success for any machine learning system depends on the quality of features fed to the algorithm. Extracting the best features requires an expert-level domain knowledge and understanding the morphology of various blood components and the parasite life cycle. Various set of features have been extracted to solve the problem and have enjoyed a varied level of success. Savkare and Narote [15] extract textual and morphological features based on mean, standard deviation. kurtosis, area, perimeter, major/minor axis of the erythrocytes. Bairagi and Charpe [21] apply Gray-Level Co-occurrence Matrix (GLCM) to extract the textual features like the contrast, pixel's correlation with the neighbors and energy (sum of the squares of the elements in the GLCM matrix). Park et al. [25] extract features based on morphology, geometry as elongation, equivalent diameter, eccentricity and statistical features of skewness, kurtosis.

Bashir et al. [26] extract features related to skewness based on the pixel intensity values. Dong et al. [17] use the Kullback-Leibler distance to determine the best features. Bibin et al. [27] used histogram-based features, color coherence vector and textual features like the Haralick features derived from the GLCM matrix, Local Binary Pattern features and Gray Level Run Length Matrix features to train the system. Gezahegn et al. [28] utilize the Scale Invariant Feature Transformation to create a bag-of-features. Nag et al. [24] extract features based on texture and morphology to filter the white blood cells.

The extracted features are fed to the algorithms which try to learn the relations and representations of the features that enable to detect the presence of parasites. Many algorithms have been implemented relishing a varied level of success. Ross et al. [29] implemented a two-stage backpropagation feed-forward network system classifying the erythrocytes as infected and not infected in the 1st stage and classify the parasite species in the 2nd stage achieving a sensitivity of 85% and positive predictive value of 81%. Tek et al. [30] propose a generic sliding window technique on top of the Ada-Boost algorithm. Anggraini et al. developed an algorithm completely based on image processing and thresholding techniques to detect the parasite in the erythrocytes [13]. May et al. [14] propose a segmentation based system to segment out the infected cells and achieve the positive predictive value of 98.90%. Khan et al. [20] take the unsupervised learning approach and implement the K-means clustering algorithm to extract two clusters - the malarial parasite infected cells and the other components achieving a precision of 0.95 and recall of 0.93. Savkare and Narote [15] implement the Support Vector Machine with the Radial Basis Function kernel in a two stage classification system. The correct identification rate of the binary classifier being 99.43% while the SVM classifier achieves an accuracy of 96.42% in identifying the life cycle stage of the parasite. Park et al. [25] conduct several experiments based on the Linear Discriminant Classification, Logistic Regression, and K-Nearest Neighbors algorithms achieving a classification accuracy greater than 95%. Bairagi and Charpe [21] apply the SVM algorithm and achieve an accuracy of 97.7%. Gezahegn et al. [28] proposed SVM-RBF based classification system. The system performs sub-optimally achieving an accuracy of 78.89%. Nag et al. [24] conduct multiple experiments based on K-Nearest Neighbors, Naïve Bayes, and Support Vector Machine with RBF kernel. The SVM model gives the best results with an accuracy of 97.59%. An ensemble of the above 3 models is created which achieves an accuracy of 98.74%.

The deep learning framework eliminates the need to extract hand-crafted features. Liang et al. [31] propose a 17-layer custom CNN model and a transfer learning approach utilizing the pre-trained AlexNet [32] as feature extractor linked to the SVM classifier for final classification achieving a classification accuracy of 97.37% and 91.99% respectively. Dong et al. [17] evaluate the performance of pre-trained deep neural models like LeNet, AlexNet and GoogleNet. GoogleNet outperformed others owing to its

greater depth and thus the ability to extract more detailed features. Bibin et al. [27] implement a deep belief network with 4 hidden layers each trained as a Restricted Boltzmann Machine to distinguish the stained objects. The system outperforms most state-of-the-art models with a small error rate of 0.0379. GI et al. [22] utilize the focus stack of images on a CNN model achieving sensitivity of 97.06%, specificity of 98.50% and Matthews Correlation Coefficient of 0.7305. Pan et al. [23] evaluate the effect of data augmentation on pre-trained LeNet-5 model and achieve an accuracy above 90%. Rajaraman et al. [33] assess the performance of pre-trained state-of-the-art models on malaria images. Features were extracted from various layers of these models and observed that the final layer of the models does not always provide optimal features. ResNet-50 outperformed others achieving an accuracy of ~95%. Hung et al. [34] propose a model based on faster region based network where convolution is done only once per image. A Region Proposal Network proposes various object regions with bounding boxes and the Object Detector based on the AlexNet performs the final classification of the objects as schizonts, trophozoites achieving an accuracy of 98%. A lightweight CNN model composed of 11 layers is proposed by Yang et al. [35] achieves an average accuracy of 93.46%. Rahman et al. [36] perform transfer learning experiments with VGG-16 pre-trained model achieving a test accuracy of 97.77% and a hybrid model combining CNN and SVM. The ensemble of the above models achieves an accuracy of 97.78%. Rajaraman et al. [37] propose ensemble of deep networks with various transfer learning models acting as the weak learners. The ensemble of VGG-19 and SqueezNet outperforms every other model with an accuracy of 99.51%. Vijayalakshmi et al. [38] take the amalgamated training approach by combining the feature extraction power of the VGG-19 and the classification power of the SVM achieving an accuracy of 93.13%.

An extensive literature survey helps to draw out the following observations on the various techniques applied on the whole slide images-

1. Median filters are the most widely used for noise removal.
2. Otsu's and Watershed algorithm followed by morphological operations in widely used for segmentation of various components and mark the region of interest.
3. Various features based on pixel, texture, spectral, color, statistics, etc. have been extracted to enhance the pattern identification in the cellular components.
4. SVM with RBF kernel has been widely used for malaria classification.

However, the entire process of feature extraction could be error-prone and since all the tasks in the pipeline are sequential, any fault that may have happened in the initial stages, gets forwarded to the model resulting in an inaccurate diagnosis. Recent studies have started to explore deep learning which removes the process of manual feature extraction. This saves time and minimizes the need to have an expert-level domain knowledge. The lack of labelled data is avoided through the data augmentation techniques.

III. MATERIALS AND METHODS

A. DATASET

1) DESCRIPTION

The dataset to be used for this study is taken from the National Library of Medicine, part of the National Institute of Health [39]. The dataset is an archive of red blood cells segmented using the Giemsa stained slides. The samples are taken from 150 infected and 50 uninfected persons at the Malaria screener research activity in CMC hospital in Bangladesh. A total of 27,558 erythrocyte images comprising of equal instances (13779) from each category. **FIGURE 1** shows a few data samples from the parasitized and non-parasitized category.
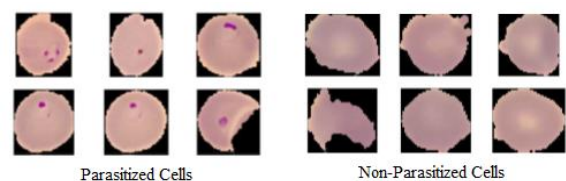

FIGURE 1 - Sample Data

2) DATA ANALYSIS

Bio-medical images are very diverse. For a similar pathological condition, images could vary invariably from person to person and even for the same person for different encounters. These differences may be attributed to the variation in lighting conditions, difference in marker stains (for pathological tests), image extraction process, image dimension, etc. Image Preprocessing ensures that all the images are in the same standard format and clean from any noise that would not add any relevance for the analysis. In this study, all images are resized to a dimension of (135, 135) and every image goes through the normalization and standardization process centering the pixel values around the mean to ensure the faster convergence in the training phase. This will result in a simple, accurate and a robust classification system.

3) DATA SPLIT

Since deep learning systems require huge amount of training data to learn all the underlying image patterns and representations, the dataset is split in the ratio 75:15:10. The training set gets 75% of the data while the test set and the validation set get 15% and 10% of the data respectively.

B. SYSTEM DESIGN AND ARCHITECTURE

Convolution Neural Network (CNN) form the base of all the models and experiments to be conducted. An image in its raw format translates to an array of pixel values. The neighboring pixels are highly correlated and usually form the basis of feature extraction. CNNs exploit this correlation by applying the techniques like local receptive fields, weight sharing, pooling and use of multiple layers making the entire architecture deep.

1)   PROGRAMMING RESOURCES

Programming task in the study is done on a cloud based system with CUDA enabled Nvidia Tesla K80 GPU, 4 core CPUs, 20 GB RAM. The programs are written in Python 3.6 using the web-based Anaconda Jupyter environment. The deep learning models are created using Keras library with Tensorflow 2.2 backend enabled with GPU acceleration. All the image processing and computer vision tasks are carried out using the Open Source Computer Vision (OpenCV) library.

2)   CUSTOM MODEL

A custom model with a total of 13 layers is designed having 3 convolutional layers with 32, 64, and 128 filter units respectively, 2 max-pooling and a Global Average Pooling layer is designed from scratch. Each convolutional unit is ReLU activated with Batch Normalization. Filters of size (3x3) with padding set to 'same' convolves the input. The Max Pool layers have pool window of (2x2). The final classification layer is a dense layer with one unit and sigmoid activation. The model design is shown in **FIGURE 2**.

3)   TRANSFER LEARNING

CNNs are basically feature extractors that learn the representation of an image. The state-of-the-art CNN models have learnt to extract features from millions of images. The skills of these models can be reused and applied to solve a related problem known as transfer learning. The main reasons to reuse the knowledge of the pre-trained models are lack of proper annotated data and computational resources. The lack of proper annotated bio-medical images makes transfer learning an important aspect when applying deep learning in the field of digital pathology. The initial layers of a CNN extract the generic features (corners and edges), the middle layers extract abstract features by aggregating the corners and edges, while the last few layers are utilized for classification of images based on the features extracted. The final layers of the pre-trained models might not be useful in classifying the pathological images and could be removed. A smaller learning rate is chosen to ensure all the previous knowledge of the base model is retained and reused. State-of-the-art architectures like ResNet50-V2, DenseNet-121, Inception-v3, Xception, InceptionResNet-V2, and EfficientNet have attained optimal performance by training over millions of images and will be utilized for transfer learning. **FIGURE 3** shows a transfer learning model implementing EfficientNet-B0 network as the convolutional base.

The dataset consists of only 13779 images of each category respectively. Deep learning systems need huge amount of data to learn all possible representations and be able to perform optimally in the real-world environment. To increase the diversity and amount of training data, various augmentation techniques like flipping, rotation, cropping, translation, illumination, scaling, shift, and zoom are applied.
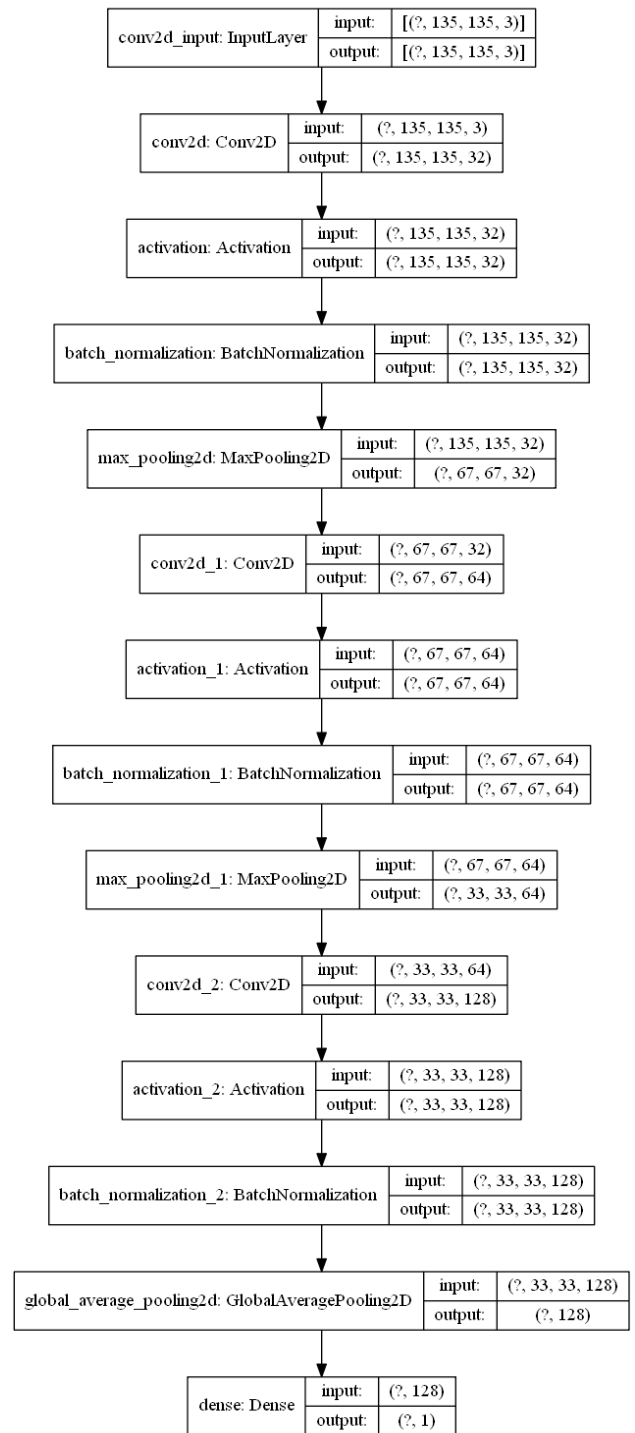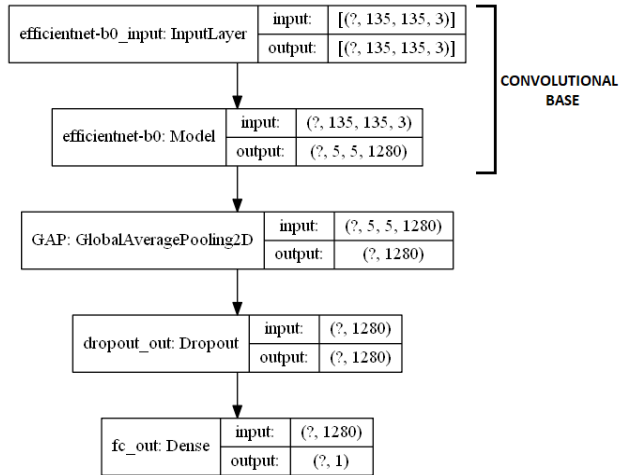

FIGURE 2 - Custom Model Design

FIGURE 3 EfficientNet-B0 Transfer Learning Model

4) SNAPSHOT ENSEMBLE

The overhead cost of training multiple deep neural networks could be very high in terms of the training time, hardware, and computational resource requirement and often acts as obstacle for creating deep ensembles. To overcome these barriers Huang et al. [40] proposed a unique method to create ensemble which at the cost of training one model, yields multiple constituent model snapshots that can be ensembled together to create a strong learner. The core idea behind the concept is to make the model converge to several local minima along the optimization path and save the model parameters at these local minima points. During the training phase, a neural network would traverse through many such points. The lowest of all such local minima is known as the Global Minima. The larger the model, more are the number of parameters and larger the number of local minima points. This implies, there are discrete sets of weights and biases, at which the model is making fewer errors. So, every such minimum can be considered a weak but a potential learner model for the problem being solved. Multiple such snapshot of weights and biases are recorded which can later be ensembled to get a better generalized model which makes the least amount of mistakes.
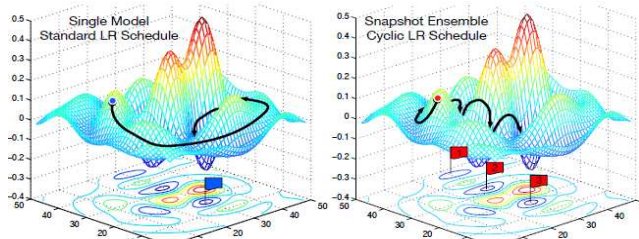

FIGURE 4. Snapshot Ensemble Cyclic LR

The left plot in **FIGURE 4** shows the global minimum point (blue flag) where the model converges to at the end of the training phase. The plot on the right, shows the model converging to 3 different local minimums (red flags) along its optimization path. At each local minimum point the model

weights and biases are saved. The training is then restarted by increasing learning rate. The idea in increasing the learning rate is to take a big step and escape the local minimum. The learning rate follows a cyclic pattern as suggested by Loshchilov and Hutter [41] where the learning rate is quickly raised and then lowered to follow the cosine pattern as seen in **FIGURE 5**.
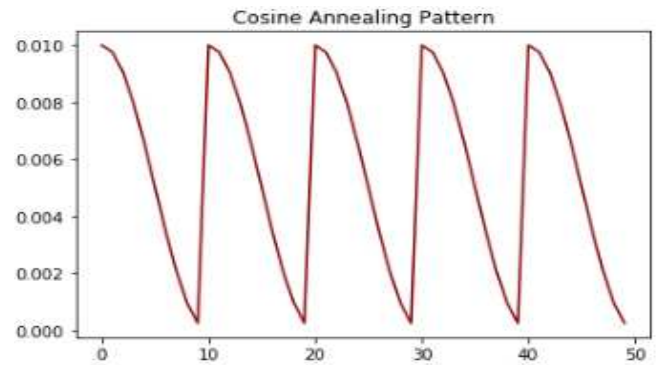

FIGURE 5. Cosine Annealing

Cosine Annealing proposed in [41] is used to decay the learning rate is defined as -

$$\propto (t) = \frac{\propto_0}{2} \left(\cos \left(\frac{\pi \, mod(t-1, \lceil T/M \rceil)}{\lceil T/M \rceil} \right) + 1 \right) \tag{1}$$

where, $\propto(t)$ is the learning rate at epoch t, $\propto 0$ is the maximum learning rate, T is the total number of epochs, M is the number of cycles, mod is the modulo operation, and square brackets indicate a floor operation.

At the end of full training cycle, there will be 'M' different models available ready to be ensembled with no additional overhead cost. During the test time, "the last 'm' model snapshots are always considered for the final ensemble as these tend to have the least error rate". The final output is the average of the snapshot models and is calculated as per the following equation defined in [40].

$$h_{Ensemble} = \frac{1}{m} \sum_0^{m-1} h_{M-i}(x) \tag{2}$$

where, x is the sample test set, $h_i(x)$ is the softmax output score of the ith snapshot.

For this experiment, (1) a custom model with 4 convolution layers as seen in **FIGURE 6** is implemented to create snapshots and (2) a snapshot ensemble model to be created out of the best performing transfer learning model.

5) GRAD-CAM - VISUAL EXPLANATIONS

The application of AI systems in healthcare domain is a challenging task mainly because the factors involved in arriving at a decision by the machines are not explainable. Questions like, how did the machine arrive at the decision? or what did the machine see to classify the blood smear as infected by malaria parasite? will always be asked to understand the machine's thought process to arrive at decisions in healthcare based AI systems. To understand the dynamics of deep networks Selvaraju et al. [42] proposed a technique to visually explain an AI system's decision and

make it more transparent. The Grad-CAM technique explains a model's decision making process answering, *"why they predict what they predict"*. Grad-CAM relies on convolutional layer as they tend to retain the spatial information. It utilizes the gradient activations coming into the last convolutional layer as these layer contain more class specific features. The gradient score of the concerned class is calculated with respect to the feature map activation of the convolution layer given by the formula -

$$\propto_k^c = \frac{1}{Z} \sum_i \sum_j \frac{\partial y^c}{\partial A_{ij}^k} \qquad (3)$$

Once the gradient scores are calculated, ReLU is applied to the weighted linear combination of the forward activation maps to consider only those features that have a positive impact on the class of interest.

$$L_{Grad-CAM}^c = ReLU(\sum_k \propto_k^c A^k) \qquad (4)$$

The Custom Model in **FIGURE 2**, will be utilized to visualize the internal dynamics of what a model sees to arrive at the decision.

6)   HYPERPARAMETERS & CALLBACKS
All the model experiments will be implemented using the hyperparameters and call backs mentioned in **TABLE 1**

TABLE 1
Model Training Parameters

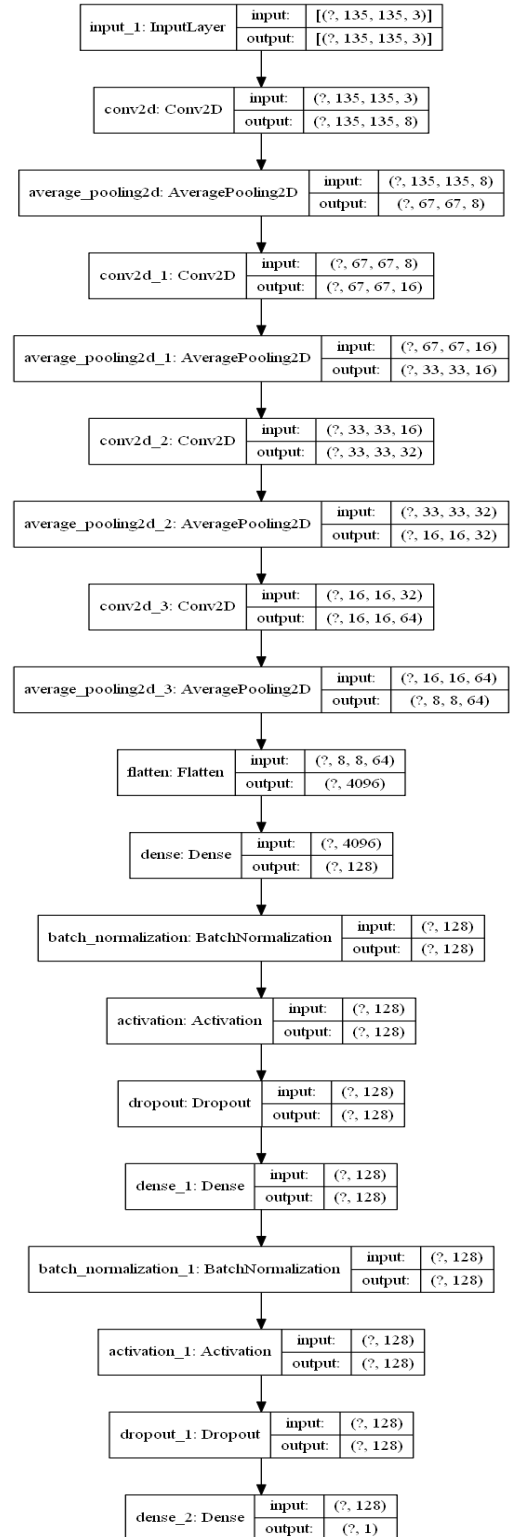| Hyperparameter	Value
Activation Function	ReLU, Sigmoid
Cost Function	Binary Cross Entropy
Learning Rate	1×10^{-3}
Optimizer	Adam
Epochs	50
Dropout Ratio	0.3 / 0.5
Batch Size	64
Training Callbacks	ModelCheckpoint, ReduceLROnPlateau Early Stopping Cosine Annealing

FIGURE 6. Custom Model for Snapshot Ensemble

TABLE 2
Custom Model Test Evaluation

Model	Confusion Matrix			F1 Score	Accuracy	Precision	Recall	MCC	AUC ROC	AUC PR
Custom Model	0	2016	45	96.93	96.95	97.79	96.09	93.92	96.95	97.92
	1	81	1992							
		0	1							

IV. RESULTS AND DISCUSSIONS

Several experiments were conducted with custom models, transfer learning and snapshot ensemble strategies. Custom model experiment was extended to analyze the network gradients and understand the dynamics of the model's approach in arriving at a decision. (**Note**: all the evaluation metrics are measured in percentage.)

A. CUSTOM MODEL

The custom model is set to train for 50 epochs with an early stopping strategy after 7 epochs. The model trained for 34 epochs before being early stopped. The training and validation history is seen in   **FIGURE 7.** The model converges well enough on training and validation data. The performance metrics on the test data of the model is summarized in **TABLE 2**. The model achieves an accuracy of 96.95% which is reasonably well for the baseline model. However, the number of false negatives is 81 which is quite high for a healthcare model for identifying disease. False negatives occur when a person having malaria is declared healthy. This will hamper the timely diagnosis, treatment for the patient and may have fatal consequences. The target is to reduce the number of false negatives.
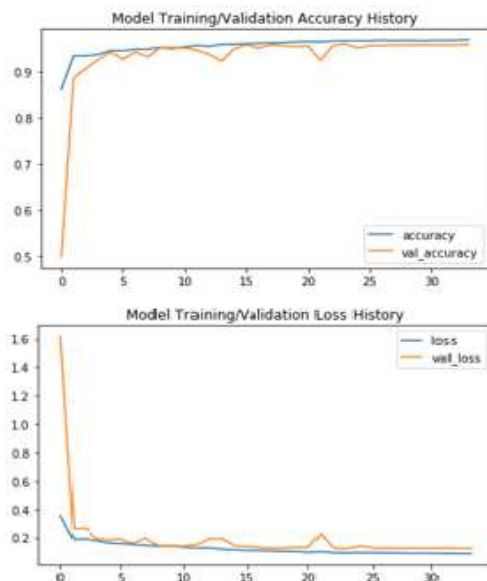

FIGURE 7. Custom Model Training History

B. GRAD-CAM VISUAL EXPLANATIONS

The custom model from the previous experiment was extended to visualize the gradients that help a model come to a decision. **FIGURE 8,** rows 1 & 3, show parasitized cells, where the model is looking and activating the stained parasitic region in the images correctly. For the image in the 2nd row, which is the non-infected cell, the model has no regions activated within the cell.

Visuals like these will help in designing explainable AI systems which are able to provide meaningful explanations as to why a particular decision was made? Such developments and capabilities would make the adoption of AI into the healthcare domain realistic and smoother.
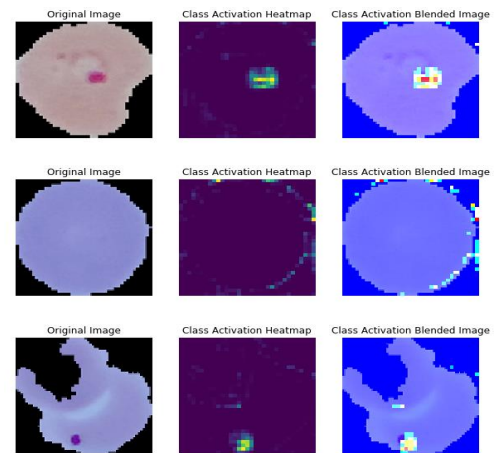
FIGURE 8. Grad-CAM Visual Explanations

C. TRANSFER LEARNING

Pre-trained model like 'DenseNet121', 'EfficientNet-B0', 'Inception-v3', 'InceptionResNetV2', 'ResNet50V2', and 'Xception' were utilized to check the efficiency of transfer learning in malaria classification problem. The last layer before the final classification layer were selected as feature extractors and the model were trained for 50 epochs with early stopping patience level set at 7. The training and validation summary for all the models are summarized in **FIGURE 9**.

The evaluation metrics obtained on the test data is summarized in **TABLE 3**. EfficientNet-B0 outperformed every other state-of-the-art model and achieved the highest F1 score of 97.95% followed by the DenseNet121 with an F1 score of 97.81%. EfficientNet-B0 records the least number of false negatives.

FIGURE 9. Transfer Learning Validation History

TABLE 3
Transfer Learning Test Evaluation Metrics
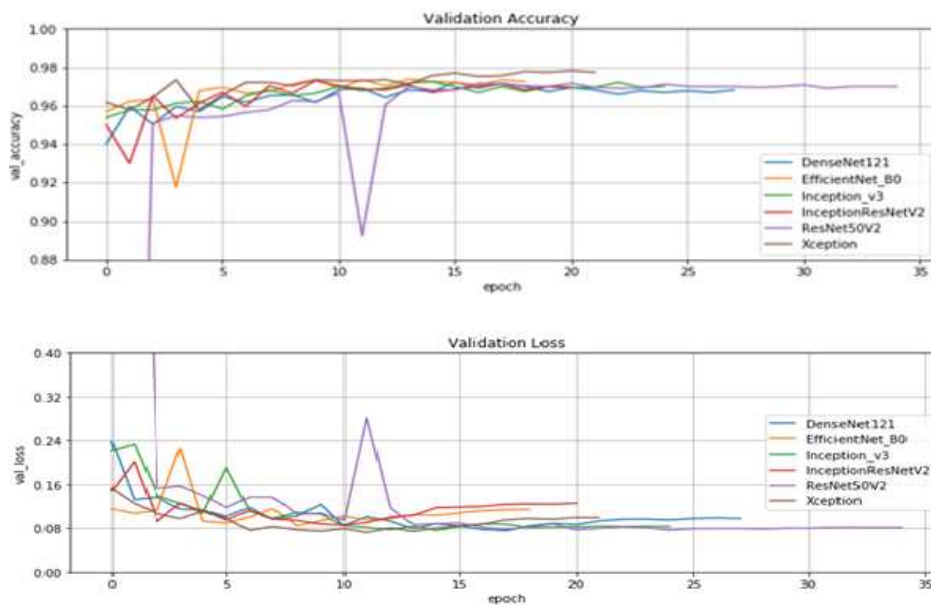
Model	Confusion Matrix			F1 Score	Accuracy	Precision	Recall	MCC	Loss	ROC AUC	AUC PR
DenseNet121	0	2015	46	97.81	97.80	97.78	97.83	95.60	5.82	97.80	98.35
	1	450	20281								
EfficientNet-B0	**0**	**2015**	**46**	**97.95**	**97.94**	**97.79**	**98.12**	**95.89**	**6.82**	**97.94**	**98.43**
	1	**390**	**20341**								
Inception-V3	0	2019	42	97.83	97.82	97.97	97.68	95.65	6.75	97.82	98.41
	1	480	20251								
InceptionResNet-V2	0	2016	45	97.76	97.75	97.83	97.68	95.50	6.75	97.75	98.34
	1	480	20251								
ResNet50-V2	0	2018	43	97.78	97.77	97.92	97.64	95.55	6.54	97.77	98.37
	1	490	20241								
Xception	0	2013	48	97.71	97.70	97.69	97.73	95.40	6.76	97.70	98.28
	1	470	20261								

D. SNAPSHOT ENSEMBLE

The ensemble experiment was performed on two different models to test the effectiveness of the ensembles and ensure that ensembles perform better than the single model. One, a custom model represented by **FIGURE 6** and second, the **EfficientNet-B0** model since it turned out to be the best performing model in the transfer learning experiment.

TABLE 4
Custom Model Snapshots Test Evaluation Metrics

Snapshot	Confusion Matrix			F1 Score	Accuracy	Precision	Recall	MCC	Loss	ROC AUC	AUC PR
Snapshot 1	0 1	2024 116	37 1957	96.24	96.30	98.14	94.40	92.67	11.45	97.94	98.43
Snapshot 2	0 1	2024 93	37 1980	96.82	96.86	98.17	95.51	93.75	8.99	96.86	97.96
Snapshot 3	0 1	2037 81	24 1992	97.43	97.46	98.81	96.09	94.96	7.19	97.46	98.43
Snapshot 4	0 1	2048 66	13 2007	98.07	98.09	99.36	96.82	96.21	6.06	98.09	98.88
Snapshot 5	**0** **1**	**2048** **48**	**13** **2025**	**98.52**	**98.52**	**99.36**	**97.68**	**97.06**	**5.98**	**98.53**	**99.10**

TABLE 5
Test Evaluation for Ensembled Models

Ensemble	Confusion Matrix			F1 Score	Accuracy	Precision	Recall	MCC	Loss	ROC AUC	AUC-PR
Snapshot Ensemble [1, 2, 3, 4, 5]	0 1	2044 75	17 1998	97.75	97.77	99.16	96.38	95.59	6.60	97.78	98.68
Snapshot Ensemble [3,4,5]	0 1	2048 62	13 2011	98.17	98.19	99.36	97.01	96.40	5.67	98.19	98.93
Snapshot Ensemble [4,5]	**0** **1**	**2048** **51**	**13** **2022**	**98.44**	**98.45**	**99.36**	**97.54**	**96.92**	**5.65**	**98.45**	**99.07**

1) CUSTOM MODEL SNAPSHOTS

The custom model was configured to utilize cosine annealing cyclic learning rate scheduler. The model was set to train for 50 epochs and the number of cycle was set to 5 since the ensembled model is most effective when the snapshots are extracted far apart during the training phase. The snapshots were recorded each at the end of 10th, 20th, 30th, 40th and 50th epochs respectively. Each of the snapshot was evaluated on the test data and the results are summarized in **TABLE 4**. The evaluation metrics gradually increases towards the final snapshots while the number of false predictions decrease steadily. Ensembled models were created by combining the snapshots. As mentioned in [40], "at the test time, the last 'm' models are always considered for the final ensemble as these tend to have the least error rate." snapshots 4 and 5 with test time loss of 6.06% and 5.98% respectively were always considered for creating the ensembled models.

Three different combinations of the snapshots were combined to create the ensembles. Ensemble 1 combined every snapshot [1 to 5], Ensemble 2 combined snapshots [3, 4, 5] and Ensemble 3 combined snapshots [4, 5]. Each of the ensembled model was evaluated on the test data and the results are summarized in **TABLE 5**. Ensemble 3 outperforms the other ensembles and the individual model obtained at the end of 50 epochs. Ensemble [4, 5] achieves a f1 score of 98.44% which is an increase of around 2.9% from the fully trained single model. Snapshots 1 and 2 do not contribute enough towards the ensemble performance. This could be accounted to a relatively higher loss as the model is still learning during the initial cycles of the training phase. Snapshots 3, 4 and 5 have much of the contribution to the ensembled model.

2) EFFICIENTNET-B0 SNAPSHOTS

The callback settings for cosine annealing was kept same as the custom model with 50 epochs and number of cycles set to 5. The snapshots were recorded each at the end of 10th, 20th, 30th, 40th and 50th epochs respectively. Each of the snapshot was evaluated on the test data and the results are summarized in **TABLE 6.** Here, five different combinations of the snapshots were tested to create the ensembled model. Ensemble 1 combined every snapshot [1 to 5], Ensemble 2 combined snapshots [2, 3, 4, 5], Ensemble 3 combined snapshot [3, 4, 5], Ensemble 4 combined snapshot [3, 4], and Ensemble 5 combined snapshot [4, 5]. Each of the ensembled model was evaluated on the test data and the results are summarized in **TABLE 7.**

E.   INTERPRETING PREDICTION SCORE

The model prediction on a test image gives the probability score for the cell being infected with the parasite. There are two samples in **FIGURE 10** with infection probability of 20.70% & 29.40% respectively. In both images a few

TABLE 6
EfficientNet-B0 Snapshots Test Evaluation

Snapshot	Confusion Matrix			F1 Score	Accuracy	Precision	Recall	MCC	Loss	AUC ROC	AUC PR
Snapshot 1	0 1	2039 46 0	22 2027 1	98.35	98.36	98.93	97.78	96.72	4.09	98.36	98.91
Snapshot 2	0 1	2047 28 0	14 2045 1	98.98	98.98	99.32	99.65	97.97	3.21	98.99	99.32
Snapshot 3	0 1	2047 21 0	14 2052 1	99.15	99.15	99.32	98.99	98.31	2.92	99.15	99.41
Snapshot 4	**0** **1**	**2048** **17** **0**	**13** **2056** **1**	**99.28**	**99.27**	**99.37**	**99.18**	**98.55**	**2.79**	**99.27**	**99.48**
Snapshot 5	0 1	2050 19 0	11 2054 1	99.28	99.27	99.47	99.08	98.55	3.19	99.27	99.51

TABLE 7
EfficientNet-B0 Ensembled Models

Ensemble	Confusion Matrix			F1 Score	Accuracy	Precision	Recall	MCC	Loss	AUC ROC	AUC PR
Ensemble [1, 2, 3, 4, 5]	0 1	2051 20 0	10 2053 1	99.27	99.27	99.52	99.04	98.55	2.53	99.28	99.52
Ensemble [2, 3, 4, 5]	0 1	2050 19 0	11 2054 1	99.28	99.27	99.47	99.08	98.55	2.50	99.27	99.51
Ensemble [3,4,5]	0 1	2049 18 0	12 2055 1	99.28	99.27	99.42	99.13	98.55	2.48	99.27	99.49
Ensemble [3,4]	0 1	2050 18 0	11 2055 1	99.32	99.32	99.52	99.13	98.65	2.56	99.32	99.54
Ensemble [4,5]	**0** **1**	**2051** **16** **0**	**10** **2057** **1**	**99.37**	**99.37**	**99.52**	**99.23**	**98.74**	**2.58**	**99.37**	**99.57**

TABLE 8
Infection Severity Score

Probability Score Range (%)	Infection Severity Level
0	Uninfected
0 - 5	Very Low
5 - 25	Low
25 - 50	Moderate
>50	High

scattered, and not so clear stained spots are visible. Such samples with smaller parasitic density could be in the initial stages of the infection which the model is unable to detect fairly. Such samples cannot be left as uninfected. This scoring aspect can be improved by **(a)** collecting more samples that are in the initial stages of infection **(b)** introduce an infection severity scoring system based on the prediction probability. The infection severity levels proposed are **High**, **Moderate**, **Low** and **Very Low**. The severity levels would correspond to the probability scores defined in **TABLE 8**. Based on the infection severity and new scoring system, **FIGURE 11** shows a few samples with probability score and infection severity.

F. FALSE PREDICTION ANALYSIS
All the experiments conducted report some amount of false predictions as false positives and false negatives. An analysis to verify these false predictions and examine a possible reason for false reports was conducted. The Snapshot Ensembled [4,5] model in **TABLE 7** was chosen for this task. The confusion matrix of this model looks like **[[2051, 10] [16, 2057]]**. The model reports a total of 26 false predictions of which 10 are false positives and 16 are false negatives. **FIGURE 12** and **FIGURE 13** show all the misclassified samples with their ground truth. As seen in the false positives, few samples clearly indicate the presence of the parasite after staining. However, they are labelled as 0 (uninfected). Similar pattern is seen in case of false negatives where samples that do not show presence of parasite after staining being labelled as 1 (parasitized). Thus, there is some amount of error in the data annotation phase. Since the data is labelled manually, errors could happen. Correcting the labels and re-training the models could help to improve the model classification performance.
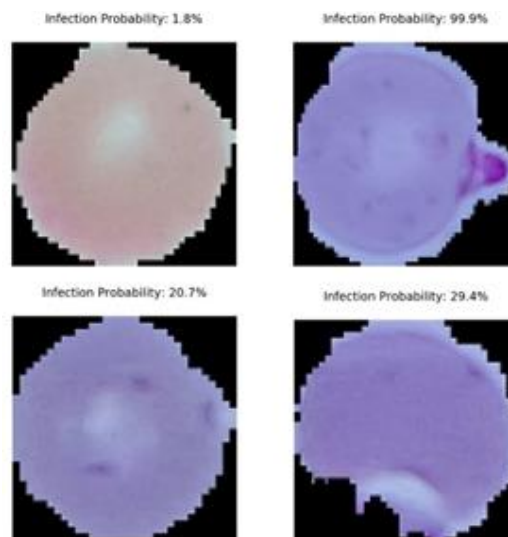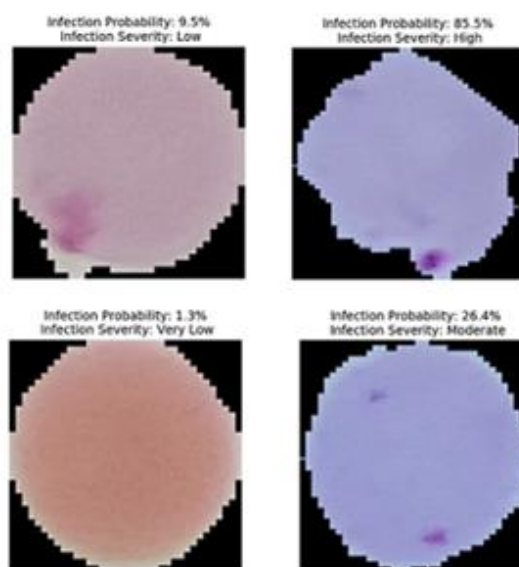

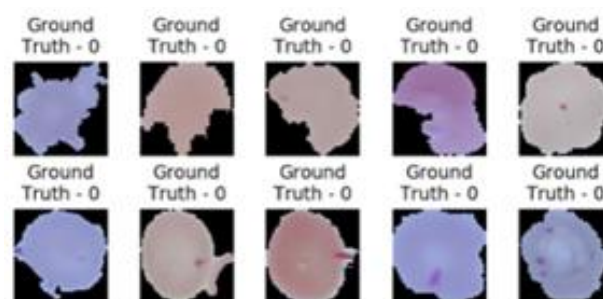**FIGURE 10.** Prediction Scores

FIGURE 11. Infection Severity
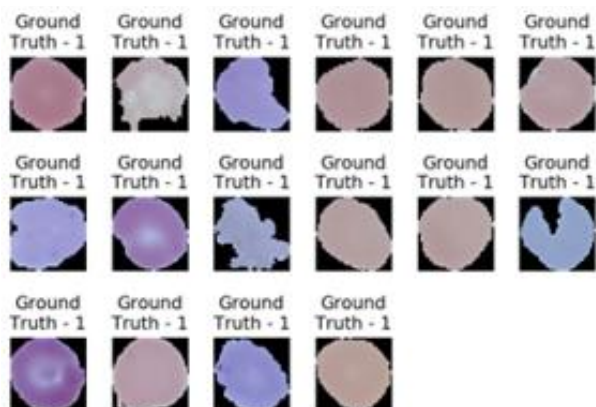

FIGURE 12. False Positives

FIGURE 13. False Negatives

G.   RESULT COMPARISON

The results obtained in this study are compared with the results of the previous work that have been conducted towards malaria classification and discussed as part of literature review. The comparison is shown in **TABLE 9**. The EfficientNet snapshot ensemble model outperforms most of the previous work is all aspect and achieves comparable result with the work of Rajaraman et al. [37] who also experiment the efficiency of neural network ensemble. However, a major improvement of this work is attributed to the fact that snapshot ensemble creates better performing ensembles by training only a single model. This not only saves training time but also creates ensembles with a limited training budget in terms of hardware resources and computational power.

The malaria detection system developed in this study will help in the accurate and timely diagnosis of malaria. The system can be deployed in remote regions which do not have access to skilled and experienced pathologists. Timely diagnosis can help design a smooth recovery path and eventually bring down the mortality rate. The integration of an explainable AI module via the GradCAM technique will offer an added advantage to encourage the usage of AI systems not only in the healthcare experts but also the patients.

However, this study is currently limited to the detection of only the P. falciparum parasite. Malaria can be caused by five different species of the parasite. The system once extended to support all the five species in a multi-class classification approach would become a complete end-to-end malaria detection system.

V. CONCLUSION

A series of experiments were conducted as part of this study building end-to-end deep learning systems for diagnosing malaria from thin blood smear whole slide images. Several custom network architectures were designed

TABLE 9
Result Comparison

Model	F1 Score	Accuracy	Precision	Sensitivity	Specificity	MCC
EfficientNet Snapshot Ensemble	99.37	99.37	99.52	99	99	98.74
EfficientNet Transfer Learning	97.78	97.77	97.97	98	98	95.55
Rahman et al., 2019	97.09	97.77	97.19	--	--	94.42
Vijayalakshmi and B, 2019	91.66	93.13	89.95	93.44	92.22	--
Rajaraman et al., 2019	99.5	99.5	99.8	--	--	99
Rajaraman et al., 2018b	98.7	98.6	--	98.1	99.2	97.2
Yang et al., 2019	80.81	97.26	78.98	82.73	98.39	--
Bibin et al., 2017	89.66	96.3	--	97.60	95.92	--
Gopakumar et al., 2017	--	97.37	--	97.06	98.5	73.05
Liang et al., 2016	97.36	97.37	97.73	96.99	97.73	94.75

along with transfer learning approach and snapshot ensemble technique. The healthcare domain where lack of proper annotated data is often considered as a limitation, transfer learning using state-of-the-art models which have a good knowledge base and serve as an efficient alternative to designing high performance deep learning models for the healthcare domain. In the transfer learning experiment, EfficientNet-B0 model turned out to be most efficient with a f1 score of 97.95% and MCC 95.89%. The ensemble learning implemented a relatively new approach of snapshot ensemble where multiple snapshots of the model are recorded during the training phase. Every snapshot is different from the other as they tend to converge at different local minima points recording unique set of weights and biases. Two different model architectures - a custom model and the Efficientnet-B0 model were setup to extract five distinct snapshots respectively. Many different ensembles were created using various combination of the snapshots from each architecture. Each ensembled model proved to be stronger than the single model. The EfficientNet-B0 ensemble with the final two snapshots gave the best classification scores achieving a f1 score of 99.37% and MCC score of 98.74%. This model also has the least number of false predictions meaning it generalizes well on unseen data and proves to be a great model in developing better clinical solution for malaria detection. Also, the fact that we get ensembled model by training only one model not only removes the need of high efficiency computational resources but also saves valuable amount of time required to train multiple large models to combine and create ensembles. The GradCAM experiment shows where exactly a model looks in the image to arrive at a decision. It is observed that the model activates the region around the parasite in the infected cells and utilizes it to differentiate the parasitized cells from unparasitized cells. The GradCAM visualizations make the models more transparent, explainable, and trustworthy which are very much essential for deploying AI based models in the healthcare network.

This study considers only the P. falciparum infected cells for malaria classification. Since, there are five different species of the malarial parasite namely, P. falciparum, P. malariae, P. ovale, P. vivax and P. knowlesi. The experiments conducted in this study can be extended to detect the other species of the parasite which can be taken up as the future work. Another important factor to consider the adoption of AI systems into healthcare practices is testing. Validating the model with external datasets i.e., samples from a different set of population or distribution would give the most correct measure for generalizability of the model. The model can be fine-tuned on the new population accordingly.

Code Repository: The python codes for the experiments in this study can be found at - https://github.com/sauravmishra1710/Malaria-Detection-Using-Deep-Learning-Techniques.

REFERENCES

[1] D. Posfai *et al.*, "Plasmodium parasite exploits host aquaporin-3 during liver stage malaria infection.," *PLoS Pathog.*, vol. 14, no. 5, p. e1007057, May 2018, doi: 10.1371/journal.ppat.1007057.

[2] C.-C. for D. C. and Prevention, "CDC - Malaria - About Malaria - Biology," 2020, Accessed: July 26, 2021. [Online]. Available: https://www.cdc.gov/malaria/about/biology/index.html.

[3] E. Worrall, S. Basu, and K. Hanson, "Is malaria a disease of poverty? A review of the literature," *Trop. Med. Int. Heal.*, vol. 10, no. 10, pp. 1047–1059, Oct. 2005, doi: 10.1111/j.1365-3156.2005.01476.x.

[4] D. Gollin and C. Zimmermann, "Malaria: Disease Impacts and Long-Run Income Differences," *Dep. Econ. Williams Coll. Dep. Econ. Work. Pap.*, no. 2997, Jan. 2008.

[5] WHO 2020, *World Malaria Report 2020*, vol. 73, no. 1. 1997.

[6] C. for D. C. Prevention, "CDC - Malaria - Diagnosis & Treatment (United States) - Treatment (U.S.) - Guidelines for Clinicians (Part 1)," 2020, Accessed: July 27, 2021. [Online]. Available: https://www.cdc.gov/malaria/diagnosis_treatment/clinicians1.html.

[7] "CDC - Malaria - Diagnosis & Treatment (United States) - Diagnosis (U.S.)," 2020.

[8] "Shortage of pathologists inhibits progress on UHC | The Daily Star," 2018. https://www.thedailystar.net/health/shortage-pathologists-inhibits-progress-uhc-1549609.

[9] UN and Gates Foundation, "UN and Gates Foundation presents vision for eradicating malaria by 2040 – United Nations Sustainable Development," 2015. https://www.un.org/sustainabledevelopment/blog/2015/09/report-by-un-and-gates-foundation-presents-vision-for-eradicating-malaria-by-2040/ (accessed July 27, 2021).

[10] Ministry of health & family welfare goi, "National Framework For Malaria Elimination In India (2016-2030) Directorate Of National Vector Borne Disease Control Programme (NVBDCP) Directorate General Of Health Services (DGHS) Ministry Of Health & Family Welfare Government Of India," 2016. Accessed: July 27, 2021. [Online].Available: http://www.indiaenvironmentportal.org.in/files/file/National-framework-for-malaria-elimination-in-India-2016–2030.pdf.

[11] C. Di Ruberto, A. Dempster, S. Khan, and B. Jarra, "Morphological Image Processing for Evaluating Malaria Disease BT - Visual Form 2001," 2001, pp. 739–748.

[12] G. Díaz, F. A. González, and E. Romero, "A semi-automatic method for quantification and classification of erythrocytes infected with malaria parasites in microscopic images," *J. Biomed. Inform.*, vol. 42, no. 2, pp. 296–307, Apr. 2009, doi: 10.1016/J.JBI.2008.11.005.

[13] D. Anggraini, A. S. Nugroho, C. Pratama, I. E. Rozi, A. A. Iskandar, and R. N. Hartono, "Automated status identification of microscopic images obtained from malaria thin blood smears," in *Proceedings of the 2011 International Conference on Electrical Engineering and Informatics*, 2011, pp. 1–6, doi: 10.1109/ICEEI.2011.6021762.

[14] Z. May, S. S. A. M. Aziz, and R. Salamat, "Automated quantification and classification of malaria parasites in thin blood smears," in *2013 IEEE International Conference on Signal and Image Processing Applications*, 2013, pp. 369–373, doi: 10.1109/ICSIPA.2013.6708035.

[15] S. S. Savkare and S. P. Narote, "Automated system for malaria parasite identification," in *2015 International Conference on Communication, Information & Computing Technology (ICCICT)*, 2015, pp. 1–4, doi: 10.1109/ICCICT.2015.7045660.

[16] J. E. Arco, J. M. Górriz, J. Ramírez, I. Álvarez, and C. G. Puntonet, "Digital image analysis for automatic enumeration of malaria parasites using morphological operations," *Expert Syst. Appl.*, vol. 42, no. 6, pp. 3041–3047, 2015, doi: https://doi.org/10.1016/j.eswa.2014.11.037.

[17] Y. Dong *et al.*, "Evaluations of deep convolutional neural networks for automatic identification of malaria infected cells," in *2017 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI)*, 2017, pp. 101–104, doi: 10.1109/BHI.2017.7897215.

[18] A. Loddo, C. Di Ruberto, and M. Kocher, "Recent Advances of

Malaria Parasites Detection Systems Based on Mathematical Morphology," *Sensors (Basel).*, vol. 18, no. 2, p. 513, Feb. 2018, doi: 10.3390/s18020513.

[19] K. Prasad, J. Winter, U. M. Bhat, R. V Acharya, and G. K. Prabhu, "Image Analysis Approach for Development of a Decision Support System for Detection of Malaria Parasites in Thin Blood Smear Images," *J. Digit. Imaging*, vol. 25, no. 4, pp. 542–549, 2012, doi: 10.1007/s10278-011-9442-6.

[20] N. A. Khan, H. Pervaz, A. K. Latif, A. Musharraf, and Saniya, "Unsupervised identification of malaria parasites using computer vision," in *2014 11th International Joint Conference on Computer Science and Software Engineering (JCSSE)*, 2014, pp. 263–267, doi: 10.1109/JCSSE.2014.6841878.

[21] V. K. Bairagi and K. C. Charpe, "Comparison of Texture Features Used for Classification of Life Stages of Malaria Parasite," *Int. J. Biomed. Imaging*, vol. 2016, p. 7214156, 2016, doi: 10.1155/2016/7214156.

[22] G. GI, M. Swetha, S. S. Gorthi, and G. Sai Subrahmanyam, "CNN based Malaria Diagnosis from Focus-stack of Blood Smear Images Acquired using Custom-built Slide Scanner," *J. Biophotonics*, vol. 11, p. e201700003, Aug. 2017, doi: 10.1002/jbio.201700003.

[23] W. D. Pan, "Classification of Malaria-Infected Cells Using Deep Convolutional Neural Networks," Y. Dong, Ed. Rijeka: IntechOpen, 2018, p. Ch. 8.

[24] S. Nag, N. Basu, and S. K. Bandyopadhyay, "Application of Machine Intelligence in Digital Pathology: Identification of Falciparum Malaria in Thin Blood Smear Image BT - Advancement of Machine Intelligence in Interactive Medical Image Analysis," O. P. Verma, S. Roy, S. C. Pandey, and M. Mittal, Eds. Singapore: Springer Singapore, 2020, pp. 65–97.

[25] H. S. Park, M. T. Rinehart, K. A. Walzer, J.-T. A. Chi, and A. Wax, "Automated Detection of P. falciparum Using Machine Learning Algorithms with Quantitative Phase Images of Unstained Cells," *PLoS One*, vol. 11, no. 9, p. e0163045, Sep. 2016, [Online]. Available: https://doi.org/10.1371/journal.pone.0163045.

[26] A. Bashir, Z. A. Mustafa, I. Abdelhameid, and R. Ibrahem, "Detection of malaria parasites using digital image processing," in *2017 International Conference on Communication, Control, Computing and Electronics Engineering (ICCCCEE)*, 2017, pp. 1–5, doi: 10.1109/ICCCCEE.2017.7867644.

[27] D. Bibin, M. S. Nair, and P. Punitha, "Malaria Parasite Detection From Peripheral Blood Smear Images Using Deep Belief Networks," *IEEE Access*, vol. 5, pp. 9099–9108, 2017, doi: 10.1109/ACCESS.2017.2705642.

[28] Y. G. Gezahegn, Y. H. G. Medhin, E. A. Etsub, and G. N. G. Tekele, "Malaria Detection and Classification Using Machine Learning Algorithms BT - Information and Communication Technology for Development for Africa," 2018, pp. 24–33.

[29] N. E. Ross, C. J. Pritchard, D. M. Rubin, and A. G. Dusé, "Automated image processing method for the diagnosis and classification of malaria on thin blood smears," *Med. Biol. Eng. Comput.*, vol. 44, no. 5, pp. 427–436, May 2006, doi: 10.1007/s11517-006-0044-2.

[30] F. B. Tek, A. G. Dempster, and I. Kale, "Computer vision for microscopy diagnosis of malaria," *Malar. J.*, vol. 8, no. 1, p. 153, 2009, doi: 10.1186/1475-2875-8-153.

[31] Z. Liang *et al.*, "CNN-based image analysis for malaria diagnosis," in *2016 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 2016, pp. 493–496, doi: 10.1109/BIBM.2016.7822567.

[32] A. Krizhevsky, I. Sutskever, and G. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," *Neural Inf. Process. Syst.*, vol. 25, Jan. 2012, doi: 10.1145/3065386.

[33] S. Rajaraman *et al.*, "Pre-trained convolutional neural networks as feature extractors toward improved malaria parasite detection in thin blood smear images," *PeerJ*, vol. 6, p. e4568, 2018, doi: 10.7717/peerj.4568.

[34] J. Hung *et al.*, *Applying Faster R-CNN for Object Detection on Malaria Images*. 2018.

[35] F. Yang *et al.*, "Deep Learning for Smartphone-Based Malaria Parasite Detection in Thick Blood Smears," *IEEE J. Biomed. Heal. Informatics*, vol. PP, p. 1, Sep. 2019, doi: 10.1109/JBHI.2019.2939121.

[36] A. Rahman *et al.*, "Improving Malaria Parasite Detection from Red Blood Cell using Deep Convolutional Neural Networks," Jul. 2019, Accessed: July 28, 2021. [Online]. Available: https://arxiv.org/abs/1907.10418.

[37] S. Rajaraman, S. Jaeger, and S. K. Antani, "Performance evaluation of deep neural ensembles toward malaria parasite detection in thin-blood smear images," *PeerJ*, vol. 7, p. e6977, 2019, doi: 10.7717/peerj.6977.

[38] A. Vijayalakshmi and R. K. B, "Deep learning approach to detect malaria from microscopic images," *Multimed. Tools Appl.*, 2019, doi: 10.1007/s11042-019-7162-y.

[39] S. Jaeger, "Malaria Datasets," *National Library of Medicine*, 2018. https://lhncbc.nlm.nih.gov/LHC-publications/pubs/MalariaDatasets.html (accessed July 28, 2021).

[40] G. Huang, Y. Li, G. Pleiss, Z. Liu, J. E. Hopcroft, and K. Q. Weinberger, "Snapshot ensembles: Train 1, get M for free," in *5th International Conference on Learning Representations, ICLR 2017 - Conference Track Proceedings*, Mar. 2017, Accessed: July 18, 2021. [Online]. Available: http://arxiv.org/abs/1704.00109.

[41] I. Loshchilov and F. Hutter, "SGDR: Stochastic Gradient Descent with Warm Restarts," Aug. 2016, Accessed: July 28, 2021. [Online]. Available: http://arxiv.org/abs/1608.03983.

[42] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization," Oct. 2016, doi: 10.1007/s11263-019-01228-7.

Saurav Mishra was born in India in 1987. He received the Bachelor in Technology degree in Computer Science & Engineering in 2009, M.S. degree in Wireless Embedded Technology from the Manipal University, India in 2012, and a Master's degree in Artificial Intelligence & Machine Learning from the Liverpool John Moore's University, U.K., in 2020 .He is currently working as a software developer in the healthcare domain and aims to contribute towards the research and development space in the adoption of Artificial Intelligence in healthcare domain. His research interests include the adoption of Deep Learning and Computer Vision in the medical diagnostics and human biology space.