Title : Towards a timely and accurate AI-based Computer-Aided Diagnosis approach in healthcare

Full name of Speakers : Quang-Vu Nguyen and Vu-Thu-Nguyet Pham

Institution :
The University of Danang – Vietnam-Korea University of Information and Communication Technology

Project Introduction

Background :

- The healthcare landscape in the ASEAN region faces numerous challenges
- The integration of AI into healthcare systems has the potential to mitigate these challenges

Targets:

This proposal seeks to adapt and optimize these technologies for the unique healthcare environment of ASEAN:

- Develop a cutting-edge Computer-Aided Diagnosis (CAD) system which can efficiently analyze diverse datasets, including medical images (such as X-rays and MRIs), electronic health records (EHRs), pathology reports, and laboratory results
- Develop standardized interfaces and data exchange protocols that enable healthcare professionals to effortlessly incorporate the CAD system into their existing workflows

Proposed Method

1. Data Aggregation and Integration

Clinical

Two clinical datasets were utilized: one comprising 150 disorder categories with approximately 1,000 different symptoms, and another containing 132 distinct symptoms and 41 different diseases.

The data was divided into three groups: 60 percent for training, 20 percent for model validation, and 20 percent for testing.

Para-clinical

- The VinDr-CXR dataset consists of over 100,000 DICOM images, sourced from two major Vietnamese hospitals. It incorporates 18,000 labeled chest X-ray images, featuring 22 significant abnormalities and 6 diagnoses, all expert-annotated.

This dataset is partitioned into two segments: the training set, comprising 15,000 images, and the test set, encompassing 3,000 images.

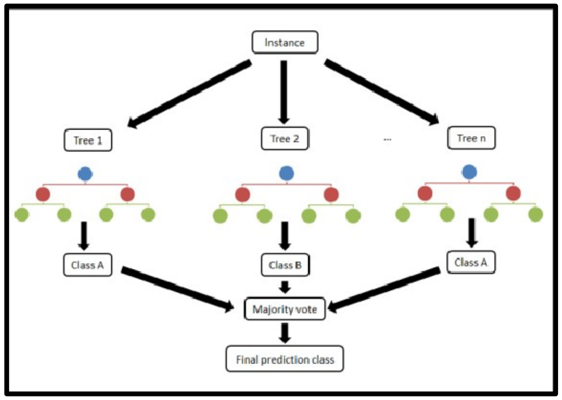
- There are four datasets used for the malaria diagnosis system: The first, Broad Bioimage Benchmark Collection consists of 1364 blood smear images stained with Giemsa, total 80,000 cells and available in the ".jpg" or ".png" format; The second Inherited from the study by Y. M. Kassim et al, this dataset was collected at Chittagong Medical College Hospital in Bangladesh, comprising 965 images with dimensions of 5312x2988 obtained from test results of 193 patients, containing approximately 162,700 cells; Two other datasets consisting of approximately 6000 blood smear cells obtained from 300 patients with malaria and 50 non-infected patient.

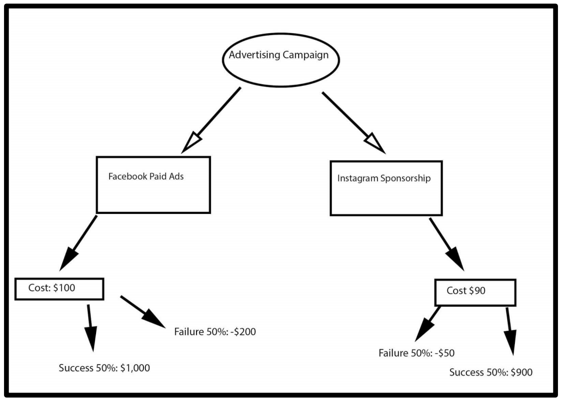
Proposed Method

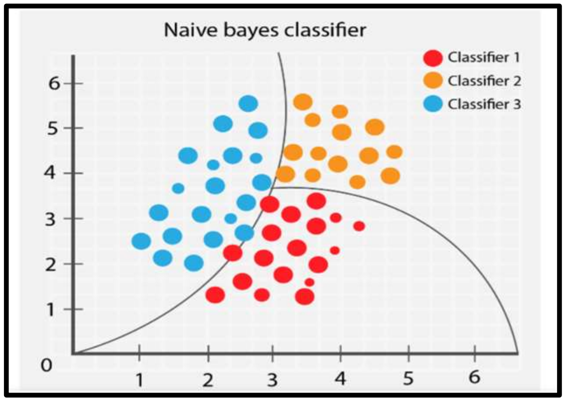
2. Algorithms

Clinical diagnosis - Models

The study involves experimentation with three machine learning models: Random Forest, Decision Trees, and Naïve Bayes.





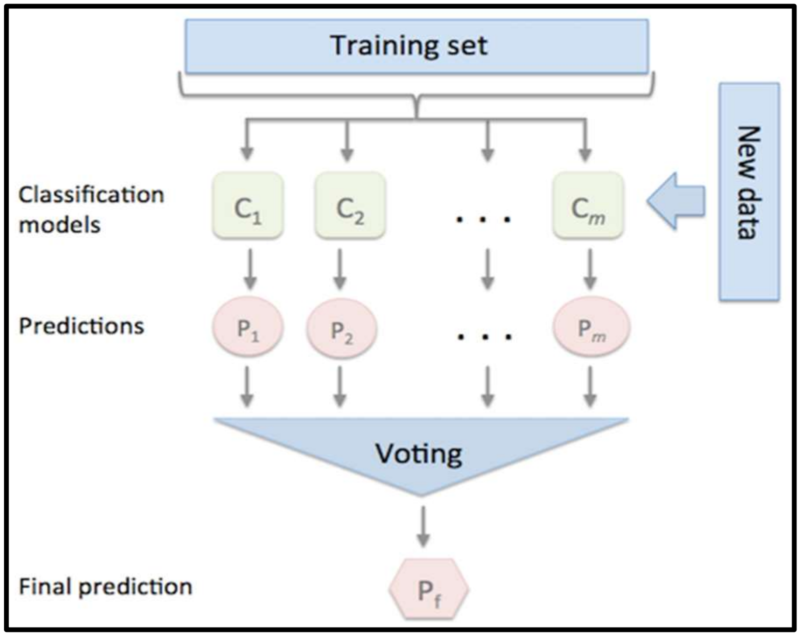


Proposed Method

2. Algorithms

Clinical diagnosis - Models

- We created a Voting Ensemble Model, which aims to enhance overall model performance by combining the predictions of multiple models.
- The Voting Ensemble aggregates predictions for each label, and the label with the majority vote becomes the final prediction.



Proposed Method

2. Algorithms

Clinical diagnosis - Evaluation metrics

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \#(1)$$

$$Precision = \frac{TP}{TP + FP} \#(2)$$

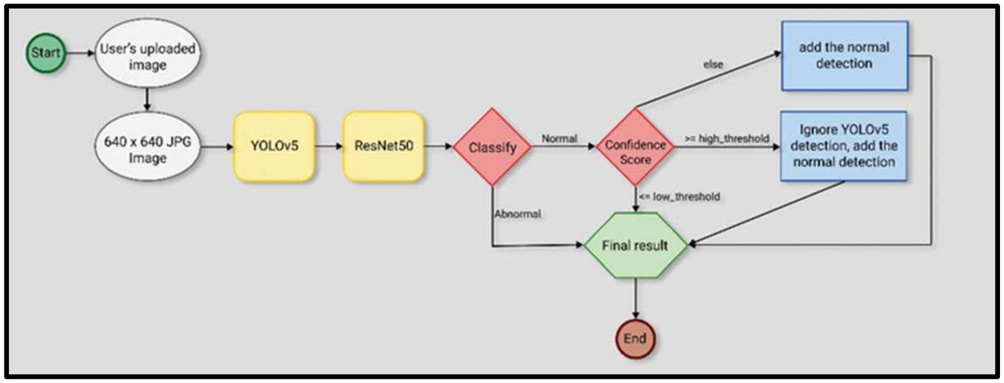
$$Recall = \frac{TP}{TP + FN} \#(3)$$

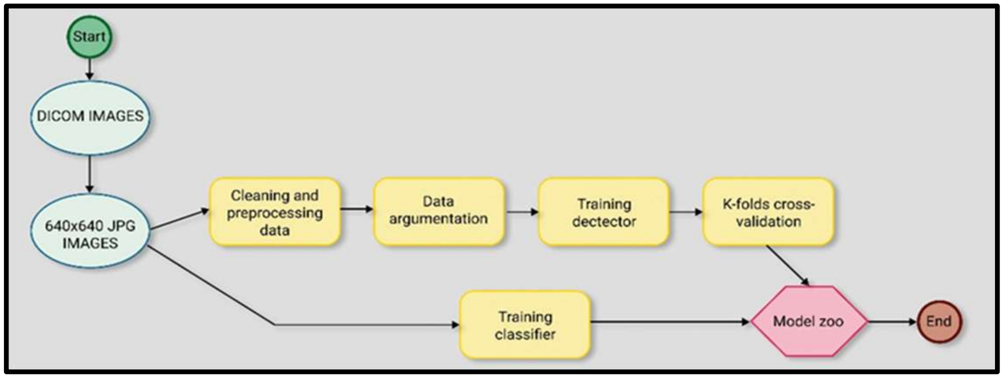
$$F1 - score = \frac{2 * (precision * recall)}{(precision + recall)} \#(4)$$

Proposed Method

2. Algorithms

Chest X-ray diagnosis - flow diagrams



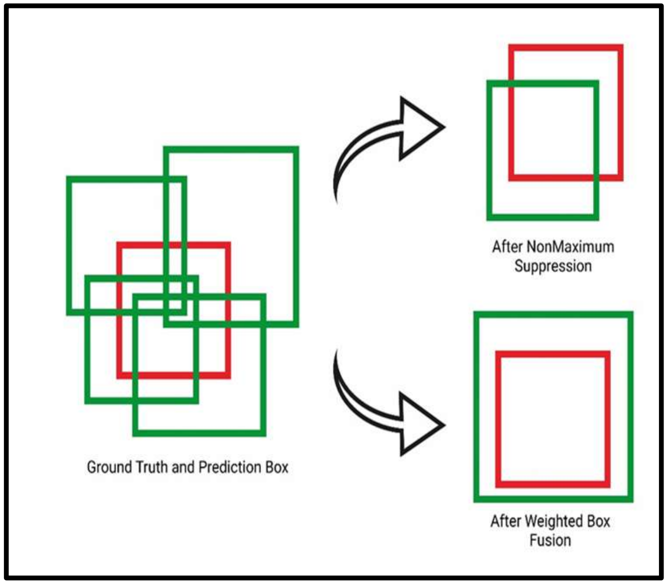


Proposed Method

2. Algorithms

Chest X-ray diagnosis - preprocessing

- Convert DICOM images to JPG format with a specific size for further processing.
- Utilize the Weighted Boxes Fusion (WBF) method to streamline the labeling process and enhance accuracy in combined predictions.

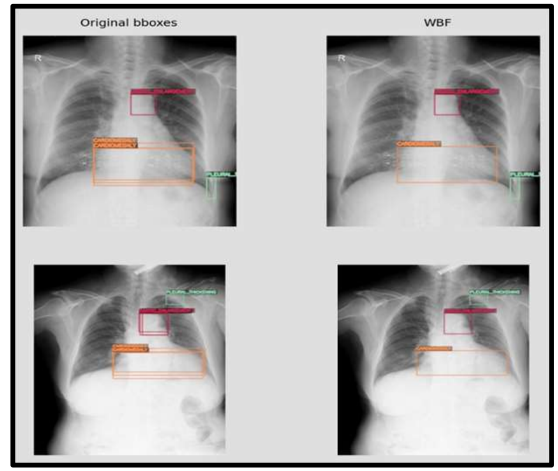


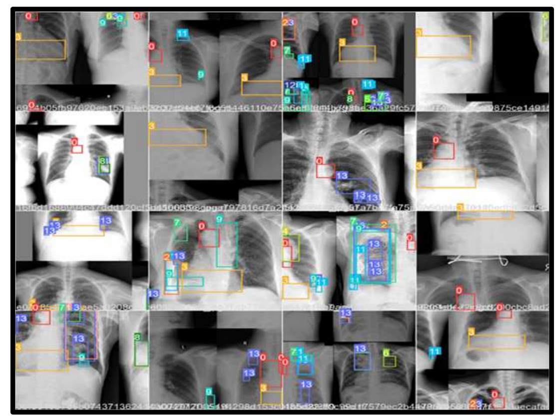
Proposed Method

2. Algorithms

Chest X-ray diagnosis - preprocessing

- Sort the bounding boxes based on confidence scores and create combinations using the Intersection over Union (IoU) criteria.
- Use the Albumentations library to enhance the diversity of the training dataset and improve the performance and robustness of machine learning models.





Proposed Method

2. Algorithms

Chest X-ray diagnosis - Evaluation metrics

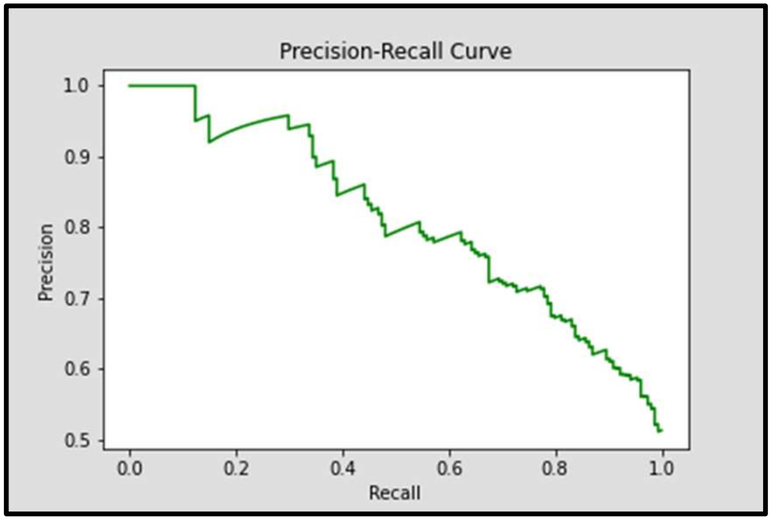
K-fold cross-validation with k=5

$$mAP = \frac{1}{n}\sum_{k=1}^{n} AP_k \,.\#(5)$$

IoU = 0.5

(mAP@0.5)

$$IoU(A,B) = \frac{area(A \cap B)}{area(A \cup B)}\#(6)$$



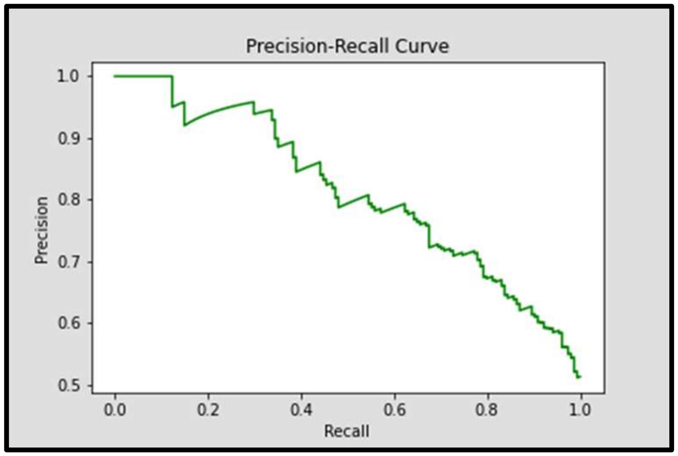
Proposed Method

2. Algorithms

Chest X-ray diagnosis - Evaluation metrics

$$AP = \Sigma(recall_n - recall_{n-1}) \cdot precision_{interp}(recall_n) \#(7)$$

$$precision_{interp}(recall_n) = \max_{\tilde{r} \geq r_n} p(\tilde{r}) \#(8)$$



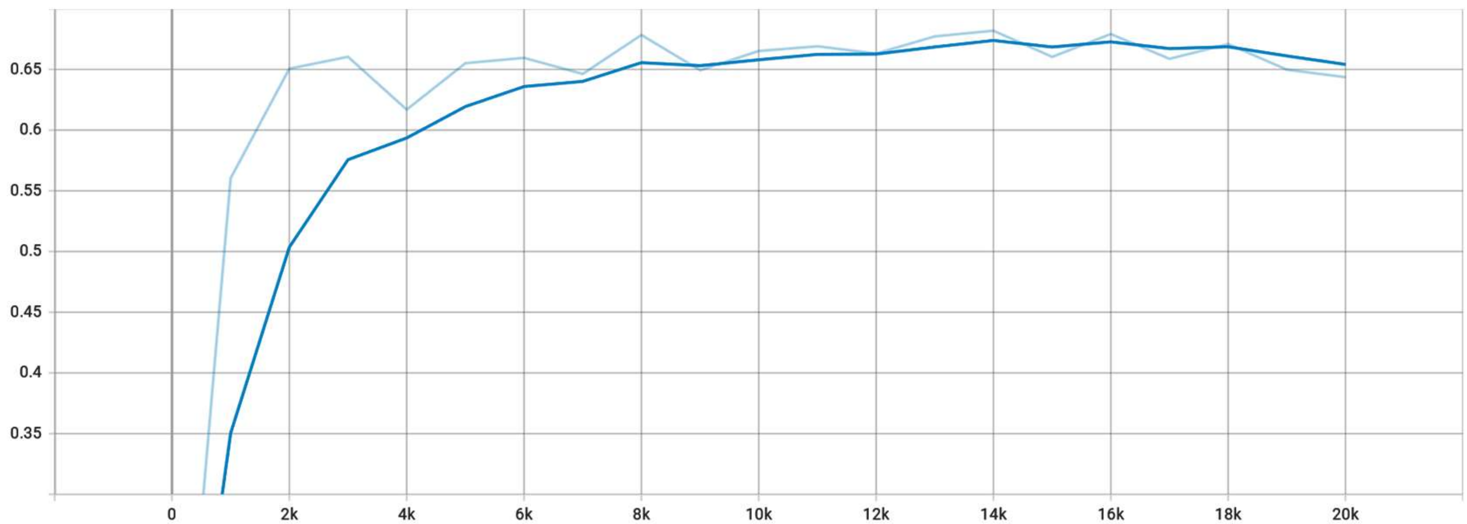
Proposed Method

2. Algorithms

Malaria diagnosis - Approach methods

Two-stage approach

Utilized the ResNet101 model as the feature extractor and initialized the Faster R-CNN model with pre-trained weights from ImageNet. The model was trained for 20,000 epochs



Evaluation results of the Faster R-CNN with the mAP@.5 matrix

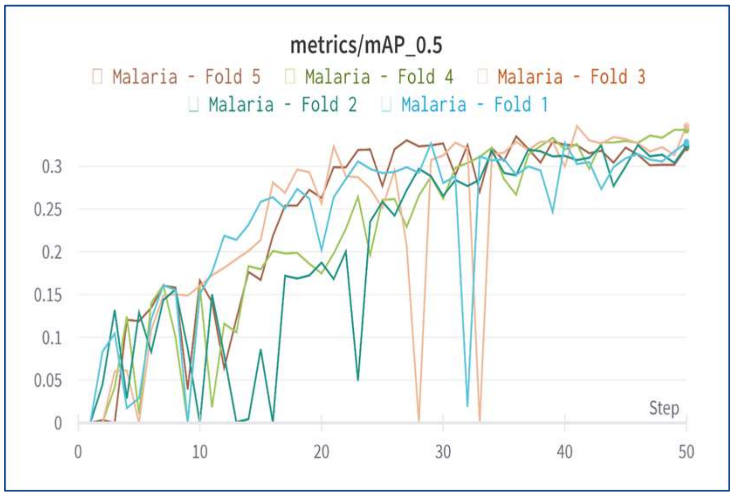
Proposed Method

2. Algorithms

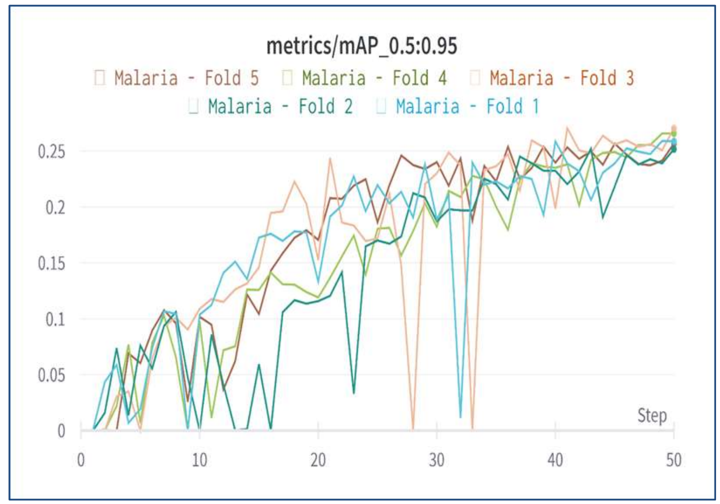
Malaria diagnosis - Approach methods

One-stage approach

YOLOv5 as the representative of the one-stage object detection models. The training set was shuffled and divided into 5 equal folds, with each fold using 4 sets for training and one set for validation.The model was trained for 50 epochs for each fold



Training results of the YOLOv5 model on mAP@0.5



Training results of the YOLOv5 model on mAP@0.5:0.95

Proposed Method

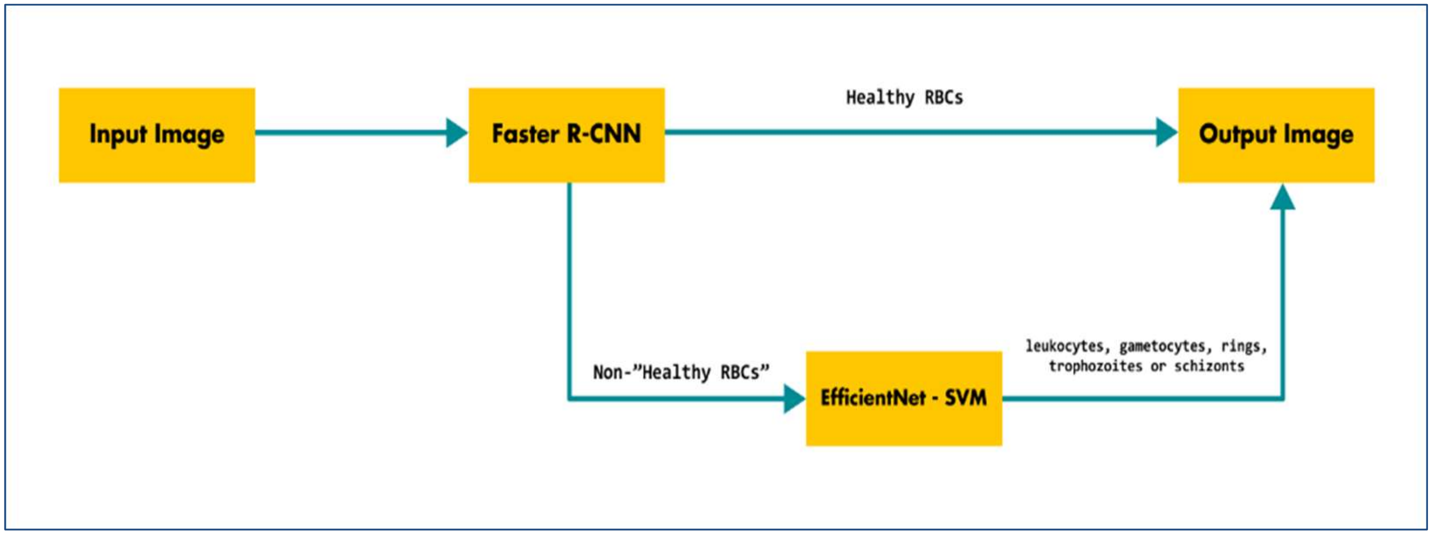
2. Algorithms

Malaria diagnosis - Approach methods

Proposed approach

Stage one: Using Faster R-CNN as binary classifier to classify Heathy RBCs or Non – Health RBCs

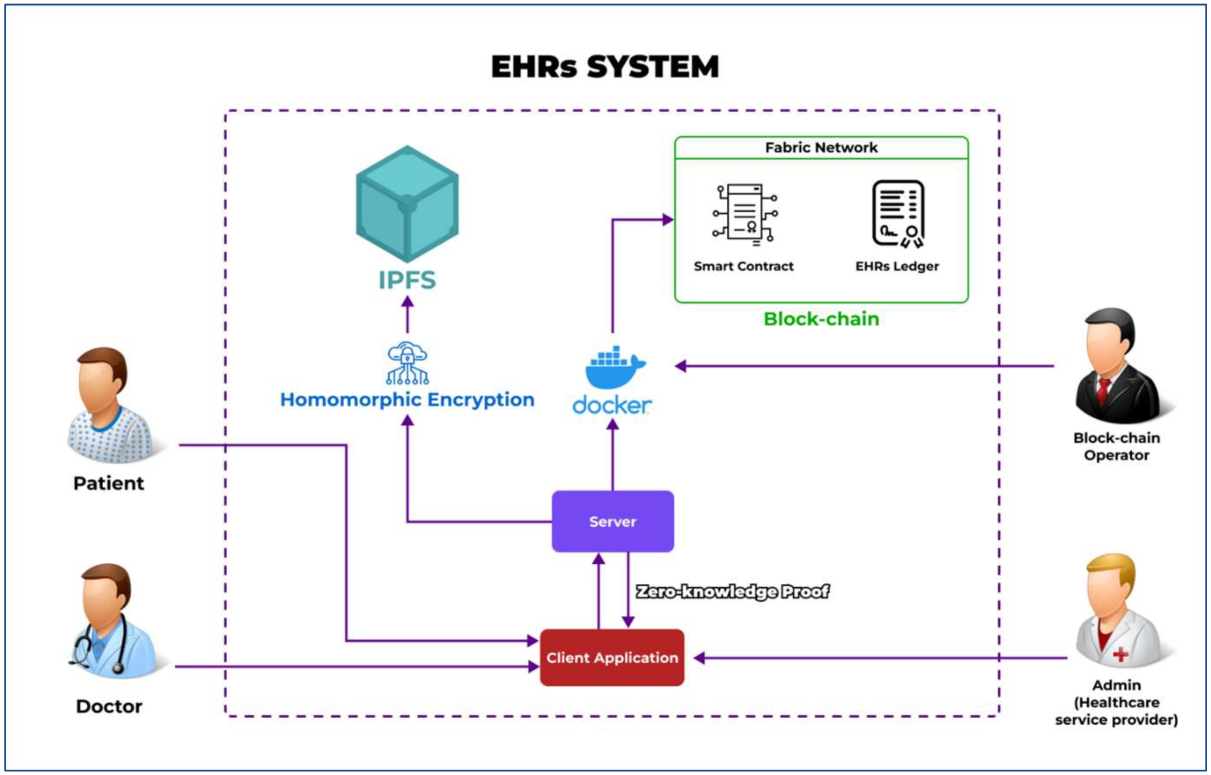
Stage two: A multi-class classifier is built using EfficientNet and SVM to accurately predict labels for each cell



Proposed Method

2. Algorithms

Blockchain-based EHRs storage system



Output/Outcome

3. Results

Clinical diagnosis

Authors	Models used	Maximum accuracy (%)
Mir et al.	Naive Bayes, SVM,Random Forest and Simple CART	79.13
Vijayarani et al.	SVM	79.66
Mohan et al.	HRFLM	88.4
Sriram et al.	Random forest	90.26
Rinkal Keniya et al.	Coarse, Medium, Fine and Weighted KNN;Gaussian Naive Bayes;Kernel Naive Bayes;Coarse, Medium and Fine Decision Tree;SubSpace KNN;RUSBoost algorithm	93.5
Our proposed method	Naïve Bayes Decision Trees Random Forest	**98.3**

Output/Outcome

3. Results

Chest X-ray diagnosis

Model	F1 – score	Best confidence threshold	mAP@0.5
CheXNet	0.435		
Faster R-CNN	0.62		
VGG16			0.4179
DetNet59			0.4751
YOLOv5	0.76	0.203	0.8
YOLOv5 + ResNet50	**0.77**	**0.213**	**0.812**

Output/Outcome

3. Results

Malaria diagnosis

Class	Precision	Recall	F1-score
Gametocyte	**1.00**	0.91	0.95
Leukocyte	**1.00**	0.93	0.96
Ring	**1.00**	0.96	**0.98**
Schizont	0.99	0.93	0.96
Trophozoite	0.97	**1.00**	**0.98**

Output/Outcome

3. Results

Malaria diagnosis

Model \ Metrics	mAP@.50	mAP@[.5:.95]
Yolov5	~0.33	~0.26
Faster R-CNN	0.65	
Proposed approach (Faster R-CNN + EfficientNet-SVM)	0.887	0.65

Title : Towards a timely and accurate AI-based Computer-Aided Diagnosis approach in healthcare

Thank you so much for your attention.