

Received 21 July 2023, accepted 5 August 2023, date of publication 10 August 2023, date of current version 17 August 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3303966

## RESEARCH ARTICLE

# ICNN-Ensemble: An Improved Convolutional Neural Network Ensemble Model for Medical Image Classification

JAVOKHIR MUSAEV<sup>1</sup>, ABDULAZIZ ANORBOEV<sup>1</sup>, YEONG-SEOK SEO<sup>1</sup>, (Member, IEEE),  
NGOC THANH NGUYEN<sup>2,3</sup>, (Senior Member, IEEE), AND DOSAM HWANG<sup>1</sup>

<sup>1</sup>Department of Computer Engineering, Yeungnam University, Gyeongsan 38541, South Korea

<sup>2</sup>Faculty of Information and Communication Technology, Wroclaw University of Science and Technology, 50-370 Wroclaw, Poland

<sup>3</sup>Faculty of Information Technology, Nguyen Tat Thanh University, Ho Chi Minh City 70000, Vietnam

Corresponding author: Yeong-Seok Seo (ysseo@yu.ac.kr)

This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea Government (MSIT) (No. NRF-2023R1A2C1008134). Also, this work was supported by the 2017 Yeungnam University Research Grant.

**ABSTRACT** Deep learning (DL) classification has become a major research topic in the areas of cancer prediction, image cell classification, and image classification in medicine. Furthermore, DL classification is the core of other subfields. Owing to various forms of ensemble models, DL models have achieved state-of-the-art performances in fields such as medicine. However, the existing models cannot solve the problem of generalization perfectly and proposed solutions only for tasks with specific datasets. Most state-of-the-art classification models presented their results in ImageNet dataset, and models elaborate on the insights of the dataset. Nonetheless, model architectures or pretrained models cannot provide the same accurate results for datasets with different classes than ImageNet. Hence, this research proposes an improved convolutional neural network ensemble (ICNN-Ensemble) based on the representation of high-resolution image channels (RHRIC) and a systematic model dropout ensemble (SMDE). ICNN-Ensemble exploits image channels after applying RHRIC and RGB images in their original forms, which accesses more residual feature connections and represents more insight into image channels. Furthermore, SMDE is applied to choose ensemble members, considering the changes of the accurate prediction field (APF) in the ICNN-Ensemble model. In addition, the proposed model executes ensembling during the test set prediction, which allows the model to be trained with larger batches and images compared to ensemble model's final results during training, allowing maximal effective usage of the graphics processing unit (GPU). Despite the small size of the model, the results of benchmarking for Malaria cell images dataset clearly illustrated that the ICNN-Ensemble model achieved significantly more accurate results than other base state-of-the-art models.

**INDEX TERMS** Representation of high-resolution image channels, systematic model dropout, accurate prediction field.

## I. INTRODUCTION

Image classification is a deep learning (DL) computer vision task that specialized in categorizing input images into pre-defined image classes by obtaining and representing insights of the images beneficial for further preprocessing. Distinct

The associate editor coordinating the review of this manuscript and approving it for publication was Donato Impedovo<sup>1</sup>.

from image segmentation and detection tasks with means of classification [1], [2], the image classification problem is a considerable part of all DL tasks. Medical image classification task is a crucial part of DL in which a small error may lead to colossal damage to human life; therefore, any improvements serve as a contribution to future advances in the field. Numerous classification models have real-life applications in astronomical research [3], the food industry,

agriculture, and farming [4], [5], [6]. The most beneficial and noteworthy application of image classification is in the field of medicine, where it has been applied in clinical treatment and medical diagnosis [7]. Currently, the main application of image classification is in diagnosis. Its real-life implementation requires very accurate results to make a clear diagnosis of the disease. The role of DL image classification is to save time for feature extraction and classification. However, most current DL models propose solutions for specific datasets that are outside the scope of medicine. This creates a huge gap in DL image classification for medical image datasets. Furthermore, the application and development of a combination of new and previous methods has become a secondary research interest in the field. These factors encouraged us to conduct research on improvements in medical image classification. The main goal of this study was to provide an accurate solution for medical image classification by proposing a more effective and accurate model. In this study, an improved convolutional neural network ensemble (ICNN-Ensemble) model was proposed to classify the Malaria cell images dataset; the model managed to achieve the most accurate results with respect to other baseline models. From the early steps of DL evolution, the classification task was the basis for all the other subfields of DL.

To increase the probability of a high number of true predictions, systematic model dropout ensemble (SMDE) requires the difference of the accuracy of the ensemble members to be in an interval of 3%, which leads to a decrease in the error rate of the model. In the proposed model, four ensemble members were used prior to the SMDE. The SMDE, which filters models according to SMDE rules, was developed considering most of the factors in ensemble learning that decrease the error rate in the predictions. A more detailed introduction of the SMDE is presented in Section III. Each step of this work presented valuable tools to develop an accurate DL classification model for medical image classification. In this work, SMDE was developed, provided a detailed view of the accurate prediction field (APF), and used the representation of high-resolution image channels (RHRIC) to increase the number of training images, which resulted in 99.67% accuracy for the Malaria cell images dataset. The result was advantageous differ from baseline models as long as it reached too close to the perfect accuracy. Moreover, parts of the proposed model with demonstrated effect can be used in different tasks. This work brings the following contributions to medical image classification area:

- 1) An increased amount of training data was obtained using representation of high-resolution image channels (RHRIC). Models were introduced with four times more data than baseline models, which led to learning more insights of the images than using RGB images only.
- 2) A clear understanding of the behavior of the accurate prediction field (APF) during improvements in model and data representation. A detailed presentation of the

APF allowed the manipulation of the data to achieve better results.

- 3) Using properties of systematic model dropout ensemble (SMDE) to increase APF. Another crucial part of the work is selecting the best possible combination of ensembles to avoid the loss of APF. False predictions of each model can transform the predictions of other ensemble members into false predictions.
- 4) More accurate classification of the dataset. By combining all the abovementioned tools and methods, better accuracy was achieved than the baseline models that were pretrained with ImageNet.

Overview of the proposed method and short presentation of the results of the proposed model was presented in Section I. Section II included representation of the main ideas of the related works, initial motivation and the gaps in the researches which lead to conduct the research. Section III illustrated the detailed introduction of the proposed method including comprehensive tools of the proposed method. Furthermore, each part of the proposed method was illustrated graphically. In-depth analysis of the experiments and results was covered in Section IV. In addition, evaluation metrics, dataset for training and testing, baseline model, training setup were introduced in this section. The last section gave conclusion to the proposed method and conducted research.

## II. RELATED WORK

Medical image analysis in DL is becoming a hot topic of state-of-the-art research; depending on the research problem, it primarily uses image segmentation and detection models in addition to image classification. In recent years, many machine learning [8], [9], and DL models [10], [11], including convolutional neural networks (CNNs) [12], [13], and recurrent neural networks, have been evaluated in image classification tasks. To address the limitations of conventional classification approaches, [14] proposed a novel ensemble learning paradigm for medical diagnosis with imbalanced data, which consisted of three phases: data preprocessing, training base classifier, and final ensemble. Visual transformers outperformed the residual neural network (ResNet) in terms of accuracy on ImageNet. In addition, in [15], the authors introduced a multilayer perceptron (MLP)-mixer that proved that CNNs or an attention-based models are not necessary. The MLP-mixer also used image patches as an input but included two types of layers: one MLP applied independently to each patch and another MLP applied across image patches. [7] used feature fusion to merge data from sensor data and electronic medical records, and only relevant and meaningful data was chosen for further training with the DL ensemble model. Modern healthcare systems are likely to be major designers and purchasers of DL goods. Qummar et al. [16] employed pretrained Resnet50, Inceptionv3, Xception, Dense121, and Dense169 models and ensembled them to achieve the best results in the field for diabetic retinopathy identification. Unlike existing models, the suggested ensemble model could predict all stages of

diabetic retinopathy and outperformed them. The ensemble of CNNs and deep residual neural networks is an effective tool for hyperspectral image classification, outperforming other state-of-the-art models [10]. Depending on the quality of the dataset, each of the aforementioned models has distinct advantages. Unbalanced, complex, high-dimensional, and noisy data are common research obstacles. Ensemble learning [17] is a useful approach for addressing these issues. Various methods, such as unified data fusion, modeling, and mining, have been developed to help build models more effectively and get better outcomes. The ultimate goal of all models was knowledge discovery and better predicting results. Existing ensemble learning approaches were divided into four categories to make it easier to describe ensemble models: supervised ensemble classifications, semi-supervised ensemble classifications, clustering ensembles, and semi-supervised clustering ensembles [18]. Reference [19] presented multi-label learning to improve the results of drug-target interaction tasks performed with binary classification models. Researchers demonstrated excellent results in this work by facilitating multi-label categorization with a community detection strategy for drug-target interactions. The semi-supervised rotation forest technique, proposed by Lu et al. [20], is an improved rotation forest algorithm that uses semi-supervised local discriminant analysis as a feature rotation tool. The gained knowledge of the discriminative and local structural information of small labeled and big unlabeled data samples was the cause for superior results utilizing this technique. It was demonstrated that [21] investigated the effect of image size on model accuracy. In addition, many other researches [23], [24], [25] focused on providing solution to the problems in the medicine using DL tools and ensemble learning techniques. Reference [24] used EfficientNetB0 architecture to prove that the architecture can serve as reliable model to predict malaria cells in during blood cells analysis. There are many ensemble strategies and methods [26] that are used in developing DL models. Unweighted Model Averaging, Majority Voting, Bayes Optimal Classifier, Stacked Generalization, Super Learner, Consensus and Query-by-Committee strategies are used for most of the models. Considering the importance of the generalization, state-of-the-art fuzzy ensemble models were proposed in various applications of the AI. A fuzzy rank-based ensemble of CNN models for classification of cervical cytology [27] was proposed which ensembled three models during the test set evaluation. Reference [28] conducted research on human activity recognition using nonlinear fuzzy ensemble of deep neural networks. The developed model's ensemble part consisted of implementation of fuzzy score generation from Mitscherlich, Blumberg and Weilbull functions, calculation of gross rank score and gross confidence score, and rewarding of classes. The model reached to 100% accuracy overcoming all the baseline models for MHEALTH dataset. Another research [29] presented ensemble of Inception V3, VGG19, and MobileNet while using Sugeno integral method

for COVID-19 detection from X-ray images. Reference [30] proposed a concise Takagi-Sugeno-Kang fuzzy ensemble classifier for high-dimensional problems. The study integrated improved bagging and dropout which improved the trade-off between accuracy and interpretability in TSK fuzzy systems. Also, the study transformed the high-dimensional feature space into a series of low-dimensional feature random subsets. Another research [31] presented ensemble of four different pretrained models which applied three non-linear functions for the prediction scores of the ensemble members. Newly published surveys [32], [33], [34], [35] shows that heterogeneous ensemble learning models exploding on various subfields of artificial intelligence. Turbidity in water treatment plant is quit important task for different environmental and public health perspectives. Reference [36] developed neuro-fuzzy approach using two different optimizations of fuzzy inference system (FIS). As a result, proposed model increased the accuracy up 15% in test set. Reference [37] proved that using fuzzy aggregation based ensemble can help to reach accurate and stable feature selection. The framework consist of three main parts: distribution generation of feature importance, distribution ensemble using fuzzy aggregation, and defuzzification for feature ranking. To detect broken rotor bar in nonstationary environment [38] introduced an ensemble-based fuzzy rough active learning approach which includes an ensemble framework, fuzzy rough active learning, and drift detection. Another study [39] applied DL classification to predict mental state of the person using speech features. Authors used variety of the DL models, including the Naive Bayes Network, SOTA algorithm, Fuzzy Rule, Probabilistic Neural Network, Decision Tree, Gradient Boosted Tree, Random Forest, Tree Ensemble, and an ensemble approach. Results of the experiments showed that ensemble approach obtained higher accuracy than most of the single models. It was stated that the ranking of the methods depend on the choice of the training set and evaluation metric. Heterogeneous ensemble models are being used in numerous fields, including medicine. Especially, diagnosis of post brain surgery injuries is considered as arduous task by most of the professionals in the field. There numerous researches that specialized in detecting, classifying diseases in different parts of human body. One of the important part of the human body is kidney that is in the study center of AI tools such as ensemble learning. References [40], [41], [42], [43], [44], and [45] applied different ensemble methods to identify the diseases in the kidney. Reference [40] used C5.0, support vector machine, Bayes, and XGBoost algorithms to develop ensemble model and applied it to early prediction of acute kidney injury in patients admitted to the neurological intensive care unit. The other research [41] focused on detecting chronic kidney disease which is considered as one of the most serous health problem in the world. The study used dataset with 24 features and feed it to the AdaBoosting ensemble which lead to 99% at early kidney disease detection. The same field was studied in [42]. Homogeneous ensemble method

Model: "sequential"

Layer (type)	Output Shape	Param #
conv2d (Conv2D)	(None, 62, 62, 32)	896
resizing (Resizing)	(None, 32, 32, 32)	0
batch_normalization (BatchNormalization)	(None, 32, 32, 32)	128
max_pooling2d (MaxPooling2D)	(None, 16, 16, 32)	0
conv2d_1 (Conv2D)	(None, 14, 14, 64)	18496
batch_normalization_1 (BatchNormalization)	(None, 14, 14, 64)	256
max_pooling2d_1 (MaxPooling2D)	(None, 7, 7, 64)	0
conv2d_2 (Conv2D)	(None, 5, 5, 128)	73856
batch_normalization_2 (BatchNormalization)	(None, 5, 5, 128)	512
max_pooling2d_2 (MaxPooling2D)	(None, 2, 2, 128)	0
flatten (Flatten)	(None, 512)	0
dense (Dense)	(None, 256)	131328
dense_1 (Dense)	(None, 128)	32896
dense_2 (Dense)	(None, 2)	258

=====  
 Total params: 258,626  
 Trainable params: 258,178  
 Non-trainable params: 448

FIGURE 1. Convolutional neural network model architecture.

applied in this research. The authors used weighted average to ensemble the models for computer tomography images to detect the disease. Despite advances of homogeneous ensemble methods, there many heterogeneous ensemble methods that overcome many homogeneous ensemble or sing models. Reference [46] presented effectiveness of the heterogeneous ensemble model for predicting protein-DNA binding affinity. The method compared with six different methods and obtained highest, 0.84%, average correlation coefficient among them.

Reviewing the aforementioned research, a gap was found in that none of the studies considered the effect of image resolution and model dropout during ensemble could create effective tools to add more power to the classifier. This motivated us to develop a new model for medical image segmentation.

III. PROPOSED METHOD

To address the research gap found in the literature review, the ICNN-Ensemble model was proposed which effectively managed the training and achieved the best results on the dataset. Using images only in the RGB form is a classic form of training, in which other tools and methods were included to obtain more insights from the training data. This study incorporated the advantages of many studies and added effective tools to understand, evaluate, and manipulate the training and testing data of the ICNN-Ensemble model. The research focused on developing a general DL

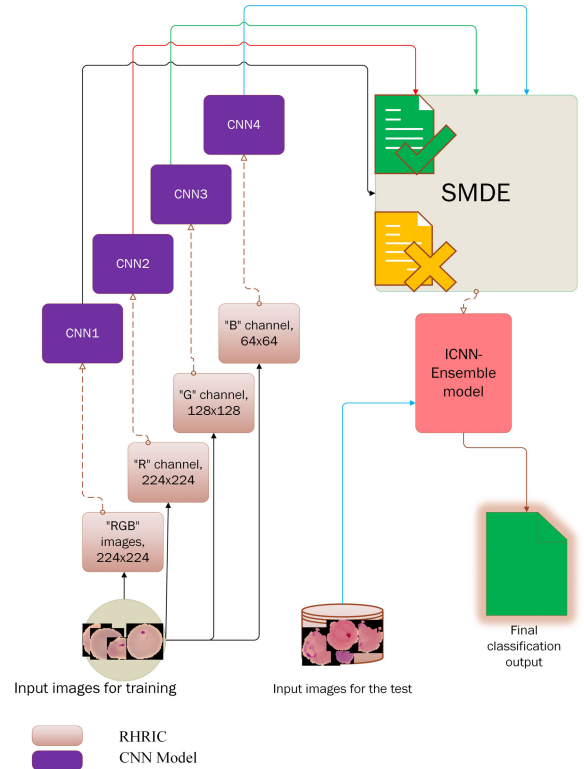


FIGURE 2. Graphical illustration of proposed methodology. (Images from Malaria cell images dataset were used to build the figure.)

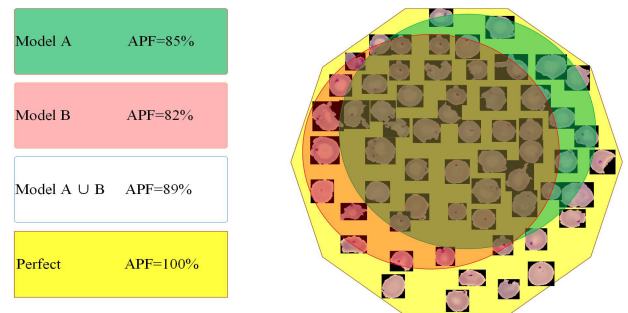
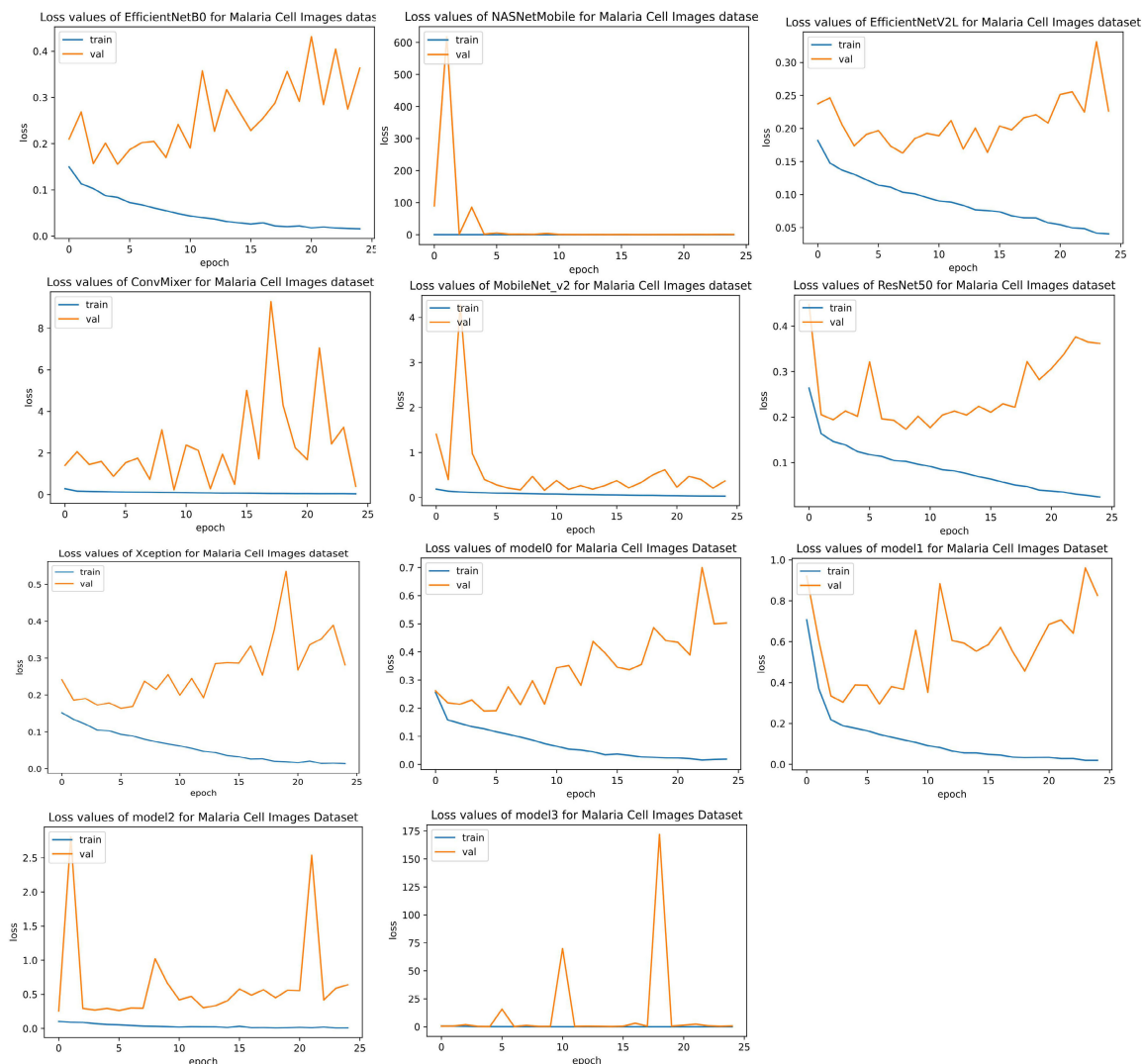


FIGURE 3. Illustration of the effect of unique true prediction (UTP), using Model A and Model B, on accurate prediction field (APF). (Images from malaria cell images dataset were used to build the figure.)

classification model that should present better results than most widely used pretrained models. The ICNN-Ensemble consists of RHRIC, CNN model architecture, and SMDE. In this section, detailed analysis and brief explanation were provided for each of them. Another important point of this work is the presentation of the reasons that explained why this model achieved better results than baseline models' results. The first step included preparing RGB images for the first model and applying RHRIC for the remaining three models. The RHRIC step consist of separating the image channels and creating three different datasets with only one image channel out of the main dataset. Different image sizes presented different features to the model, which led to preferred



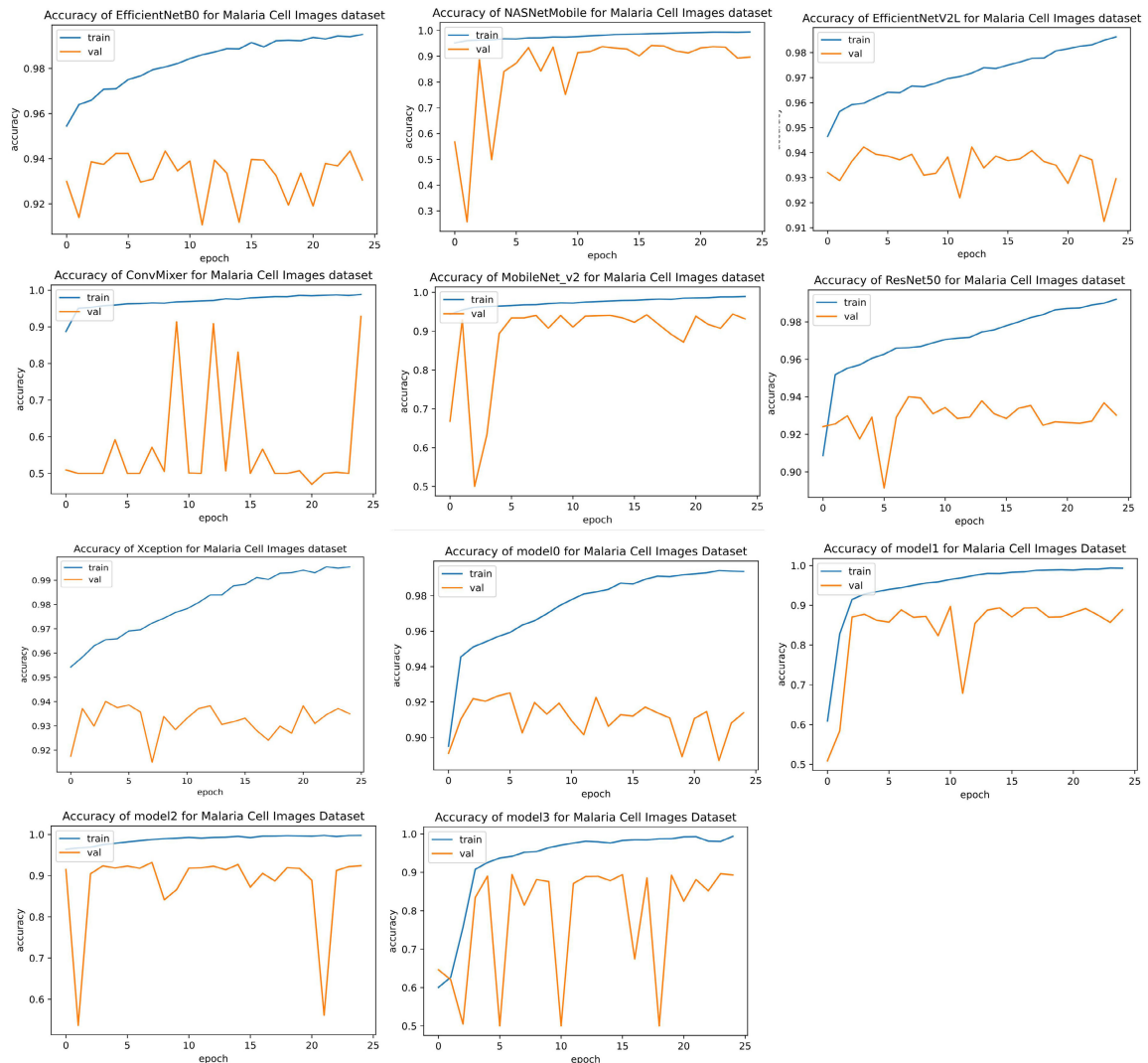
**FIGURE 4.** Loss values of baseline models and ensemble members during the training and validation.

results for the datasets. In the study, considering the positive effect of image resolution, each one-channel image was resized. Image channel sizes were changed to  $224 \times 224$ ,  $128 \times 128$ , and  $64 \times 64$  for the R, G, and B channels, respectively. The ICNN-Ensemble model includes the following steps:

- Preparing data and resizing them. In the experiments, images of the main datasets were called, resized to  $224 \times 224 \times 3$ , and re-scaled dividing each pixel by 255. This enabled a comparison of the dataset’s final classification results with the pretrained model classification.
- Preparing the R channel of the images for the second CNN model. Each image was called to the training environment, resized to  $224 \times 224 \times 3$ , and re-scaled dividing each pixel by 255. Next, out of three channels only the first R channel was chosen for the training. The final image for the training had  $224 \times 224 \times 1$  shape or in a tensor form; `image[:, :, 0]` which presents image with only the first channel. R channel images with

$224 \times 224 \times 1$  shape were saved for the training in the next steps of the training.

- Preprocessing the G channel of the image. In this step, the images were resized to  $128 \times 128 \times 3$ , rescaled each image pixel into  $[0,1]$  interval, chosen only G channel of the image by using `image[:, :, 1]` tensor which finally gave  $128 \times 128 \times 1$  image tensor for the training and saved for the future usage in the training.
- Preparing the B image channels for the training. Initially, the original images from the dataset was called and resized to  $64 \times 64 \times 3$ . In the next step, each image pixel were rescaled into  $[0,1]$  interval, chosen only B channel of the image by using `image[:, :, 2]` tensor which finally gave  $64 \times 64 \times 1$  image tensor for the training and saved for the future usage in the training.
- The next step was to feed the outputs of Steps 1, 2,3, and 4 into the CNN models. The previous steps represent the RHRIC, in which various image sizes were. In this step, all the outputs of the RHRIC were separately fed



**FIGURE 5.** Accuracy of baseline models and ensemble members during the training and validation.

into CNN models. The CNN models have the same architecture as that described in Fig. 1. To distinguish the input data, CNN models were designated as CNN1, CNN2, CNN3, and CNN4, which used the main dataset, R channel dataset, G channel dataset, and B channel dataset, respectively. All datasets were trained.

- The SMDE was applied to the models trained on four datasets. The SMDE checked for the following conditions:
  - 1) Did the model reach the highest accuracy?
  - 2) Is it the main model that trained with main dataset?
  - 3) b:=True: Is the accuracy of the model 4% less than accuracy of the best model?
  - 4) b:=False: Is the accuracy of the model 3% less than accuracy of the best model?
- After checking the models for the SMDE conditions, only those that satisfied SMDE conditions, were used to develop the ensemble model. In this step, the optimal models for ensembling were selected.

- The outputs of the previous step were ensembled using appropriate prediction probabilities for the test. Let the predictions for each dataset be denoted as  $P_0, P_1, P_2,$  and  $P_3$ .  $M_i$  represents  $i$ th image of the test set. For each of  $i$ th image in the dataset, the prediction was calculated and added them in the following form:  $P_0(M_i) + P_1(M_i) + P_2(M_i) + P_3(M_i)$ .
- $P_0$  was used to calculate the accuracy of the final step. The shape of  $P_0$  was  $2754 \times 2$ , for 2754 images and two classes.

This method was developed considering various factors that help increase the APF and decrease errors in predictions. Fig. 2 shows the architecture of the proposed model, where four different inputs and training data of different sizes were used for each CNN model. As described previously, the images were separated into different channels and used for training. The architecture of the CNN model is shown in Fig. 1. Despite the simplicity of the CNN model and the small number of training parameters, the model managed to

transfer large amounts of insights from images into the final classification layer for accurate classification. The rectified linear unit (ReLU), softmax activation functions and max pooling were used. Cross-entropy loss was used for the final loss evaluation.

$$M(b) = \frac{\gamma(b - \text{mean}(b))}{\sqrt{\text{var}(b) + \epsilon + \beta}}. \quad (1)$$

The goal of batch normalization is to keep the mean output close to zero and the standard deviation close to 1. Equation (1) presents batch normalization as a function where the input is a batch ( $b$ ), and the output is the normalized form of the input.  $\gamma$ ,  $\beta$ , and  $\epsilon$  are the small constant, learned offset factor, and configurable constant, respectively.

$$\text{ReLU} = f(a) = \max(0, a). \quad (2)$$

$$\sigma(z)_i = \frac{e^{z_i}}{\sum_{j=1}^K e^{z_j}} \quad (3)$$

Except for the final layer, all CNN model layers used the ReLU activation function. The softmax function transforms inputs into a probability distribution, whereas the ReLU function returns the positive part of its inputs. Softmax function (3) normalizes the vector  $z$  of  $K$  class predictions into  $K$  probabilities. The softmax function (3) takes the vector  $z$  of the  $K$  class predictions and normalizes it into  $K$  probabilities. The softmax function  $\delta : R^K \rightarrow (0, 1)^K$ ,  $i = 1, \dots, K$  and  $z = (z_1, \dots, z_K) \in R^K$  whereas  $K$  is greater than one.

$$\text{Cross entropy loss} = \sum_{i=1}^v y_i \log p_i \quad (4)$$

In (4),  $v$  is the number of classes,  $y$  is the label of the input, and  $p$  denotes the predicted values for the input.

An important part of this study is to understand the APF and its effect on the model. The APF is the set of all true predicted images and their indexes. If 10000 images were fed to the model and obtained the prediction for the input, all true predicted images and their indexes are chosen as APF. The APF is affected by the unique true predictions (UTP) of each model during ensembling. As given in (5),  $X$  is the prediction scope of model  $X$ , and  $Y$  is the prediction scope of model  $Y$ .

$$\text{UTP}(X, Y) = X - X \cap Y \quad (5)$$

Increasing the number of accurate predictions was the main goal of this study. To achieve this, the study focused on understanding the behavior of the APF and manipulated it using UTP. Initially, each image was marked with unique indexes, for 10000 images, numbers from 1 to 10000 were used as images' indexes. The UTP represents a unique prediction for each model with respect to the other. For instance, model1 can truly predict the label of the images with indexes [101-2000] when model2 can give true predictions for images with indexes [1001-10000]. In this example model1 has 19% accuracy and model2 reached 90% accuracy. When UTP is analyzed, model1 has advantage on predicting images from

index 101 to index 2000 and model2 cannot give true predictions for images with indexes [101-2000]. In this example,  $\text{UTP}(\text{model1}, \text{model2})=100\%$  which was explained with all true predicted images of model1 wasn't truly predicted by model2.

Every model can include different true predicted images from the same dataset. Fig. 3 illustrates the effect of UTP on the APF. True predictions of each model can differ as they represent different insights. To use this as a tool to enlarge the APF, models were ensembled and analyzed their UTP. If there were two Models A and B, and Model A had 95% accuracy, and Model B had 94% accuracy, then Model A can predict parasite image malaria cells more accurately than Model B, and Model B can predict uninfected malaria image cells better than model A. When the models were combined, their prediction powers were combined as well. This allowed us to predict both types of cells better than using only one model. Another problem in ensemble models is the effect of errors on the predictions of the ensemble. Existing features that lead to incorrect predictions in the models can also grow after ensembling them. To address this problem, SMDE was developed. This carefully selects models for ensembling and filters them based on specific requirements. Another advantage of this work is that the models were trained with the outputs of Steps 1,2,3, and 4. Subsequently, all training processes were stopped and only testing included SMDE and resulted in the ICNN-Ensemble model. Therefore, ICNN-Ensemble achieved better results than expected and added extra knowledge to the ensemble, decreasing the effect of errors on predictions. Studying the behavior of the UTP allowed us to manage the APF of the ensemble model, which provided more opportunity to modify the model and obtain better classification than with most pretrained models.

#### IV. EXPERIMENTS AND RESULTS

In this study, the Malaria cell images dataset was used to evaluate the proposed model. In this part of the study, accuracy and UTP metrics were chosen to evaluate the experiments. In this section, a comprehensive analysis of the experiments and the results is provided. Furthermore, the section includes the baseline models, training setup, and a detailed presentation of the results.

##### A. DATASET

The Malaria cell images dataset<sup>1</sup> is a dataset in the medical image classification field that provides valuable data for DL classification and is considered one of the best datasets to prove the effectiveness of medical image classification models. Considering these factors, the dataset was used. The collection contains 27558 infected and uninfected pictures. The dataset is balanced and includes photographs of various sizes, rather than images with normalized sizes. To conduct the study, the photos were enlarged to 224 224 and rescaled by 255. For the test set, 2000 photos were chosen, and

<sup>1</sup><https://lhncbc.nlm.nih.gov/LHC-downloads/dataset.html>

the remaining 25558 images were utilized for the training and validation sets; the validation set was 10%. The dataset of 27558 photos was divided into three separate datasets. Each of the three datasets constructed featured only one picture channel, deleting the other two channels from the 27558 photos. The first dataset was constructed using R channels from the 27558 photographs, the second using G channels from the 27558 images, and the third using B channels from the 27558 images.

**B. BASELINE MODEL**

In this study, pretrained EfficientNetB0, NASNetMobile, EfficientNetV2L, MobileNetV2, ResNet50 and Xception models, and ConvMixer were used to compare the results and analyze the impact of the proposed method. These models are the most used in DL classification tasks because they represent pretrained features of the ImageNet dataset and easily reach accurate predictions. The main goal was to obtain a higher accuracy than that of these models.

**C. TRAINING SETUP**

Python 3.6.12 and TensorFlow 2.1.0 were used to develop the CNN model architecture for the suggested method’s training. The trials were carried out using a 12 GB GeForce RTX 3080 Ti with CUDA 10.2 on a PC with an Intel core-i9 11900F CPU and 64 GB of RAM. The model’s weights were randomly initialized and trained for specific epochs during training. The Adam optimizer with a default learning rate of 0.001 and the cross entropy loss function were used to train the ICNN-Ensemble. The model was trained using the Malaria image cell dataset.

**D. EVALUATION METRICS**

To present comparable and useful results, the model was evaluated using accuracy, precision, recall and f1 score. The metrics were mainly chosen to meaningfully explain the method’s achievements for the dataset. UTP was used, as described in Section III, and accuracy metrics. The main goal of this study is to achieve as high an accuracy as possible with minimal loss of image knowledge during training. The accuracy was defined as the ratio of the number of true predictions to the total number of cases used to evaluate the model.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{6}$$

$$Precision = \frac{TP}{TP + FP} \tag{7}$$

$$Recall = \frac{TP}{TP + FN} \tag{8}$$

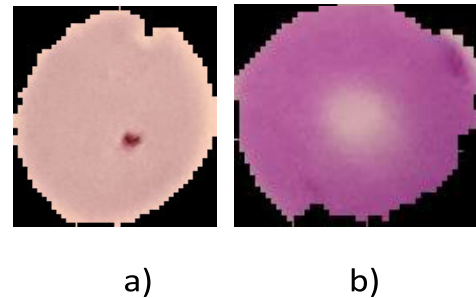
$$F1\ score = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{9}$$

TP represents true predicted positive outcomes, TN represents true predicted negative results, FP represents false predicted positive results, and FN represents false predicted negative results.

$$UTP(X, Y) = X - X \cap Y \tag{10}$$

**TABLE 1. UTP of the models trained on “R”, “G”, “B” image channels and main dataset (MD).**

	MD (CNN1)	R (CNN2)	G (CNN3)	B (CNN4)
MD (CNN1)	0.0	0.03	0.023	0.024
R (CNN2)	0.004	0.0	0.005	0.005
G (CNN3)	0.023	0.03	0.0	0.024
B (CNN4)	0.073	0.08	0.072	0.0



**FIGURE 6. Examples of test samples from the Malaria cell images dataset. a and b belong to “Infected” and “Uninfected” classes, respectively.**

To present detailed explanation of the reasons and tools of the research success, UTP was used. X and Y represents two different models’ APF by including index of the inputs. The exact role of the model in the field can be determine by calculating the accuracy. F1, precision and recall scores helps to find other gaps and includes practical insights to continue the research for further developments.

**E. EXPERIMENTAL RESULTS AND DISCUSSIONS**

Our research primarily focused on improving accuracy and maintaining a small model size. The baseline models have millions of parameters that require a huge amount of time and computation for training and testing. To propose an affective light model with the highest accuracy, the APF was analyzed and UTP of the models used for the ICNN-Ensemble. As presented in Table 1, each model trained on different channels of the image had UTPs that differed from the other models. These differences were very small. When they were ensembled, the main part of the unique prediction delivered to the ensemble model.

**1) ERROR ANALYSIS**

Fig. 6 presents some examples from the Malaria cell images dataset where most of the base classifiers made wrong predictions on the sample, but the proposed ensemble model made correct prediction. Fig. 6a is a sample from the “Infected” class of the dataset, which is classified as “Uninfected” with 91% confidence by ICNN1, “Uninfected” with 88% confidence by ICNN2, “Infected” with 100% confidence by ICNN3, and “Infected” with 99.5% accuracy ICNN4. Despite the low confidence and wrong predictions, SMDE find a way to ensemble the ensemble members to reach true prediction label with 99.75% confidence. Effective selection of ensemble members presented its advance in Fig. 6b.



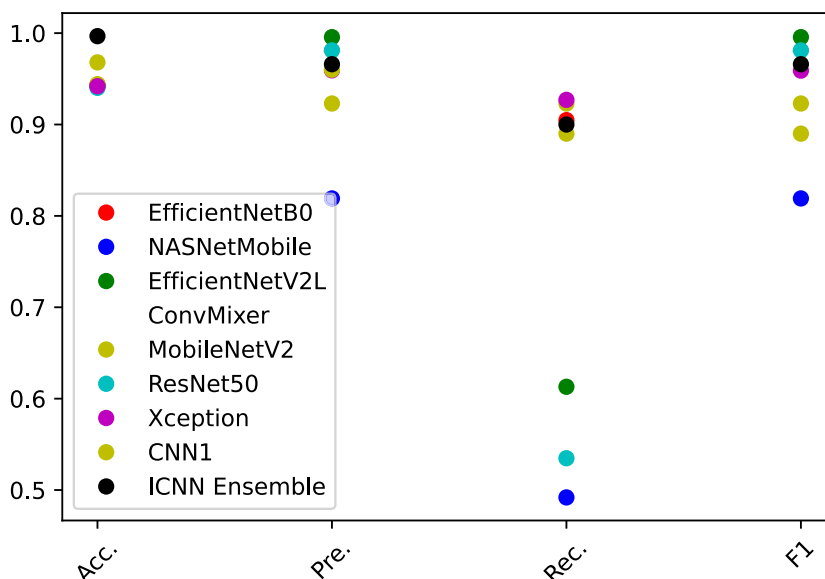


FIGURE 7. Test set evaluation of the baseline and proposed models using Malaria cell images dataset. Acc., Pre., Rec., and F1 represents accuracy, precision, recall, and f1 scores, respectively.

TABLE 2. Test set accuracy of the models using malaria cell images dataset.

Models	Acc. (%)	Pre.	Rec.	F1	Param. (mln)
EfficientNetB0	94.34	0.9803	0.9048	0.9803	5.3
NASNetMobile	94.19	0.8191	0.4919	0.8191	5.3
EfficientNetV2L	94.23	0.9956	0.613	<b>0.9956</b>	119
ConvMixer	92.81	0.977	0.889	0.977	0.6
MobileNetV2	94.41	0.923	0.923	0.923	4.3
ResNet50	94.01	<b>0.9811</b>	0.5348	0.9811	25.6
Xception	94.19	0.9593	0.927	0.959	22.9
CNN1	96.8	0.96	0.89	0.89	<b>0.26</b>
CNN2	99.4	0.94	0.94	0.9399	<b>0.26</b>
CNN3	96.8	0.9654	<b>0.9652</b>	0.9653	<b>0.26</b>
CNN4	91.93	0.9619	0.9619	0.9619	<b>0.26</b>
ICNN-Ensemble	<b>99.67</b>	0.966	0.90	0.966	<b>0.52</b>

The image was taken from the “Uninfected” images in the dataset. Nevertheless, ICNN3 classified it as “Infected” image with 100% confidence and ICNN4 classifies it as “Infected” with 98.1% confidence. But, the other two classifiers, ICNN1 and ICNN2, classified it as “Uninfected” with 93% and 99% confidence. The role of SMDE in accurate final predictions was demonstrated in this situation as well. The proposed ICNN-Ensemble model predicted this image as “Uninfected”.

2) STATISTICAL ANALYSIS

To analyze the proposed model, McNemar’s [47] statistical test was performed. McNemar’s test is a non-parametric analysis of the distribution of paired nominal data. The “p-value” represents the likelihood of two models being similar; consequently, a lower p-value is preferred. The p-value must be less than 5% to reject the null hypothesis that the two models are similar. That example, if p-value < 0.05, we may confidently conclude that the two models under study are statistically distinct. The result of the performed statistical

TABLE 3. Results of the McNemar’s test performed between the proposed ensemble model (ICNN-Ensemble) and the base learners used: null hypothesis is rejected for all cases.

Models	ICNN-Ensemble
CNN1	0.000
CNN2	0.000
CNN3	0.000
CNN4	0.000

analysis showed that all the base classifiers are different than ICNN-Ensemble model. Table 3 shows that p-value for all ensemble members are equal to zero when p-values were rounded to the nearest thousandth.

Combining the UTPs of each model led to a larger APF. The APF is the main source of achieving high accuracy in the model. All the baseline models achieved lower accuracy on the dataset; however, their accuracies were suitable for many fields. A classification accuracy of more than 90% is considered an effective model for implementation. Fig. 4 and Fig. 5 shows the loss and accuracy of the proposed and baseline models. Analyzing them it can be seen that, in some cases, there are overfitting and irregular trend of the accuracy. Despite it, proposed method applied and obtained marvelous results on test set. As shown in Table 2 and Fig. 7, proposed model achieved 99.67% accuracy, whereas the other models achieved an accuracy of approximately 95%. EfficientNetB0, NASNetMobile, EfficientNetV2L, MobileNetV2 and Xception reached around 94% accuracy on the given dataset. Despite having huge amount of parameters, the models cannot produce better outcomes than ICNN-ensemble. When true positive predictions were examined, ResNet50 provided the best prediction on true positives, with a precision score of 98%. When recall score of the models were studied, CNN3 presented the highest score outcoming all of the models. F1 score were calculated to obtain stronger behavior of the

TP, TN, FP, and FN. In this case, EfficientNetV2L with 119 mln trainable parameters reached the highest score. This can be described with the help of UTP and APF changes. All baseline models had an average of 10 times more trainable parameters. With a very small model, the number of inputs was increased and decreased incorrect predictions in the ensemble model. To achieve these results, the proposed model was divided into two-sections: training and testing. The training part continued only up to the SMDE part of the ICNN-Ensemble, while the test set went through all steps of the ICNN-Ensemble. Nevertheless, the test set evaluation required a smaller model than the training set because the ineffective ensemble members were excluded. In addition, only the best epoch from the training part of the model was used. All of these factors yielded expected goal of the study. Looking at the accuracies of the ensemble members, ICNN1, ICNN2, ICNN3, and ICNN4, it can be seen that their accuracies for fed input are lower than the accuracy of ICNN-Ensemble model. By contrast, when SMDE was applied, the results were improved. This can only be explained by the positive effect of UTP on the APF, which yielded the highest accuracy for the dataset.

## V. CONCLUSION

The study proved that using UTP and enlarging the APF can lead to a considerable change in the true predictions of the model. Systematic dropout and its rules (SMDE) to choose ensemble members were great tools for developing the ICNN-Ensemble model, which reached a final accuracy of 99.67%, outperforming all baseline models that were pre-trained with one million images. None of the baseline models could achieve the results of the ICNN-Ensemble, despite the fact that the ICNN-Ensemble had a much smaller size and was trained only with the Malaria cell images dataset. The dataset contains less than 30000 images, and training with this number of images does not create the effect of a large dataset, such as that with one million images. The obtained results can be accepted as an outcome of the ICNN-Ensemble architecture. There is still a gap that other research can address. It can be tried to decrease the model size and, at the same time, obtain better results on the dataset. In addition, a generalization problem remains to be addressed: the model was not checked on other datasets, and the tools in the ICNN-Ensemble should be generalized.

In the future work we plan to use collective intelligence for achieving higher accuracy of the DL classification-based method by means of using user opinions in social media. We will use ontology tools for building a social system and consensus methods for integrating the user opinions. This can help much in increasing the effectiveness of classification [48], [49], [50].

## REFERENCES

- [1] B. Olimov, S.-J. Koh, and J. Kim, "AEDCN-Net: Accurate and efficient deep convolutional neural network model for medical image segmentation," *IEEE Access*, vol. 9, pp. 154194–154203, 2021, doi: 10.1109/ACCESS.2021.3128607.
- [2] B. A. U. Olimov, K. C. Veluvolu, A. Paul, and J. Kim, "UzADL: Anomaly detection and localization using graph Laplacian matrix-based unsupervised learning method," *Comput. Ind. Eng.*, vol. 171, Sep. 2022, Art. no. 108313, doi: 10.1016/j.cie.2022.108313.
- [3] M. Pritt and G. Chern, "Satellite image classification with deep learning," in *Proc. IEEE Appl. Imag. Pattern Recognit. Workshop (AIPR)*, Oct. 2017, pp. 1–7, doi: 10.1109/AIPR.2017.8457969.
- [4] A. Martinazzo, M. Espadoto, and N. S. T. Hirata, "Self-supervised learning for astronomical image classification," in *Proc. 25th Int. Conf. Pattern Recognit. (ICPR)*, Jan. 2021, pp. 4176–4182, doi: 10.1109/ICPR48806.2021.9412911.
- [5] S. R. N. M. Ayyub and A. Manjramkar, "Fruit disease classification and identification using image processing," in *Proc. 3rd Int. Conf. Comput. Methodologies Commun. (ICCMC)*, Mar. 2019, pp. 754–758, doi: 10.1109/ICCMC.2019.8819789.
- [6] Y. Adhitya, S. W. Prakosa, M. Köppen, and J.-S. Leu, "Feature extraction for cocoa bean digital image classification prediction for smart farming application," *Agronomy*, vol. 10, no. 11, p. 1642, Oct. 2020, doi: 10.3390/agronomy10111642.
- [7] F. Ali, S. El-Sappagh, S. M. R. Islam, D. Kwak, A. Ali, M. Imran, and K.-S. Kwak, "A smart healthcare monitoring system for heart disease prediction based on ensemble deep learning and feature fusion," *Inf. Fusion*, vol. 63, pp. 208–222, Nov. 2020, doi: 10.1016/j.inffus.2020.06.008.
- [8] D. P. Gaikwad and R. C. Thool, "Intrusion detection system using bagging ensemble method of machine learning," in *Proc. Int. Conf. Comput. Commun. Control Autom.*, Feb. 2015, pp. 291–295, doi: 10.1109/ICCUA.2015.61.
- [9] T. Alqurashi and W. Wang, "Clustering ensemble method," *Int. J. Mach. Learn. Cybern.*, vol. 10, no. 6, pp. 1227–1246, Jun. 2019, doi: 10.1007/s13042-017-0756-7.
- [10] Y. Chen, Y. Wang, Y. Gu, X. He, P. Ghamisi, and X. Jia, "Deep learning ensemble for hyperspectral image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 6, pp. 1882–1897, Jun. 2019, doi: 10.1109/JSTARS.2019.2915259.
- [11] Z. Chen, J. Duan, L. Kang, and G. Qiu, "Class-imbalanced deep learning via a class-balanced ensemble," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 10, pp. 5626–5640, Oct. 2022, doi: 10.1109/TNNLS.2021.3071122.
- [12] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," 2015, *arXiv:1505.04597*.
- [13] R. Xin, J. Zhang, and Y. Shao, "Complex network classification with convolutional neural network," *Tsinghua Sci. Technol.*, vol. 25, no. 4, pp. 447–457, Aug. 2020, doi: 10.26599/TST.2019.9010055.
- [14] N. Liu, X. Li, E. Qi, M. Xu, L. Li, and B. Gao, "A novel ensemble learning paradigm for medical diagnosis with imbalanced data," *IEEE Access*, vol. 8, pp. 171263–171280, 2020, doi: 10.1109/ACCESS.2020.3014362.
- [15] I. Tolstikhin, N. Houlsby, A. Kolesnikov, L. Beyer, X. Zhai, T. Unterthiner, J. Yung, A. Steiner, D. Keysers, J. Uszkoreit, M. Lucic, and A. Dosovitskiy, "MLP-mixer: An all-MLP architecture for vision," 2021, *arXiv:2105.01601*.
- [16] S. Qummar, F. G. Khan, S. Shah, A. Khan, S. Shamsirband, Z. U. Rehman, I. A. Khan, and W. Jadoon, "A deep learning ensemble approach for diabetic retinopathy detection," *IEEE Access*, vol. 7, pp. 150530–150539, 2019, doi: 10.1109/ACCESS.2019.2947484.
- [17] B. Krawczyk, L. L. Minku, J. Gama, J. Stefanowski, and M. Woźniak, "Ensemble learning for data stream analysis: A survey," *Inf. Fusion*, vol. 37, pp. 132–156, Sep. 2017, doi: 10.1016/j.inffus.2017.02.004.
- [18] Z. Yu, P. Luo, J. You, H.-S. Wong, H. Leung, S. Wu, J. Zhang, and G. Han, "Incremental semi-supervised clustering ensemble for high dimensional data clustering," *IEEE Trans. Knowl. Data Eng.*, vol. 28, no. 3, pp. 701–714, Mar. 2016, doi: 10.1109/TKDE.2015.2499200.
- [19] Y. Chu, X. Shan, T. Chen, M. Jiang, Y. Wang, Q. Wang, D. R. Salahub, Y. Xiong, and D.-Q. Wei, "DTI-MLCD: Predicting drug-target interactions using multi-label learning with community detection method," *Briefings Bioinf.*, vol. 22, no. 3, May 2021, Art. no. bbaa205, doi: 10.1093/bib/bbaa205.
- [20] X. Lu, J. Zhang, T. Li, and Y. Zhang, "Hyperspectral image classification based on semi-supervised rotation forest," *Remote Sens.*, vol. 9, no. 9, p. 924, Sep. 2017, doi: 10.3390/rs9090924.
- [21] C. F. Sabottke and B. M. Spieler, "The effect of image resolution on deep learning in radiography," *Radiology: Artif. Intell.*, vol. 2, no. 1, Jan. 2020, Art. no. e190015, doi: 10.1148/ryai.2019190015.
- [22] A. Fred Agarap, "Deep learning using rectified linear units (ReLU)," 2018, *arXiv:1803.08375*.

- [23] V. Gattulli, D. Impedovo, G. Pirlo, and G. Semeraro, "Early dementia identification: On the use of random handwriting strokes," in *Proc. Int. Graphonomics Conf.*, Jan. 2022, pp. 285–300, doi: [10.1007/978-3-031-19745-1\\_21](https://doi.org/10.1007/978-3-031-19745-1_21).
- [24] G. Marques, A. Ferreras, and I. de la Torre-Diez, "An ensemble-based approach for automated medical diagnosis of malaria using EfficientNet," *Multimedia Tools Appl.*, vol. 81, no. 19, pp. 28061–28078, Mar. 2022, doi: [10.1007/s11042-022-12624-6](https://doi.org/10.1007/s11042-022-12624-6).
- [25] A. R. Laeli, Z. Rustam, and J. Pandelaki, "Tuberculosis detection based on chest X-rays using ensemble method with CNN feature extraction," in *Proc. Int. Conf. Decis. Aid Sci. Appl. (DASA)*, Dec. 2021, pp. 682–686, doi: [10.1109/dasa53625.2021.9682237](https://doi.org/10.1109/dasa53625.2021.9682237).
- [26] M. A. Ganaie, M. Hu, A. K. Malik, M. Tanveer, and P. N. Suganthan, "Ensemble deep learning: A review," 2021, *arXiv:2104.02395*.
- [27] A. Manna, R. Kundu, D. Kaplun, A. Sinitca, and R. Sarkar, "A fuzzy rank-based ensemble of CNN models for classification of cervical cytology," *Sci. Rep.*, vol. 11, no. 1, p. 14538, Jul. 2021, doi: [10.1038/s41598-021-93783-8](https://doi.org/10.1038/s41598-021-93783-8).
- [28] S. Ghosal, M. Sarkar, and R. Sarkar, "NoFED-Net: Nonlinear fuzzy ensemble of deep neural networks for human activity recognition," *IEEE Internet Things J.*, vol. 9, no. 18, pp. 17526–17535, Sep. 2022, doi: [10.1109/jiot.2022.3155560](https://doi.org/10.1109/jiot.2022.3155560).
- [29] S. Dey, R. Bhattacharya, S. Malakar, F. Schwenker, and R. Sarkar, "Covid-ConvLSTM: A fuzzy ensemble model for COVID-19 detection from chest X-rays," *Exp. Syst. Appl.*, vol. 206, Nov. 2022, Art. no. 117812, doi: [10.1016/j.eswa.2022.117812](https://doi.org/10.1016/j.eswa.2022.117812).
- [30] F. Guo, J. Liu, M. Li, T. Huang, Y. Zhang, D. Li, and H. Zhou, "A concise TSK fuzzy ensemble classifier integrating dropout and bagging for high-dimensional problems," *IEEE Trans. Fuzzy Syst.*, vol. 30, no. 8, pp. 3176–3190, Aug. 2022, doi: [10.1109/TFUZZ.2021.3106330](https://doi.org/10.1109/TFUZZ.2021.3106330).
- [31] S. D. Deb and R. K. Jha, "Breast UltraSound image classification using fuzzy-rank-based ensemble network," *Biomed. Signal Process. Control*, vol. 85, Aug. 2023, Art. no. 104871, doi: [10.1016/j.bspc.2023.104871](https://doi.org/10.1016/j.bspc.2023.104871).
- [32] O. Day and T. M. Khoshgoftaar, "A survey on heterogeneous transfer learning," *J. Big Data*, vol. 4, no. 1, pp. 1–42, Sep. 2017, doi: [10.1186/s40537-017-0089-0](https://doi.org/10.1186/s40537-017-0089-0).
- [33] H. M. Gomes, J. P. Barddal, F. Enembreck, and A. Bifet, "A survey on ensemble learning for data stream classification," *ACM Comput. Surv.*, vol. 50, no. 2, pp. 1–36, Mar. 2018, doi: [10.1145/3054925](https://doi.org/10.1145/3054925).
- [34] I. D. Mienye and Y. Sun, "A survey of ensemble learning: Concepts, algorithms, applications, and prospects," *IEEE Access*, vol. 10, pp. 99129–99149, 2022, doi: [10.1109/ACCESS.2022.3207287](https://doi.org/10.1109/ACCESS.2022.3207287).
- [35] H. R. Kadkhodaei, A. M. E. Moghadam, and M. Dehghan, "HBoost: A heterogeneous ensemble classifier based on the boosting method and entropy measurement," *Exp. Syst. Appl.*, vol. 157, Nov. 2020, Art. no. 113482, doi: [10.1016/j.eswa.2020.113482](https://doi.org/10.1016/j.eswa.2020.113482).
- [36] S. I. Abba, R. A. Abdulkadir, M. S. Gaya, M. A. Saleh, P. Esmaili, and M. B. Jibril, "Neuro-fuzzy ensemble techniques for the prediction of turbidity in water treatment plant," in *Proc. 2nd Int. Conf. IEEE Nigeria Comput. Chapter*, Oct. 2019, pp. 1–6. [Online]. Available: <https://ieeexplore.ieee.org/document/8949629>
- [37] Z. Shen, X. Chen, and J. M. Garibaldi, "A fuzzy aggregation based ensemble framework for accurate and stable feature selection," in *Proc. IEEE Int. Conf. Fuzzy Syst.*, Jul. 2021, pp. 1–7. [Online]. Available: <https://ieeexplore.ieee.org/document/9494539>
- [38] Z. Hosseinpour, M. M. Arefi, N. Mozafari, H. Luo, and S. Yin, "An ensemble-based fuzzy rough active learning approach for broken rotor bar detection in nonstationary environment," *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1–8, 2022, doi: [10.1109/tim.2022.3178743](https://doi.org/10.1109/tim.2022.3178743).
- [39] K. Kaczmarek-Majer and A. Kiersztyn, "Experimental evaluation of the accuracy of an ensemble of fuzzy methods for classification of episodes in bipolar disorder," in *Proc. IEEE Int. Conf. Fuzzy Syst.*, Jul. 2022, pp. 1–7. [Online]. Available: <https://ieeexplore.ieee.org/document/9882582>
- [40] M. Wu, X. Jiang, K. Du, Y. Xu, and W. Zhang, "Ensemble machine learning algorithm for predicting acute kidney injury in patients admitted to the neurointensive care unit following brain surgery," *Sci. Rep.*, vol. 13, no. 1, p. 6705, Apr. 2023, doi: [10.1038/s41598-023-33930-5](https://doi.org/10.1038/s41598-023-33930-5).
- [41] N. M. Suganthi, V. M. Jemin, P. Rama, and E. Chandralekha, "Chronic kidney disease detection using AdaBoosting ensemble method and K-fold cross validation," in *Proc. Int. Conf. Automat., Comput. Renew. Syst.*, 2022, pp. 979–983. [Online]. Available: <https://ieeexplore.ieee.org/document/10029047>
- [42] Z. Xiong, H. Zhang, Y. Chen, and Y. Song, "Deep ensemble learning network for kidney lesion detection," in *Proc. Chin. Automat. Congr.*, 2019, pp. 3841–3846. [Online]. Available: <https://ieeexplore.ieee.org/document/8997272>
- [43] Y. Wang, Y. Wei, Q. Wu, H. Yang, and J. Li, "An acute kidney injury prediction model based on ensemble learning algorithm," in *Proc. 10th Int. Conf. Inf. Technol. Med. Educ.*, Aug. 2019, pp. 18–22. [Online]. Available: <https://ieeexplore.ieee.org/document/8965012>
- [44] Nikhila, "Chronic kidney disease prediction using machine learning ensemble algorithm," in *Proc. Int. Conf. Comput., Commun., Intell. Syst.*, Feb. 2021, pp. 476–480. [Online]. Available: <https://ieeexplore.ieee.org/document/9397144>
- [45] V. Chaurasia, M. K. Pandey, and S. Pal, "Chronic kidney disease: A prediction and comparison of ensemble and basic classifiers performance," *Hum.-Intell. Syst. Integr.*, vol. 4, nos. 1–2, pp. 1–10, Feb. 2022, doi: [10.1007/s42454-022-00040-y](https://doi.org/10.1007/s42454-022-00040-y).
- [46] W. Yang and L. Deng, "PreDBA: A heterogeneous ensemble approach for predicting protein-DNA binding affinity," *Sci. Rep.*, vol. 10, no. 1, p. 1278, Jan. 2020, doi: [10.1038/s41598-020-57778-1](https://doi.org/10.1038/s41598-020-57778-1).
- [47] T. G. Dietterich, "Approximate statistical tests for comparing supervised classification learning algorithms," *Neural Comput.*, vol. 10, no. 7, pp. 1895–1923, Oct. 1998, doi: [10.1162/089976698300017197](https://doi.org/10.1162/089976698300017197).
- [48] T. H. Duong, N. T. Nguyen, and G. S. Jo, "A method for integration of WordNet-based ontologies using distance measures," in *Proc. Int. Conf. Knowl.-Based Intell. Inf. Eng. Syst.*, Aug. 2008, pp. 210–219, doi: [10.1007/978-3-540-85563-7\\_31](https://doi.org/10.1007/978-3-540-85563-7_31).
- [49] N. T. Nguyen, "Conflicts of ontologies—Classification and consensus-based methods for resolving," in *Proc. Int. Conf. Knowl.-Based Intell. Inf. Eng. Syst.*, Jan. 2006, pp. 267–274, doi: [10.1007/11893004\\_34](https://doi.org/10.1007/11893004_34).
- [50] N. T. Nguyen and J. Sobecki, "Using consensus methods to construct adaptive interfaces in multimodal web-based systems," *Universal Access Inf. Soc.*, vol. 2, no. 4, pp. 342–358, Nov. 2003, doi: [10.1007/s10209-003-0050-1](https://doi.org/10.1007/s10209-003-0050-1).



**JAVOKHIR MUSAUEV** received the B.S. degree in world economy and international economic relations from the University of World Economy and International Economic Relations, Uzbekistan, in 2016, and the M.S. degree from Keimyung University, South Korea, in 2020. He is currently pursuing the Ph.D. degree with the Department of Computer Engineering, Yeungnam University, South Korea. He has authored two conference papers and one patent. His research interests include computer vision and pattern recognition using deep learning techniques.



**ABDULAZIZ ANORBOEV** received the B.S. degree in finance from Bangor University, Bangor, Wales, U.K., in 2018, and the M.S. degree in business administration from Keimyung University, Daegu, South Korea, in 2020. He is currently pursuing the Ph.D. degree with the Department of Computer Engineering, Yeungnam University, South Korea. His research interests include deep learning, computer vision, and pattern recognition.



**YEONG-SEOK SEO** (Member, IEEE) received the B.S. degree in computer science from Soongsil University, Seoul, Republic of Korea, in 2006, and the M.S. and Ph.D. degrees in computer science from the Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Republic of Korea, in 2008 and 2012, respectively.

From September 2012 to December 2013, he was a Postdoctoral Researcher with the Institute for Information and Electronics, KAIST. From January 2014 to August 2016, he was a Senior Researcher with the Korea Testing Laboratory (KTL), Seoul. From September 2016 to August 2022, he was an Assistant Professor (Tenure Track) with the Department of Computer Engineering, Yeungnam University, Gyeongsan, Gyeongbuk, Republic of Korea. Since September 2022, he has been an Associate Professor (Tenure Track) with Yeungnam University. His research interests include software engineering, artificial intelligence, the Internet of Things, and big data analysis.

Prof. Seo is a member of the Board of Directors of the Software Engineering Society in South Korea. Furthermore, he is involved in international standardization activities and is a member of the Korean National Body Mirror Committee to ISO on IT Service Management and IT Governance (ISO/IEC JTC1/SC40). He also served as a Technical Committee Member for some international conferences and workshops, such as ICSE 2020 Demonstrate Track, ICSE 2020 Software Engineering in Practice, MITA 2019, QRS 2020, CSA 2020, WITC 2021, FutureTech 2021, CUTE 2021, CSA 2021, WITC 2022, FutureTech 2022, CUTE 2022, CSA 2022, MITA 2021, WITC 2023, ACIIDS 2023, and ICCCI 2023. He was a recipient of the Undang Academic Paper Award (Grand Prize), in 2022, and the 2nd JIPS Survey Paper Awards, in 2019. He received the Best Paper Award from the ASK 2022, ASK 2021, and MITA 2021 and 2019. He is an Associate Editor of *Human-Centric Computing and Information Sciences* (HCIS) (SCI indexed), *Processes* (SCI indexed), *Electronics* (SCI indexed), and *Journal of Information Processing Systems* (JIPS) (Scopus/ESCI indexed). He has served as the Chair for international conferences and workshops, such as the Proceedings Co-Chair for APSEC 2018, the Publicity Chair for WITC 2019, and the Program Chair for HCIS 2019. He was the Guest Editor of *Journal of Systems and Software* (JSS) (SCI indexed).



**NGOC THANH NGUYEN** (Senior Member, IEEE) is currently a Full Professor with the Wrocław University of Science and Technology and the Head of the Information Systems Department, Faculty of Computer Science and Management. He has given 22 plenary and keynote speeches for international conferences, and more than 40 invited lectures in many countries. He has edited more than 30 special issues in international journals, 52 books, and 35 conference proceedings.

He is the author or coauthor of five monographs and more than 350 journal articles and conference papers. His research interests include collective intelligence, knowledge integration methods, inconsistent knowledge

processing, and multi-agent systems. He serves as a member for the Council of Scientific Excellence of Poland, a member for the Committee on Informatics of the Polish Academy of Sciences, and an Expert for the National Center of Research and Development and European Commission in evaluating research projects in several programs, such as Marie Skłodowska-Curie Individual Fellowships, FET, and EUREKA. He was the general chair or the program chair of more than 40 international conferences. He also serves as the Chair for the IEEE SMC Technical Committee on Computational Collective Intelligence. He serves as the Editor-in-Chief for *International Journal of Information and Telecommunication* (Taylor&Francis), *Transactions on Computational Collective Intelligence* (Springer), and *Vietnam Journal of Computer Science* (World Scientific). He is also the Associate Editor-in-Chief of several prestigious international journals, among others, *Journal of Intelligent and Fuzzy Systems* and *Applied Intelligence*. In 2009, he was granted the title of Distinguished Scientist of ACM. He was also a Distinguished Visitor of IEEE and a Distinguished Speaker of ACM.



**DOSAM HWANG** received the Ph.D. degree from Kyoto University, Kyoto, Japan. He was a Full Professor with the Department of Computer Engineering, Yeungnam University, South Korea, from 1996 to 2023. From 2005 to 2009, he was the Head of the Computer Engineering Department, Yeungnam University, for five years. He has also held a position as a Principal Researcher with the Korea Institute of Science and Technology (KIST). He has been a Visiting Professor with the Korea

Advanced Institute of Science and Technology (KAIST). He is currently a Professor Emeritus with Yeungnam University. He has participated in several Korean national research projects, such as a project on Machine Translation System (1985–1990), the National IT Ontology Infrastructure and Technology Development Project called “CoreOnto” (2006–2009), and “Exobrain” (2013–2014), the project focused on the construction of deep knowledge base and question-answering platform. He is now in charge of an intelligent service integration based on the IoT big data as part of Korea’s other principal national research project “BK+” (since 2014). He had more than 50 publications. His research interests include natural language processing, ontology, knowledge engineering, information retrieval, and machine translation. In recognition of his great commitment and contribution to the relevant fields of study, he was honored as a Distinguished Researcher of KIST by Korea’s Ministry of Science and Technology (MoST), in 1988. He received the Prize for Good Conduct from the Kyunghee High School, in 1973. He has so far been not only the co-chair of several international conferences but also a Steering Committee Member of ICCCI and ACIIDS, and MISSI international conferences. More specifically, for example, he has been an Assistant Secretary of ISO/TC37/SC4 for language resource management, from 2005 to 2007, and also the Secretary of the Korean TC for ISO/TC37/SC4, from 2005 to 2007. In 2006, he was the Director of the Korean Society for Cognitive Science (KSCS) and the Korean Information Science Society (KISS). He has been the Society’s Director and the Mentor of a Knowledge Engineering Study Group, since 2007.

...