

Investigating the Impact of Train / Test Split Ratio on the Performance of Pre-Trained Models with Custom Datasets

Houda Bichri¹, Adil Chergui², Mustapha Hain³

AICSE lab, Department of Computer Science, ENSAM Casablanca, Casablanca, Morocco^{1,3}
ISSIEE lab, Department of Computer Science, ENSAM Casablanca, Casablanca, Morocco²

Abstract—The proper allocation of data between training and testing is a critical factor influencing the performance of deep learning models, especially those built upon pre-trained architectures. Having the suitable training set size is an important factor for the classification model's generalization performance. The main goal of this study is to find the appropriate training set size for three pre-trained networks using different custom datasets. For this aim, the study presented in this paper explores the effect of varying the train / test split ratio on the performance of three popular pre-trained models, namely MobileNetV2, ResNet50v2 and VGG19, with a focus on image classification task. In this work, three balanced datasets never seen by the models have been used, each containing 1000 images divided into two classes. The train / test split ratios used for this study are: 60-40, 70-30, 80-20 and 90-10. The focus was on the critical metrics of sensitivity, specificity and overall accuracy to evaluate the performance of the classifiers under the different ratios. Experimental results show that, the performance of the classifiers is affected by varying the training / testing split ratio for the three custom datasets. Moreover, with the three pre-trained models, using more than 70% of the dataset images for the training task gives better performance.

Keywords—Artificial intelligence; classification; MobileNetV2; ResNet50v2; sensitivity; specificity; train / test split ratio; VGG19

I. INTRODUCTION

In our daily life, we have a huge number of generated data of different forms: text, image and video obtained from cameras and sensors. This data can be analyzed efficiently by using the advanced techniques such as deep learning. In image classification, deep learning models are used to identify special features in the images characterizing a particular class and that will help the model to distinguish between different classes. They can reach the human level performance on several fields like classifying animals, different types of food, diseases.

When having a small dataset, the best way to classify images is by using the transfer learning approach. It uses one model's knowledge on a machine learning task and reuses it as a starting point for a different but a related task. The pre-trained neural network is fine-tuned to achieve the user's needs rather than being trained the model from scratch.

Using transfer learning on image classification was introduced in the literature in several areas:

- The authors in study [1] propose the use of VGG19 architecture as the base model and complement it with different state-of-the-art techniques to classify histopathological images into IDC and non-IDC classes, Invasive Ductal Carcinoma (IDC) is a type of breast cancer.
- To identify covid-19, pneumonia and lung cancer diseases using chest radiographs, researchers in paper [2] suggest the combination of VGG-19 and Convolutional Neural Networks (CNN) to improve the performance of the multi-class classification task.
- The study presented by [3] gives different fine-tuned pre-trained models such as VGG-19, ResNet50V2 and DenseNet-121 to predict sentiments using the Twitter-based images.
- Researchers in the paper [4] gave the performance evaluation of Resnet18, Resnet50, Alexnet, DenseNet121, DenseNet201 and VGG16 models in rating gravel road images obtained from self-recorded videos and from Google Street View.
- The classification of a forest fire imagery into forest fire and no-fire was introduced in the paper [5] using a new proposed approach based on the use of the VGG19 model.
- To improve the intention classification accuracy, researchers in the paper [6] used the knowledge of the ERNIE model (Enhanced Representation through Knowledge Integration) for both: the student and the teacher models.
- The paper in [7] presents image classification and image prediction for the ImageNet dataset using the pre-trained models: MobileNet, MobileNetV2, VGG16, VGG19 and ResNet50.
- For the land use and land cover classification, transfer learning was used in the study presented in the paper [8] to fine-tune the pre-trained models: VGG16 and WRNs (Wide Residual Networks). To compare the performance and computational time, some techniques were employed such as: gradient clipping, data augmentation and early stopping. The red-green-blue version of the EuroSAT dataset was used in this work.

- The paper in [9] presents a systematic review of the early detection of Alzheimer disease (AD) by using transfer learning and neuroimaging biomarkers. In this review, five datasets were used. The researchers in this paper confirm that, for the early diagnosis of AD, the use of transfer learning technique is beneficial to develop a more accurate model.
- Transfer learning studies that uses the non-medical ImageNet datasets for medical image analysis was systematically reviewed in the paper [10]. To approach medical tasks with a non-medical dataset, the researchers suggest the use of transfer learning with ImageNet dataset. They also approve that CNN model and transfer learning technique gave reasonable performance.
- To classify mangrove communities, the study presented in the paper [11] uses three transfer learning strategies and discuss the differences in the classification task. Different models were constructed with the three deep learning algorithms: DeepLabV3+, HRNet and MCCUNet,
- For Arabic tweet classification, a transformer-based model was proposed in the paper [12]. This model was constructed from a pre-trained BERT model given by the hugging face transformer library using custom dense layers. To categorize the tweets, a multi-class classification layer was built on the top of the BERT encoder. Five publicly datasets was employed to do this study.
- The paper in [13] presents a review of the diabetic retinopathy classification with deep learning models that use transfer learning technique. According to this work, transfer learning is useful with medical image classification due to the limited number of medical images.
- To perform urban sounds classification, the researchers in the paper [14] have applied transfer learning with three datasets: UrbanSound8k, ESC-10 and Air Compressor. The pre-trained models used in this study are: GoogLeNet, SqueezeNet, ShuffleNet, VGGish and YAMNet.
- The detection of the Covid-19 disease was done in the paper [15] using the transfer learning technique with a dataset formed by X-ray images. The dataset is divided in three folders: Covid-19, pneumonia and normal (healthy) cases. Although a small dataset was used, high accuracy is achieved over all the models. The proposed approach uses VGG16, VGG19 and ResNet101 architectures.
- To categorize various food products using transfer learning, a recognition model was introduced in the paper [16]. The dataset employed in this work is Food-101. The proposed model, that uses Efficientnetb0

architecture, reaches the accuracy of 80% which is the best accuracy compared to other state of the art models.

- The study presented in the paper [17] applied the transfer learning technique for classifying monkeypox skin lesions. Six different models were employed in this work: DenseNet201, InceptionResNetV2, EfficientNetB7, InceptionV3, ResNet50 and VGG16. The paper proposes also a fine-tuned version of the InceptionV3 model named PoxNet22. The proposed transfer learning-based model gives better performance in terms of accuracy, recall and precision.
- The paper [18] gives an evaluation of pre-trained models for the detection of osteoporosis which is a bone disease in knee radiographs. VGG16 and VGG16 with fine-tuning were used in this study. The models were evaluated using accuracy, sensitivity and specificity metrics. According to this study, fine-tuning improves the VGG-16 performance for the desired task.
- Different pre-trained models for bird image identification were studied and compared in the paper [19]. The models employed are: DenseNet201, InceptionV3, MobileNetV2 and ResNet52V2. The dataset contains 58388 bird images belonging to 400 species. All the implemented models give good accuracy but DenseNet201 was the best network, according to the authors.
- The researchers in the paper [20] confirm that the VGG16 gives a good performance with all the nine different chest X-ray datasets used. The datasets have various sizes and different class labels.

While using the transfer learning technique, the obtained model's accuracy can exceed 90% even when using datasets with less than 100 images in each class [21] just by using the correct implementation of the pre-trained model.

In a previous work [22], a performance comparison of three pre-trained models on the classification task using a custom dataset was performed. The models were trained on 30 epochs with 10^{-3} and 20 as values for the learning rate and the batch size parameters respectively. VGG19 achieved the highest accuracy, precision, recall and f1-score.

This work follows the perspective of those researches, by proposing a study of variation impact of train / test split ratio on the performance of three fine-tuned pre-trained models (MobileNetV2, ResNet50v2 and VGG19), while using a new dataset never seen by the models.

This paper is organized as follows:

Section II contains a literature review on machine learning pre-trained models. Section III introduces the pre-trained models. Section IV contains the description of the three datasets used in this study. The preprocessing phase and evaluation metrics employed for the performance comparison are described in Section V. Results and discussion is given in Section VI. The conclusion is in Section VII.

II. LITERATURE REVIEW ON MACHINE LEARNING PRE-TRAINED MODELS

Many researches try to understand how to enhance performance of ML pre-trained models. Here are some of those studies:

- The work presented in paper [23] gives a study of the effects of dataset size and training/testing split ratios on the performance of multiclass classifiers. The results demonstrate that XGBoost gives the best performance. The performance evaluation was done using 25 performance parameters.
- The paper in [24] presents a CNN-based automatic model for the identification of the strawberry leaf plant disease like: powdery mildew leaf, healthy leaf and caterpillar pests leaf. MobileNetV3-Large and efficientNet-B0 were implemented as architecture. The dataset contains 1336 images collected from the field and the data augmentation was applied to it.
- The paper in [25] proposed a deep learning method based on CNN architecture to classify six types of strawberry plants diseases. This study utilizes 4663 strawberry leaf disease images data.
- The work presented in the paper [26] demonstrates that the use of CNN models is useful than the non-Deep learning models to distinguish between infected and healthy strawberry leaves. Under the supervision of disease specialists, the dataset (1450 images) were collected from Balamore and Millen farms Ltd. AlexNet, SqueezeNet, GoogleNet, ResNet50, SqueezeNet-MOD1 and SqueezeNet-MOD2 were employed in this study.
- The paper [27] presents an evaluation of four ensemble learning algorithms (random forest, CatBoost, XGBoost and random forest) for the prediction of heart disease using different hyperparameter optimization techniques. Three kaggle datasets were combined which had features to augment the dataset size. Using 80% of the dataset images for training was useful because the proposed model gives better accuracy while working with the train / test split ratio of 80%-20%.
- The paper [28] studied the impact of different train / test split ratios on the model performance. The ratios: 50-50, 60-40, 70-30, 80-20 and 90-10 were used in this work. The authors recommend the use of different classifiers with different train / test split ratios.

III. PRE-TRAINED MODELS

A. MobileNetV2

MobileNetV2 [29] is a CNN model based on an inverted residual structure where the residual connections are between the bottleneck layers. The architecture incorporates shortcut connections to aid in training deeper networks without vanishing gradients. To improve efficiency, the model uses depth-wise separable convolution which is independently

performed for each input channel. Depth-wise separable convolution reduces the complexity cost and the pre-trained model size. Due to this, MobileNetV2 has higher accuracy, needs fewer operations and is much faster than the MobileNetV1 model. The MobileNetV2 architecture (see Fig. 1) consists of 17 building blocks in a row followed by 1x1 convolutional layer, global average pooling layer and classification layer. The expansion layer role is to expand the number of channels in the data. In the projection layer, high number of dimensions is reduced to a smaller one.

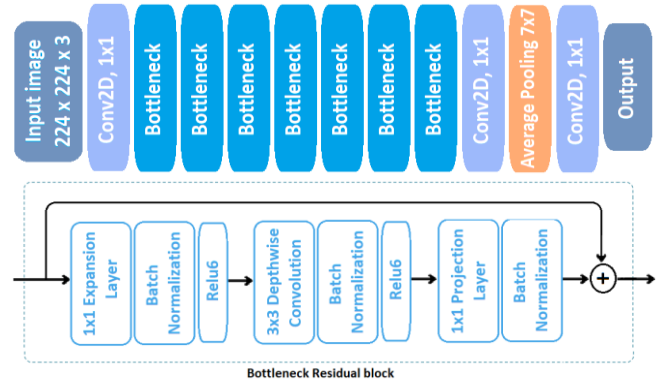


Fig. 1. MobileNetV2 architecture.

B. VGG19

VGG19 [30], part of the VGG family (Visual Geometry Group), is a CNN architecture with multiple layers. It was published by Simony and Zisserman researchers from the Oxford University in 2014. It consists of 19 layers: 16 convolutional and three fully connected layers with a filter size of 3x3 (see Fig. 2). The number of parameters is reduced due to the small kernel size; it also enables them to cover the entire image. VGG19 was trained on the ImageNet database that contains more than 14 million images belonging to 1000 categories which helps the network to capture a diverse set of features, making it a powerful tool for the transfer learning task. For downsampling, VGG19 incorporates max-pooling layers and uses fully connected ones for classification.



Fig. 2. VGG19 architecture.

C. ResNet50v2

ResNet50V2 [31], is a 50-convolutional neural network: 48 convolutional layers, one MaxPool layer and one average pool layer. It is known for its depth and skip connections which protect the model from vanishing gradient problem in much deeper networks. ResNet50V2 uses residual blocks which enhance the training efficiency in achieving both depth and accuracy in different tasks. To reduce the computational complexity and adjust the input layer to increase the performance of the network, ResNet50V2 utilizes batch normalization and bottleneck blocks. The ResNet50V2 architecture is shown in Fig. 3.

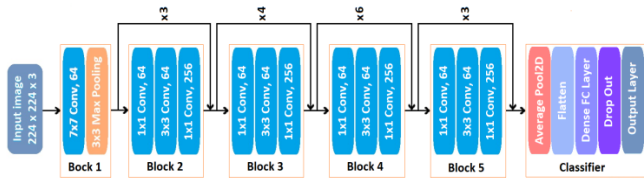


Fig. 3. ResNet50v2 architecture.

IV. DATASETS

The balance (or imbalance) of the classes, which is the diversity of samples belonging to each class is a significant factor that affects the performance of classification models. Providing imbalanced data to the classifier may bias it towards the majority class because it lacks enough data to learn about the minority, which can cause false predictions. In this context, the study presented in this work was done with three balanced datasets each containing 1000 colored images divided into two classes: class_0 (500 images) and class_1 (500 images). The description of the three datasets is in Table I.

Some samples of images in the three datasets are shown in Fig. 4, Fig. 5 and Fig. 6.



Fig. 4. Sample images from Dataset1.

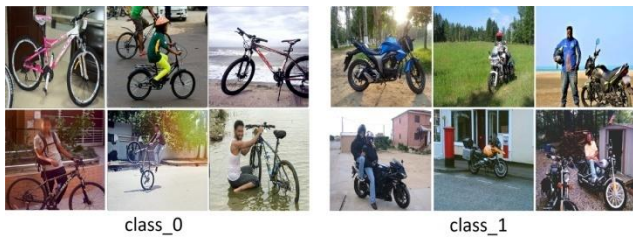


Fig. 5. Sample images from Dataset2.

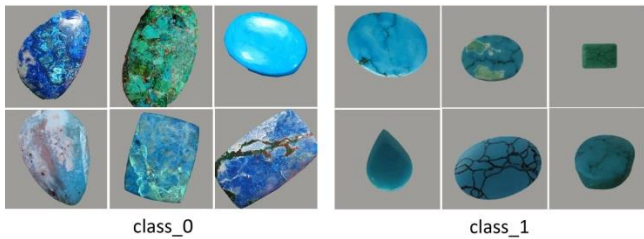


Fig. 6. Sample images from Dataset3.

TABLE I. DESCRIPTION OF THE DATASETS

Dataset name	Description	Class_0 description	Class_1 description
Dataset1 (Fig. 4)	Formed from the kaggle dataset named 'Images of Strawberry Leaves for Tipburn Detection' ¹ . It contains 1431 images divided into two classes: healthy (626 images) and calciumdeficiency (805 images).	Contains 500 images of leaves with calcium deficiency.	Contains images of 500 images healthy leaves.
Dataset2 (Fig. 5)	Formed from the kaggle dataset named 'Paribahan BD' ² . It contains 7474 images divided into two classes: local-vehicles (7474 images grouped in 8 folders) and generated images (80 images grouped in 8 folders).	Contains 500 images of bicycles.	Contains 500 images of bikes.
Dataset3 (Fig. 6)	Formed from the kaggle dataset named 'Gemstones' ³ . It contains 6043 images divided into three classes: train (3043 images grouped in 6 folders), test and validation (each one contains 1500 images grouped in 6 folders).	Contains 500 images of turquoise gemstones.	Contains 500 images of fake-turquoise gemstones.

V. PREPROCESSING AND EVALUATION METRICS

A. Preprocessing

One of the major problems when training deep learning models is to have a large dataset which is not always an easy task. It is necessary to have a huge number of images in each class of several subjects of the classification. To expand the size of the three small datasets used in this work, it was beneficial to utilize the data augmentation process with the KerasImageDataGenerator class. The summarized data augmentation description is in Table II. The images have different dimensions, for this reason, they were rescaled to 224x224 pixel resolution to make them compatible with the pre-trained model's requirement.

B. Evaluation Metrics

In machine learning, the performance evaluation of classification models needs the use of some metrics to be able to solve real-world problems.

There are several measures to test the performance of classification results [32], the three following ones were considered in this study:

¹<https://www.kaggle.com/datasets/ercanavsar/images-of-strawberry-leaves-for-tipburn-detection>

²https://www.kaggle.com/datasets/naifislam/paribahan-bd?select=generated_images

³<https://www.kaggle.com/datasets/muhammadmuzamil5500/gemstones>

$$Sensitivity = \frac{TP}{TP + FN} \quad (1)$$

$$Specificity = \frac{TN}{TN + FP} \quad (2)$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (3)$$

TP: True Positive, TN: True Negative, FP: False Positive and FN: False Negative.

TABLE II. SUMMARIZED DATA AUGMENTATION DESCRIPTION

Name	Value / Description
Rotation_range	20
Zoom_range	0.15
Width_shift_range	0.2
height_shift_range	0.2
Shear_range	0.15
horizontal_flip	True
Fill_mode	'nearest'

VI. RESULTS AND DISCUSSION

The pre-trained models were trained in Google Colab notebook with a learning rate of 10^{-3} , a batch size of 32 for 60 epochs. Graphics Processing Units (GPUs) was used as the model's hardware platform.

Adam function is the simple and time-efficient optimizer for deep neural networks; thus, it has been employed for the compilation process.

The results discussed in this work are the best ones achieved from several experiments which were carried out for each case.

A. Dataset1

From Table III, it can be observed that the MobileNetV2 achieves the best sensitivity (100%) with the ratio 90%-10% and the best specificity (99%) with the ratio 80%-20%, but the best performance in terms of accuracy is obtained when using 70% of the dataset for training and 30% for testing (97.67%).

TABLE III. MOBILENETV2 EXPERIMENTATION RESULTS ON DATASET1

Train / Test ratio	Sensitivity (%)	Specificity (%)	Accuracy (%)
60% - 40%	97	97.5	97.25
70% - 30%	96.67	98.67	97.67
80% - 20%	93	99	96
90% - 10%	100	94	97

Looking at the plot of confusion matrix (see Fig. 7), it can be seen that MobileNetV2 model accurately predicted 293 out of 300 total samples (train / test ratio: 70%-30%).

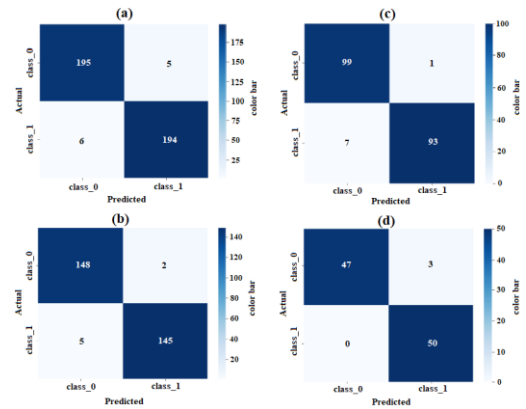


Fig. 7. Confusion matrices for MobileNetV2 with the ratios: (a) 60%-40%; (b) 70%-30%, (c) 80%-20% and (d) 90%-10%.

Resnet50V2 performs well with the Dataset1: the accuracy is greater than 97% with all train / test ratios (see Table IV). The best sensitivity score (99.5%) is observed when using 60% of the dataset for the training phase. The high specificity score (100%) is obtained with the ratio 90%-10%. But the best performance of the network is achieved with the ratio 70%-30% in terms of accuracy (98.5%).

TABLE IV. RESNET50V2 EXPERIMENTATION RESULTS ON DATASET1

Train / Test ratio	Sensitivity (%)	Specificity (%)	Accuracy (%)
60% - 40%	99.5	96.5	98
70% - 30%	98.67	97.33	98
80% - 20%	98	99	98.5
90% - 10%	96	100	98

ResNet50v2 classifier gives better performance with the ratio 80%-20%: it predicted accurately 197 samples which represent 98.5% of the total samples (200 samples) (see Fig. 8).

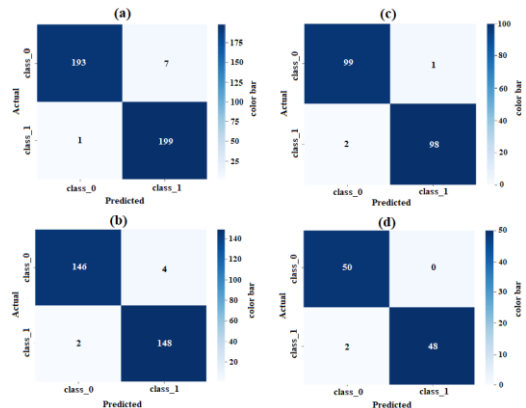


Fig. 8. Confusion matrices for ResNet50v2 (Dataset1).

Although the VGG19 classifier achieves the best sensitivity score (97%) with the ratio 80%-20% and the best specificity one (100%) with the ratio 90%-10%, the best performance of the network is observed while using 70%-30% as a train / test ratio. With this ratio (70%-30%), VGG19 reaches 97.33% for the accuracy metric (see Table V).

TABLE V. VGG19 EXPERIMENTATION RESULTS ON DATASET1

Train / Test ratio	Sensitivity (%)	Specificity (%)	Accuracy (%)
60% - 40%	94	98	96
70% - 30%	96	98.67	97.33
80% - 20%	97	96	96.5
90% - 10%	88	100	94

The VGG19’s confusion matrix, plotted in Fig. 9, shows that the model succeeds to classify 292 samples out of all samples while using the ratio 70%-30%.

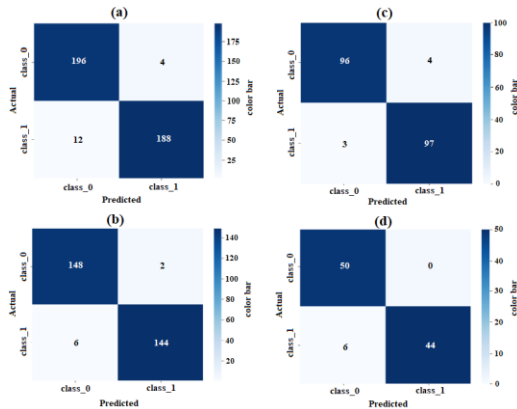


Fig. 9. Confusion matrices for VGG19 (Dataset1).

B. Dataset2

Out of all the train / test ratios, MobileNetV2 performs well with 80%-20% and 90%-10% in terms of sensitivity, specificity and accuracy with 100%, 98% and 99% respectively (see Table VI).

TABLE VI. MOBILENETV2 EXPERIMENTATION RESULTS ON DATASET2

Train / Test ratio	Sensitivity (%)	Specificity (%)	Accuracy (%)
60% - 40%	99	96.5	97.75
70% - 30%	99.33	96.67	98
80% - 20%	100	98	99
90% - 10%	100	98	99

With the ratios 80%-20% and 90%-10%, MobileNetV2 classifies accurately 99% of all the samples which is the best result obtained by this network (see Fig. 10).

The network gives better sensitivity (100%) while working with 90% of the dataset for the training phase. The better specificity and accuracy were reached with the ratio 80%-20% with scores of 100% and 99.5% respectively (see Table VII).

ResNet50v2 classifier gives better performance with the ratio 80%-20%: it predicted accurately 199 samples which represent 99.5% of the total samples (200 samples) (see Fig. 11).

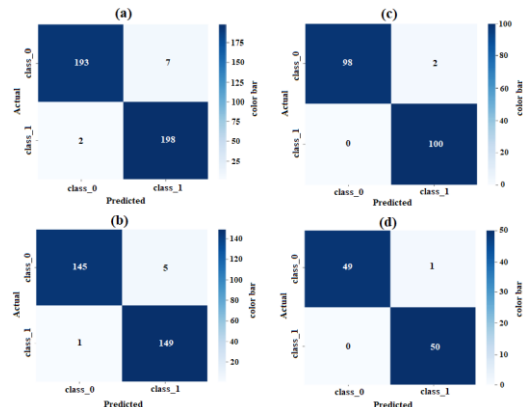


Fig. 10. Confusion matrices for MobileNetV2 (Dataset2).

TABLE VII. RESNET50V2 EXPERIMENTATION RESULTS ON DATASET2

Train / Test ratio	Sensitivity (%)	Specificity (%)	Accuracy (%)
60% - 40%	99	98	98.5
70% - 30%	99.33	96	97.67
80% - 20%	99	100	99.5
90% - 10%	100	98	99

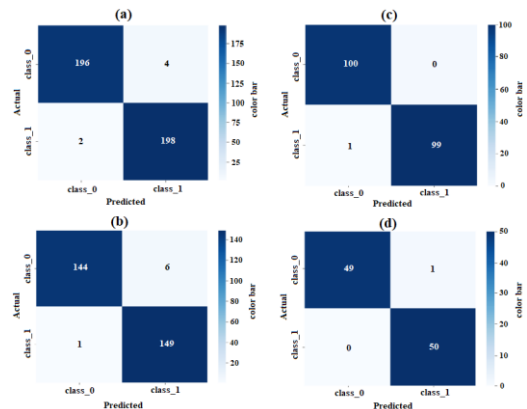


Fig. 11. Confusion matrices for ResNet50v2 (Dataset2).

The train / test ratio 90%-10% gives better sensitivity, specificity and accuracy with the score 100% for each one of them (see Table VIII).

TABLE VIII. VGG19 EXPERIMENTATION RESULTS ON DATASET2

Train / Test ratio	Sensitivity (%)	Specificity (%)	Accuracy (%)
60% - 40%	99.5	96	97.75
70% - 30%	99.33	98	98.67
80% - 20%	99	99	99
90% - 10%	100	100	100

The pre-trained model succeeds to correctly classify all the samples while taking 90% of the dataset for training which represent a good result (see Fig. 12).

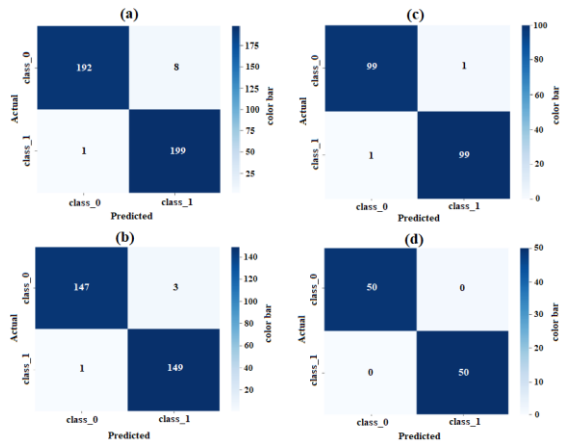


Fig. 12. Confusion matrices for VGG19 (Dataset2).

C. Dataset3

The MobileNetV2 model achieves better sensitivity (100%) with the ratios 80%-20%, and better specificity (100%) with the ratios 60%-40%. But out of all the train / test ratios, 80%-20% gives significantly better accuracy with a score of 99.5% which is a good result (see Table IX).

TABLE IX. MOBILENETV2 EXPERIMENTATION RESULTS ON DATASET3

Train / Test ratio	Sensitivity (%)	Specificity (%)	Accuracy %
60% - 40%	98.5	100	99.25
70% - 30%	98.67	100	99.33
80% - 20%	100	99	99.5
90% - 10%	100	98	99

According to the confusion matrix of MobileNetV2 model with the ratios 80%-20%, it can be observed that the pre-trained model arrives to accurately predict 99.5% of all samples (199 samples of 200) (see Fig. 13).

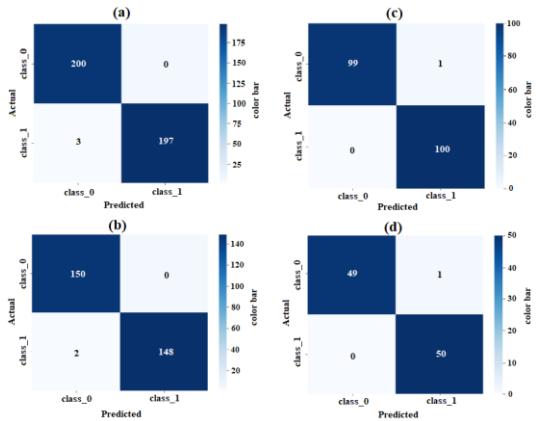


Fig. 13. Confusion matrices for MobileNetV2 (Dataset3).

With the Dataset3, the ResNet50V2 classifier achieves the perfect sensitivity, specificity and accuracy while working with 90% of the total dataset's samples for training and 10% for testing, with a score of 100% for each one of them. It also gives better specificity with the ratio 80%-20% (see Table X).

TABLE X. RESNET50V2 EXPERIMENTATION RESULTS ON DATASET3

Train / Test ratio	Sensitivity (%)	Specificity (%)	Accuracy (%)
60% - 40%	99.5	97.5	98.5
70% - 30%	99.33	99.33	99.33
80% - 20%	98	100	99
90% - 10%	100	100	100

ResNet50v2 classifier gives better performance with the ratio 90%-10%: it predicted successfully all the samples (see Fig. 14).

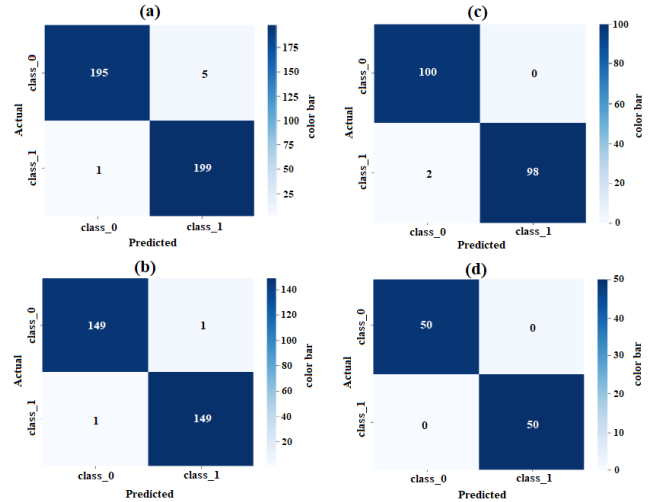


Fig. 14. Confusion matrices for ResNet50v2 (Dataset3).

With the train / test ratio 90%-10%, the VGG19 pre-trained model gives better performance in terms of sensitivity, specificity and accuracy with a score of 100% for each one of the metrics (see Table XI).

TABLE XI. VGG19 EXPERIMENTATION RESULTS ON DATASET3

Train / Test ratio	Sensitivity (%)	Specificity (%)	Accuracy (%)
60% - 40%	97.5	99.5	98.5
70% - 30%	98	99.33	98.67
80% - 20%	98	100	99
90% - 10%	100	100	100

The VGG19's confusion matrix, plotted in Fig. 15, shows that the model succeeds to classify correctly all the samples while using the ratio 90%-10%.

The results show that the best performance achieved with the networks was when using the train / test ratios 80%-20% and 90%-10%. While working with the ratio 60%-40%, all the classifiers couldn't give better scores in terms of sensitivity, specificity and accuracy, as well as the other ratios. The three networks need more than 70% of the dataset's samples for the training phase to give better results.

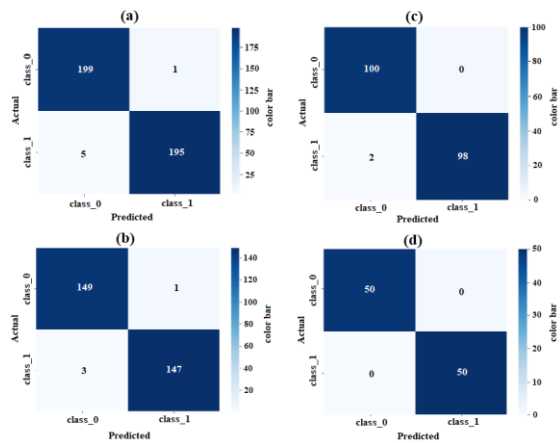


Fig. 15. Confusion matrices for VGG19 (Dataset3).

VII. CONCLUSION

This study was realized with three datasets never seen by the pre-trained models: MobileNetV2, ResNet50 v2 and VGG19. The datasets were divided in two classes (class_0 and class_1). For all experiences, the batch size was fixed at 32 and the learning rate at 10^{-3} . All the experiences were for 60 epochs. Analyzing the results, it can be observed that the train / test split ratio has a significant impact on the classification performance of the three pre-trained networks: MobileNetV2, ResNet50v2 and VGG19, the ratios 80%-20% and 90%-10% gives better results on the most cases.

All the pre-trained networks used in this study performs well with the Dataset3, this is due to its simplicity in comparison with the other datasets (Dataset2 and Dataset3). It has less features to be learned by the models which facilitate the learning process and enhance the classifier's performance.

In conclusion, increasing the size of the train data enhanced the performance of the three classifiers; more than 70% of the dataset's samples is required in the training phase to achieve better performance.

For future work, other datasets with different sizes will be studied using the three pre-trained models to have a more generalized conclusion concerning the impact of the train / test split ratio on the performance of the networks. Other architectures could be also added to the study. The impact of using several optimizers will be investigated for different pre-trained models and different datasets.

REFERENCES

- [1] Singh, R., Ahmed, T., Kumar, A., Singh, A. K., Pandey, A. K., & Singh, S. K. (2020). Imbalanced breast cancer classification using transfer learning. *IEEE/ACM transactions on computational biology and bioinformatics*, 18(1), 83-93.
- [2] Malik, H., & Anees, T. (2022). BDCNet: Multi-classification convolutional neural network model for classification of COVID-19, pneumonia, and lung cancer from chest radiographs. *Multimedia Systems*, 28(3), 815-829.
- [3] Chandrasekaran, G., Antoanela, N., Andrei, G., Monica, C., & Hemanth, J. (2022). Visual sentiment analysis using deep learning models with social media data. *Applied Sciences*, 12(3), 1030.
- [4] Saeed, N., Nyberg, R. G., & Alam, M. (2022). Gravel road classification based on loose gravel using transfer learning. *International Journal of Pavement Engineering*, 1-8.

- [5] Khan, A., Hassan, B., Khan, S., Ahmed, R., & Abuassba, A. (2022). DeepFire: A novel dataset and deep transfer learning benchmark for forest fire detection. *Mobile Information Systems*, 2022.
- [6] Guo, S., & Wang, Q. (2022). Application of knowledge distillation based on transfer learning of ERNIE model in intelligent dialogue intention recognition. *Sensors*, 22(3), 1270.
- [7] Gujjar, J. P., Kumar, H. P., & Chiplunkar, N. N. (2021). Image classification and prediction using transfer learning in colab notebook. *Global Transitions Proceedings*, 2(2), 382-385.
- [8] Naushad, R., Kaur, T., & Ghaderpour, E. (2021). Deep transfer learning for land use and land cover classification: A comparative study. *Sensors*, 21(23), 8083.
- [9] Agarwal, D., Marques, G., de la Torre-Díez, I., Franco Martín, M. A., García Zapirain, B., & Martín Rodríguez, F. (2021). Transfer learning for Alzheimer's disease through neuroimaging biomarkers: a systematic review. *Sensors*, 21(21), 7259.
- [10] Morid, M. A., Borjali, A., & Del Fiol, G. (2021). A scoping review of learning research on medical image analysis using ImageNet. *Computers in biology and medicine*, 128, 104115.
- [11] Li, Y., Fu, B., Sun, X., Fan, D., Wang, Y., He, H., ... & Yao, Y. (2022). Comparison of Different Transfer Learning Methods for Classification of Mangrove Communities Using MCCUNet and UAV Multispectral Images. *Remote Sensing*, 14(21), 5533.
- [12] Alruily, M., Manaf Fazal, A., Mostafa, A. M., & Ezz, M. (2023). Automated Arabic long-tweet classification using transfer learning with BERT. *Applied Sciences*, 13(6), 3482.
- [13] Kandel, I., & Castelli, M. (2020). Transfer learning with convolutional neural networks for diabetic retinopathy image classification. A review. *Applied Sciences*, 10(6), 2021.
- [14] Tsalera, E., Papadakis, A., & Samarakou, M. (2021). Comparison of pre-trained CNNs for audio classification using transfer learning. *Journal of Sensor and Actuator Networks*, 10(4), 72.
- [15] Karthikeyan, D., Varde, A. S., & Wang, W. (2020, December). Transfer learning for decision support in Covid-19 detection from a few images in big data. In *2020 IEEE International Conference on Big Data (Big Data)* (pp. 4873-4881). IEEE.
- [16] VijayaKumari, G., Vutkur, P., & Vishwanath, P. (2022). Food classification using transfer learning technique. *Global Transitions Proceedings*, 3(1), 225-229.
- [17] Yasmin, F., Hassan, M. M., Hasan, M., Zaman, S., Kaushal, C., El-Shafai, W., & Soliman, N. F. (2023). PoxNet22: A fine-tuned model for the classification of monkeypox disease using transfer learning. *IEEE Access*, 11, 24053-24076.
- [18] Abubakar, U. B., Boukar, M. M., & Adeshina, S. (2022). Evaluation of Parameter Fine-Tuning with Transfer Learning for Osteoporosis Classification in Knee Radiograph. *International Journal of Advanced Computer Science and Applications*, 13(8).
- [19] Manna, A., Upasani, N., Jadhav, S., Mane, R., Chaudhari, R., & Chatre, V. (2023). Bird Image Classification using Convolutional Neural Network Transfer Learning Architectures. *International Journal of Advanced Computer Science and Applications*, 14(3).
- [20] Sunyoto, A., Pristyanto, Y., Setyanto, A., Alarfaj, F., Almusallam, N., & Alreshoodi, M. (2022). The Performance Evaluation of Transfer Learning VGG16 Algorithm on Various Chest X-ray Imaging Datasets for COVID-19 Classification. *International Journal of Advanced Computer Science and Applications*, 13(9).
- [21] Zhu, W., Braun, B., Chiang, L. H., & Romagnoli, J. A. (2021). Investigation of transfer learning for image classification and impact on training sample size. *Chemometrics and Intelligent Laboratory Systems*, 211, 104269.
- [22] Bichri, H., Chergui, A., & Hain, M. (2023). Image Classification with Transfer Learning Using a Custom Dataset: Comparative Study. *Procedia Computer Science*, 220, 48-54.
- [23] Rác, A., Bajusz, D., & Héberger, K. (2021). Effect of dataset size and train / test split ratios in QSAR/QSPR multiclass classification. *Molecules*, 26(4), 1111.
- [24] Pramudhita, D. A., Azzahra, F., Arfat, I. K., Magdalena, R., & Saidah, S. (2023). Strawberry Plant Diseases Classification Using CNN Based

- on MobileNetV3-Large and EfficientNet-B0 Architecture. *Jurnal Ilmiah Teknik Elektro Komputer dan Informatika*, 9(3), 522-534.
- [25] Dinata, M. I., Nugroho, S. M. S., & Rachmadi, R. F. (2021, June). Classification of strawberry plant diseases with leaf image using CNN. In *2021 International Conference on Artificial Intelligence and Computer Science Technology (ICAICST)* (pp. 68-72). IEEE.
- [26] Shin, J., Chang, Y. K., Heung, B., Nguyen-Quang, T., Price, G. W., & Al-Mallahi, A. (2021). A deep learning approach for RGB image-based powdery mildew disease detection on strawberry leaves. *Computers and electronics in agriculture*, 183, 106042.
- [27] Asif, D., Bibi, M., Arif, M. S., & Mukheimer, A. (2023). Enhancing Heart Disease Prediction through Ensemble Learning Techniques with Hyperparameter Optimization. *Algorithms*, 16(6), 308.
- [28] Muraina, I. (2022). Ideal dataset splitting ratios in machine learning algorithms: general concerns for data scientists and data analysts. In *7th International Mardin Artuklu Scientific Research Conference* (pp. 496-504).
- [29] Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L. C. (2018). Mobilenetv2: Inverted residuals and linear bottlenecks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 4510-4520).
- [30] Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.
- [31] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Identity mappings in deep residual networks. In *Computer Vision—ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part IV 14* (pp. 630-645). Springer International Publishing.
- [32] Baker, S., & Kandasamy, Y. (2023). Machine learning for understanding and predicting neurodevelopmental outcomes in premature infants: a systematic review. *Pediatric Research*, 93(2), 293-299.