RMIT

UNIVERSITY

COSC2753 Machine Learning - Assignment 2

Group Report

May 16, 2025

Group: S3_G5
Members:
- Nguyen Vu Trong Nhan - s4028019@rmit.edu.vn
- Le Duy Quang - s3912105@rmit.edu.vn
- Do Khoa Nguyen - s3978796@rmit.edu.vn
- Duong Hoang Anh Khoa - s3864111@rmit.edu.vn

1. Exploratory Data Analysis

1.1. Metadata and Images Validation

Here, metadata and images are validated to ensure they have the correct number of relations in-between. Missing values in metadata are checked to be **non-empty**, so imputation is not needed.

1.2. Statistical Overview

There are **10 unique `label` types** and **10 `variety` types**, with `age` being in range of 45 to 82. With that, this data could be categorized into **categorical** and **discrete** feature types, where `variety` and `label` being limited - thus **categorial**, and `age` being **discrete** as it is countable.

1.2.1. Data Shape

Visualize the dataset with distribution of each features in Figure 1, observations could be seen:

- `age`: fairly Gaussian-like distribution, no outliers, and a suitable inter-quartile range.
- `disease`: about most dataset has high contribution, while other disease contribute less than half.
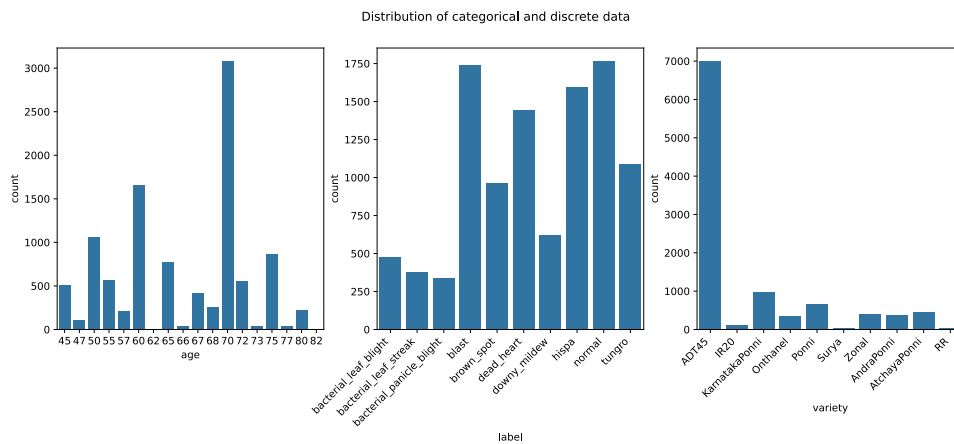- `variety`: one feature (ADT45) is the majority gaining 67% contribution of all plant types.

Figure 1: Distribution of `age`, `disease` and `variety` in dataset.

1.2.2. Relationships between Features

Factorize categorial values and plot correlation heat-map, features seems to correlate quite poorly.

When compare `age` to disease `label`, unhealthy disease tend to has larger inter-quartile range, while healthy `label` has small variation and small amount of outliers could be seen. Compare `age` to `variety`, ADT45 has more distributed data while some variety are found to has only one `age` value, and others has more "controlled" values.

Finally, the contribution of data-points of ADT45 to the dataset, most of data comes from only one plant `variety`, and its distribution could be spot as the majority. By that, further resampling method might be used to improve the performance of trained models later on.

1.3. Image Analysis

1.3.1. Metadata and Duplicates

Common issues on images are analyzed with CleanVision [1] to ensure they have standard lighting, identical aspect-ratio and image size, and image duplicates. Here, it is found to have identical aspect ratio and size, however consists of around 400 duplicates. Moreover, some of the duplicate set are found to consists of different metadata.

Removing the duplicates and unmatched metadata at this stage saw a minor dataset reduction of 2%.

1.3.2. Color Featuring

Analyzing mean color distributions as shown in Figure 2, in higher color intensity ($x > 150$), the green color-space is the majority due to the fact that images are capture of plants [2].
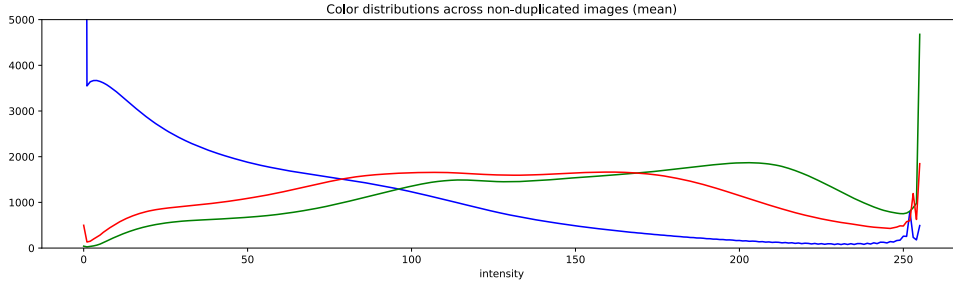

Figure 2: Mean color distribution of non-duplicated images in dataset

Sample some images that has low green intensity, due to environmental color changes or capturing with incorrect white balance setting resulting in color has more significant blue hues [3]; experiment could be conduct to see if model performance would improve with some of the images removed.
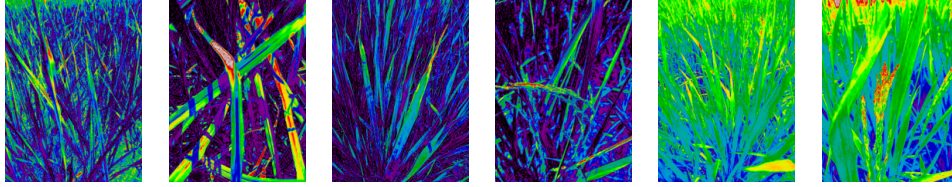

Figure 3: Image samples after applying `nipy` color-map

Flatten out the color-space and apply `nipy` color-map [4], the intensity of the red channel could be highlight and seen as a differentiate factor that could plays a major role in training our model.

Experimenting with a single color-channel disabled [5], information about the image could be gain in small edges, especially `brown_spot` where clear distinction of colors near the edge could be seen.

1.4. Handling the Imbalance

As `label` and `variety` are imbalanced, SMOTE [6] is performed to over-sample the dataset [7].

Here, over-sampling is chosen due to the large gap in both features, where under-sampling would remove many samples of the dataset and might become a factor in training. Moreover, the decision to use SMOTE over ADASYN [8] is due to overall performance in many researches [9], [10].

2. Training

2.1. Task 1. Paddy Multi-class Label Classification

2.1.1. Data Preparation

We perform training with image at original size of `640x480` and square `256x256` size, to assess the impact of spatial dimensions on classification accuracy. For image color channel diversity, we tested images with standard RGB, red-removed (G+B), green-removed (R+B), blue-removed (R+G), and spectral projection. We evaluate the original train images with the variant of colors, while generate an addition non-imbalance SMOTE [6] dataset taken from the train validation split to handle up-sampling.

2.1.2. Model Architecture and Justification

ResNet [11], AlexNet [12], ConvNext [13] and MobileNetV3 [14] were tested across various color channels. Standard RGB images results were at least 10% better than the remaining channel in the majority of model. While MobileNetV3 showed signs of overhitting, ConvNeXT was resource-

intensive with limited benefits. On the other hand, AlexNet and ResNet offered solid accuracy with minimal over-fitting, making them the most suitable option.

From these two, we enhanced AlexNet with Inception-style parallel branches. As a result, accuracy was improved by 7%, peaking at 90% with only 0.2 in loss. Furthermore, adding a Squeeze-and-Excitation (SE) [15] block further refined feature attention, stabilized training after initial fluctuations and achieved 93.953% validation accuracy with reduced over-fitting.

Finally, Grad-CAM [16] was applied to guide feature selection and select augmentation strategy. Despite slightly drop in accuracy, using coarse dropout from Albumentations [17] method helps improve real usage scenario generalization. Additionally, we experimented with SMOTE-based oversampling, but it significantly reduced the model's accuracy and increased over-fitting, suggesting limited effectiveness in this context.

2.1.3. Performance Evaluation

After 50 epochs, the model achieved strong performance with a validation accuracy of 91.52% and training accuracy of 90.96%, indicating good generalization. The close loss values (training: 0.2733, validation: 0.2866) further confirm stable learning. We observe that the all the weight average from precision, recall, and f1-score all reach 0.92.
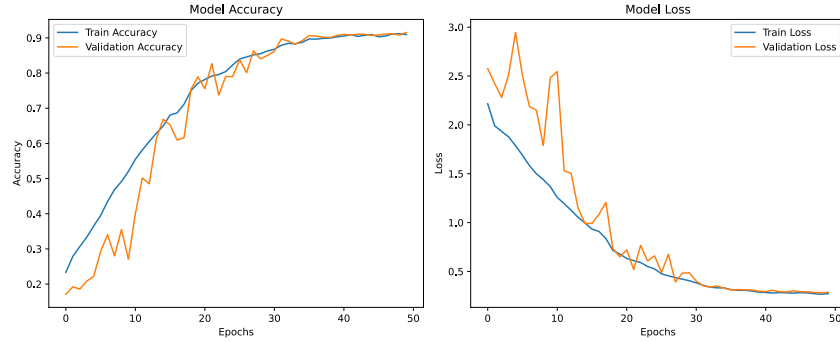

Figure 4: Epoch Accuracy

2.1.4. Ultimate Judgement

Testing with different models with different color channels at the beginning help understand the model strength as well as the effectiveness of that color channel among the tested models. Thus, the choice of AlexNet become more robust for furthure examination and development. The final model in combination of AlexNet, Inception technique, and SE block has significantly increase compared to the orignial AlexNet model by up to 7% in validation accuracy and achieve better generalization.

2.2. Task 2. Paddy Variety Classification using 4-Stack Color CNN

2.2.1. Data Preparation

At this stage, images were resized into 128x128, and categorical features was numerically encoded.

2.2.2. Model Architecture and Justification

MobilenetV2, VGG16, and a custom-built Convolutional Neural Network (CNN) were implemented for model selection phase. Both MobileNetV2 [18] and VGG16 [19] shared limitations in performance due to their capability of only accepting only 3-channel RGB inputs. While these two pretrained models offered strong transfer learning capabilities [20], they were not capable of leveraging the additional spectral channel (nipy_spectral) that are provided by our data, they tended to underperform, particularly on minority classes.

In contrast, CNN was designed to support 4-channel input [21], in this case stacking RGB and spectral images allowing it to extract richer features across visual and spectral domains. Hence, The 4-stack color CNN was selected as the most suitable architecture for our data.

2.2.3. Performance Evaluation

After training for 27 epochs which is due to early stopping implemented, the final result were:

```
loss: 0.1358 - accuracy: 0.9524 - val_loss: 0.2974 - val_accuracy: 0.9289
```

The small gap between 95.2% of training accuracy and 92.9% of validation accuracy indicates that the model generalizes well without significant over-fitting. The moderate validation loss (0.2974) suggests that the model assigning high probability to the correct class, aligning with expected performance for moderate-complexity CNNs trained on limited-size datasets without pretraining.
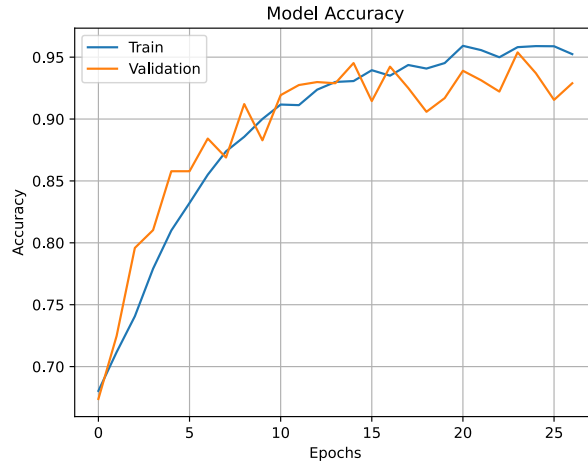

Figure 5: Epoch Accuracy

2.2.4. Ultimate Judgement

The training history, confusion matrix and the classification report were used to evaluate the 4-channel CNN Model. Training and validation accuracy steadily improved, reaching 95.7% and 94% respectively, with minimal divergence, indicating good generalization and effective regularization via data augmentation and early stopping. The confusion matrix shows strong dominant classes in ADT45 and IR20, on the other hand, likely due to visually similarity, minor confusion were shown in class like Ponni and RR, confirmed with the classification report's macros F1-score of 0.87 and weighted average of 0.94, reflecting balanced performance across all classes.

From this, we conclude that the model is well-generalized, handles class imbalance robustly, and is suitable for practical deployment in rice variety classification tasks. However, the model's performance on classes with very few samples such as Surya and RR still lags slightly, indicating a limitation in data availability rather than the architecture's capability. Nevertheless, this custom 4-channel CNN is the most effective model for our problem space considered on the accuracy, class balance, training stability and the efficiency of the model.

2.3. Task 3. Age Estimation using EfficientNetB0

2.3.1. Data Preparation

The pipeline loads metadata from `meta_train.csv` and collects valid (`image_path`, `age`) pairs from spectral images (`nipy_spectral.jpg`) stored in labeled subdirectories. Only images with matching metadata are included for training.

2.3.2. Model Architecture and Justification

InceptionV3 showed high memory usage and slow convergence with minimal performance gains for age regression. We switched to EfficientNetB0 for its lightweight design, faster processing, and lower resource consumption, making it a more practical and effective choice.

We use a convolutional neural network based on EfficientNetB0 as the backbone architecture. This model was chosen for its high performance-to-parameter ratio and its efficient scaling of depth, width, and resolution [22]. The EfficientNetB0 model is used without pretrained weights (weights=None) to allow full training on our domain-specific dataset.

Architecture Overview Input Layer: 128×128×3, EfficientNetB0 Base: Feature extractor without top classification layers, Two dense layers (256 & 128 units, ReLU) with L2 regularization, Output Layer: Single unit (linear activation) for age regression

This strategy gradually reduces the learning rate during training to avoid overshooting minima and helps with model convergence [23].

The model was trained using the **Adam** optimizer and **Mean Absolute Error (MAE)** as both the loss function and evaluation metric, which is suitable for continuous regression tasks like age prediction.

2.3.3. Performance Evaluation

After training for 60 epochs, the final results are:
- Training MAE: 3.78
- Validation MAE: 5.53

The low difference between training and validation MAE indicates the model generalizes relatively well. This level of MAE is consistent with baseline performance in age estimation tasks from facial or biological imagery when not using pretrained weights or large datasets.
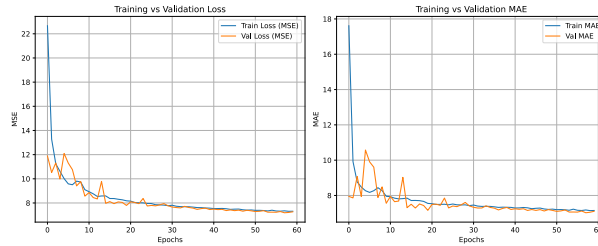

Figure 6: Epoch logs showing steady decrease in MAE and loss.

2.3.4. Ultimate Judgement

The model demonstrates strong potential for age prediction from spectral images. Achieving a validation MAE of 5.53 is promising, especially given that:
- No pretrained weights were used,
- The image resolution was relatively low (128×128),
- Only basic preprocessing was applied.

These results suggest that the current pipeline provides a solid foundation for further improvement. With fine-tuning, transfer learning, and additional data or augmentation, the model could potentially reach a lower MAE, which is competitive in age regression tasks.

Appendix

Rice Plant Disease Prediction

Age: 68.88999938964844
Label: normal
Variety: AndraPonni

↑
200009.jpg

Predict

Figure 7: Single-page application for onsite model prediction

image_id	label	variety	age
200001.jpg	dead_heart	ADT45	69.30999755859375
200002.jpg	normal	ADT45	67.41000366210938
200003.jpg	normal	ADT45	66.9800033569336
200004.jpg	dead_heart	Zonal	68.8499984741211
200005.jpg	blast	AndraPonni	35.88999938964844
200006.jpg	normal	ADT45	63.52000045776367
200007.jpg	normal	ADT45	68.08999633789062
200008.jpg	dead_heart	ADT45	65.66999816894531
200009.jpg	normal	AndraPonni	68.88999938964844
200010.jpg	normal	ADT45	64.36000061035156
200011.jpg	dead_heart	ADT45	69.4000015258789
200012.jpg	normal	KarnatakaPonni	53.970001220703125
200013.jpg	normal	Ponni	70.04000091552734
200014.jpg	normal	IR20	69.36000061035156
200015.jpg	normal	ADT45	57.310001373291016
200016.jpg	normal	ADT45	69.0999984741211
200017.jpg	dead_heart	ADT45	67.12999725341797
200018.jpg	normal	ADT45	59.52000045776367
200019.jpg	normal	ADT45	68.93000030517578

Table 1: First 20 results of model predictions

References

[1] Cleanlab, "Cleanlab/cleanvision: Automatically find issues in image datasets and practice data-centric computer vision.." [Online]. Available: https://github.com/cleanlab/cleanvision

[2] O. Yaman and T. Tuncer, "Plant Classification Method Using Histogram and Machine Learning for Smart Agriculture Applications," *Acta Infologica*, vol. 7, no. 1, pp. 17–28, 2024, doi: 10.26650/acin.1070261.

[3] M. Afifi, B. Price, S. Cohen, and M. S. Brown, "When Color Constancy Goes Wrong: Correcting Improperly White-Balanced Images," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2019.

[4] K. Gorgolewski *et al.*, "Nipype: A Flexible, Lightweight and Extensible Neuroimaging Data Processing Framework in Python," *Frontiers in Neuroinformatics*, 2011, doi: 10.3389/fninf.2011.00013.

[5] E. Hamuda, M. Glavin, and E. Jones, "A survey of image processing techniques for plant extraction and segmentation in the field," *Computers and Electronics in Agriculture*, vol. 125, pp. 184–199, 2016, doi: https://doi.org/10.1016/j.compag.2016.04.024.

[6] K. W. Bowyer, N. V. Chawla, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic Minority Over-sampling Technique," *CoRR*, 2011, [Online]. Available: http://arxiv.org/abs/1106.1813

[7] G. Lemaître, F. Nogueira, and C. K. Aridas, "Imbalanced-learn: A Python Toolbox to Tackle the Curse of Imbalanced Datasets in Machine Learning," *Journal of Machine Learning Research*, vol. 18, no. 17, pp. 1–5, 2017, [Online]. Available: http://jmlr.org/papers/v18/16-365

[8] H. He, Y. Bai, E. A. Garcia, and S. Li, "ADASYN: Adaptive synthetic sampling approach for imbalanced learning," in *2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence)*, 2008, pp. 1322–1328. doi: 10.1109/IJCNN.2008.4633969.

[9] J. Brandt and E. Lanzén, "A Comparative Review of SMOTE and ADASYN in Imbalanced Data Classification," 2021. [Online]. Available: https://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-432162

[10] I. Dey and V. Pratap, "A Comparative Study of SMOTE, Borderline-SMOTE, and ADASYN Oversampling Techniques using Different Classifiers," in *2023 3rd International Conference on Smart Data Intelligence (ICSMDI)*, 2023, pp. 294–302. doi: 10.1109/ICSMDI57622.2023.00060.

[11] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778. doi: 10.1109/CVPR.2016.90.

[12] Z. Li *et al.*, "Improved AlexNet with Inception-V4 for Plant Disease Diagnosis," *Computational intelligence and neuroscience*, vol. 2022, no. 1, p. 5862600, 2022.

[13] A. T. N. Rohman, F. Sthevanie, and K. N. Ramadhani, "Classification of Mango Plant Leaf Diseases Using Optimized ConvNeXt," in *2024 International Conference on Intelligent Cybernetics Technology & Applications (ICICyTA)*, 2024, pp. 443–448.

[14] S. Qian, C. Ning, and Y. Hu, "MobileNetV3 for Image Classification," in *2021 IEEE 2nd International Conference on Big Data, Artificial Intelligence and Internet of Things Engineering (ICBAIE)*, 2021, pp. 490–497. doi: 10.1109/ICBAIE52039.2021.9389905.

[15] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 7132–7141.

[16] R. R. Selvaraju, A. Das, R. Vedantam, M. Cogswell, D. Parikh, and D. Batra, "Grad-CAM: Why did you say that?," *arXiv preprint arXiv:1611.07450*, 2016.

[17] A. Buslaev, V. I. Iglovikov, E. Khvedchenya, A. Parinov, M. Druzhinin, and A. A. Kalinin, "Albumentations: fast and flexible image augmentations," *Information*, vol. 11, no. 2, p. 125, 2020.

[18] K. Team, "MobileNetV2 - Keras Applications." 2024.

[19] K. Team, "VGG16 - Keras Applications." 2024.

[20] F. D. Adhinata, N. A. F. Tanjung, W. Widayat, G. R. Pasfica, and F. R. Satura, "Comparative Study of VGG16 and MobileNetV2 for Masked Face Recognition," *Jurnal Ilmiah Teknik Elektro Komputer dan Informatika*, vol. 7, no. 2, p. 230, Jul. 2021, doi: 10.26555/jiteki.v7i2.20758.

[21] R. Gopalapillai, D. Gupta, M. Zakariah, and Y. A. Alotaibi, "Convolution-based encoding of depth images for transfer learning in RGB-D scene classification," *Sensors (Basel)*, vol. 21, no. 23, p. 7950, Nov. 2021.

[22] M. Tan and Q. V. Le, "EfficientNet: Rethinking model scaling for convolutional Neural Networks," *arXiv [cs.LG]*, 2019.

[23] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016.