Comparison of Classification Methodologies using Convolutional Neural Networks in a Dataset of Plant Leaf Diseases

Capstone Project for Master of Science in Data Analytics

Ruairi O’Donohoe – sba23348

Image Classification

- Methodologies
  - Multi-class (MC): Classify an image with a single predicted label
  - Multi-label (ML): Classify an image with multiple potential label
- Common Parameters:
  - See Table below (Lydia 2020)
  - Differences in calculation methods of probability and metrics.
  - Parameters for ML are often inconsistent in research
  - Loss and Accuracy less suitable for comparison. F1 score used instead.

	Multi-Class	Multi-Label
Classifier function	Softmax	Sigmoid
Loss function	Categorical CE	Binary CE
Accuracy	Categorical Acc	Binary Acc
F1 Score	No threshold. Chooses argmax	0.5

FieldPlant research

(Moupojou 2023)

- Dataset:
  - ML dataset containing images of leaf disease.
  - Recent (2023*), in-field images, expertly annotated.
  - Class imbalance:
    - 27 disease classes
    - Top 3 classes combined: 58.7%
- Experiments
  - 4 CNN models: VGG16, MobileNetV2, InceptionV2, InceptionResNetV2
  - Transfer learning and fine-tuning.
  - Used MC methodology, despite ML dataset
  - No data stratification, despite class imbalance.
- Conclusion:
  - "... existing models are not sufficiently accurate for plant disease ... classification of images collected directly from the field, although the ... results for FieldPlant are better than those for PlantDoc."

*Correction of 2013 in thesis

Methodology

1. Stratify dataset
2. Train 4 models with each (2) methodology.
   - Transfer Learning (FT) and Fine tuning
   - Yielding 8 sets of results
3. Hyperparameter Tuning (HT):
   - Using best performing model.
   - RandomSearch algorithm
   - Train top layers only

Q1: Does dataset stratification improve the performance of an image classification CNN model?

- Experiment:
  - MobileNetV2, MC methodology.
  - Unstratified vs Stratified datasets.
- Results:
  - Both unstratified datasets outperformed the stratified.
- Conclusion
  - Possibly due to distribution of rare diseases in unstratified dataset.

Dataset	Categorical Acc	F1 Score
FP (unstratified)	82.9%	-
Unstratified	82.36%	81.37%
Stratified	78.40%	79.10%

Q2: Does using multi-label classification methodology improve performance on a multi-label dataset as compared to multi-class methodology?

- Experiment:
  - Train all 4 models with both methodologies
- Results
  - All ML models outperform their MC counterparts for Loss and Accuracy
  - Contrary for F1 Score
- Conclusion
  - Likely due to methods of calculations

Model	Method.	Loss	Acc.	F1
InceptionResNetV2	ML	0.05	98.47%	79.67%
	MC	0.63	85.58%	83.49%
InceptionV3	ML	0.04	98.67%	80.94%
	MC	0.61	86.63%	85.37%
MobileNetV2	ML	0.04	98.48%	77.41%
	MC	0.67	78.40%	78.10%
VGG16	ML	0.05	98.32%	78.95%
	MC	0.59	85.60%	83.61%

Q3: What is the best combination of model and methodology?

- Experiment:
  - Identify best model and methodology combination from Experiment 2.
- Results
  - Best models per metric all differ.
  - All ML models outperform all MC models for Loss and Accuracy.
  - All but one MC model outperforms all ML models for F1 Score.
  - Top model loss, is lowest model F1 Score.
- Conclusion
  - Best model depends on metric.
  - ML models performed poorly using F1 score.

	Model Loss		Model Accuracy		Model F1 score	
1	Mobilenet _ML	0.042	Inception _v3_ML	98.67%	Inception_ V3_MC	85.37%
2	Inception_ v3_ML	0.044	Mobilenet _ML	98.48%	vgg16 _MC	83.61%
3	Vgg16 _ML	0.050	Inception_ resnetv2_ML	98.47%	Inception_ resnetv2_MC	83.49%
4	Inception_ resnetv2_ML	0.052	vgg16 _ML	98.32%	Inception_ v3_ML	80.94%
5	Vgg16 _MC	0.594	Inception _v3_MC	86.63%	Inception_ resnetv2_ML	79.67%
6	Inception _v3_MC	0.607	vgg16 _MC	85.6%	vgg16 _ML	78.95%
7	Inception_ resnetv2_MC	0.634	Inception_ resnetv2_MC	85.58%	Mobilenet _MC	78.1%
8	Mobilenet _MC	0.674	Mobilenet _MC	78.4%	Mobilenet _ML	77.41%

Q4: Does hyperparameter tuning improve model performance over standard transfer learning?

- Experiment:
  - MobileNetV2 – best loss
  - RandomSearch algorithm
  - Compare model performance:
    - Standard Model (SM): TL and fine tuning
    - Hyperparameter Tuning (HT): Top layers only
- Results:
  - Best hyperparameters (Table 1)
  - HT model outperforms SM for all metrics (Table 2)
- Conclusions:
  - HT improves model performance.
  - Best parameters include Sigmoid activation
    - Search process based on optimizing loss

Hyperparameter	Selection
Activation	**Sigmoid**
Dropout Rate	0.2
Units	256
Epochs	47

Table 1: Best hyperparameters

Model	Loss	Acc.	F1
SM	0.042	98.48%	77.41%
HT	0.0394	98.54%	81.56%

Table 2: Compare model metrics

Conclusion

1. Data stratification did not improve performance.
2. ML models outperform MC models for Loss and Accuracy, but not F1 Score.
3. The best model depends on metric.
4. Hyperparameter tuning chose Sigmoid function. Outperformed standard model.

References

1. Lydia, A.A., Francis, F.S., 2020. Multi-Label Classification using Deep Convolutional Neural Network. 2020 Int. Conf. Innov. Trends Inf. Technol. ICITIIT 1–6. https://doi.org/10.1109/ICITIIT49094.2020.9071539

2. Moupojou, E., Tagne, A., Retraint, F., Tadonkemwa, A., Wilfried, D., Tapamo, H., Nkenlifack, M., 2023. FieldPlant: A Dataset of Field Plant Images for Plant Disease Detection and Classification With Deep Learning. IEEE Access PP, 1–1. https://doi.org/10.1109/ACCESS.2023.3263042

Acknowledgments

- Vikas Tomer - Thesis supervisor.
- Lecturers of MSc Data Analytics in CCT

THANK YOU