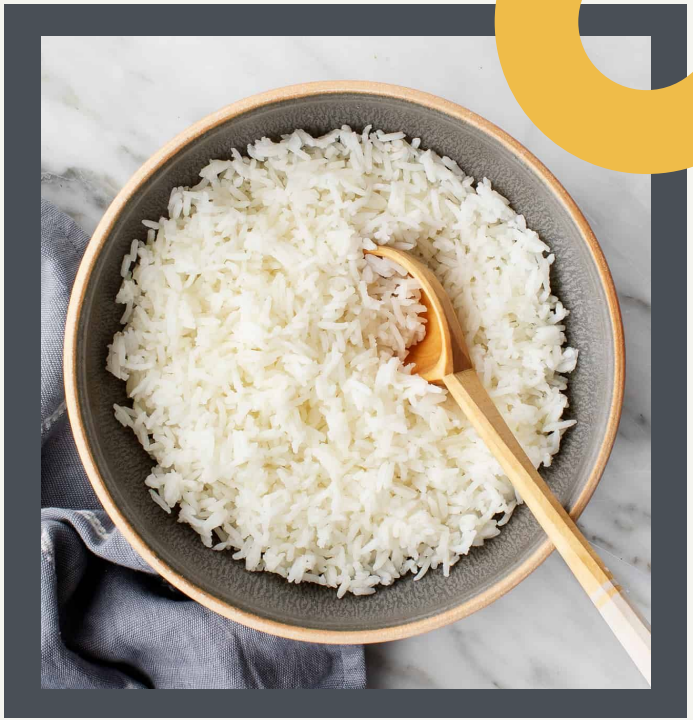
Fine Grain Analysis

A Project on Rice Classification

rice.jpeg



Amy Zhao Cassidy Xu Marios Tsotras Spencer Morga

Sup-Rice Pop Quiz!

Can you classify these grains of rice?

Arborio

Basmati

Ispala

Karacadag

Jasmine

WHY ARE WE DOING THIS?

LOGISTIC REGRESSION + KNN

FEATURE DATA PREPROCESSING

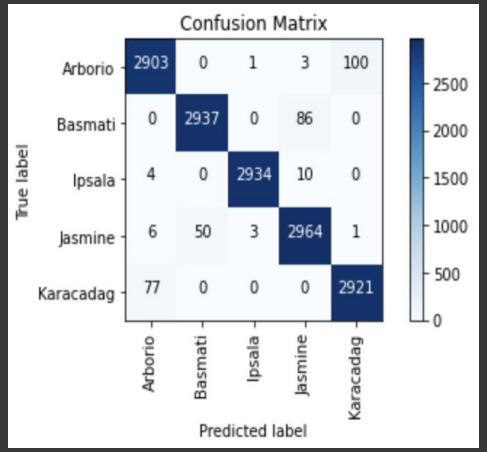
- 5 types: arborio, basmati, karacadag, ipsala, jasmine
- quantitative feature metrics measured from rice
- unnecessary features removed
- 60k - 15k train test split
- 3 models run - LR, LR-L2, kNN

PROCEDURE

- preprocess metrics dataset
- scale, impute, encode data with pipeline
- fit, train, and test log reg / KNN models
- create confusion matrices
- k-fold cross validation

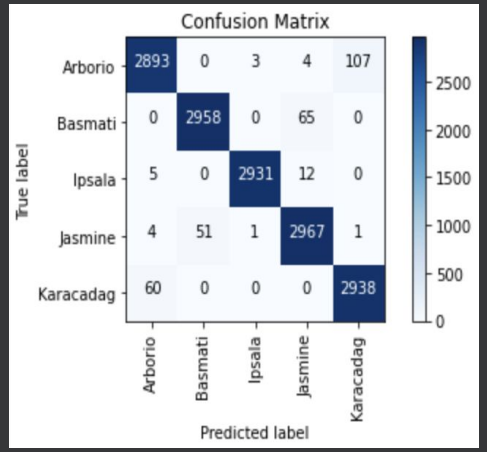
model performance

LOG REG NO PENALTY



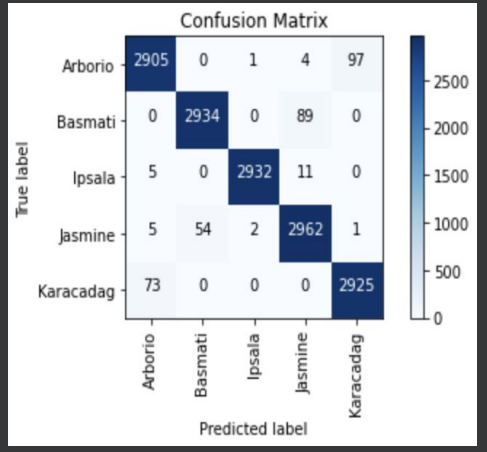
accuracy	0.977267
precision	0.977412
recall	0.977338
f-1 score	0.977355

LOG REG L2 PENALTY



accuracy	0.97720
precision	0.97736
recall	0.97727
f-1 score	0.97729

KNN CLASSIFIER



accuracy	0.979133
precision	0.979272
recall	0.979191
f-1 score	0.979204

k-fold cross validation

KNN mean accuracy : **97.94%**

Log Reg mean accuracy :
(no penalty) **97.76%**

kNN when k=5 performs best with accuracy of 97.91%

CONCLUSIONS

- Accurate, but impractical
- Hard to gather data
- Easy implementation

THE PROBLEM

We can analyze the data quantitatively, but can we figure out a way to visually classify the rice?

And is it better?

yes.

BENEFITS of visualization

- Practical
- Classification Accessibility
- Improves Accuracy

CONVOLUTIONAL NEURAL NETWORK

PROCEDURE

- gather image dataset
- convert all images to tensors
- split into train, test, and validation data
- create the convolutional neural network
- optimize then train the model
- evaluate data - loss and accuracy

model performance

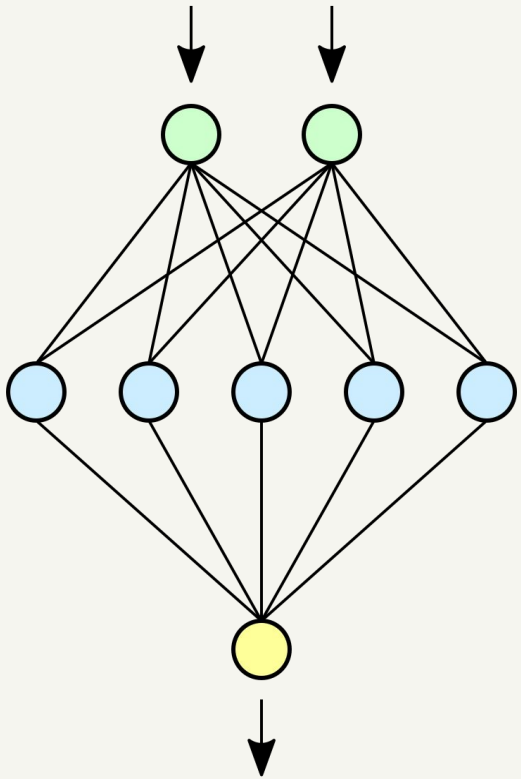
7,157,901 trainable parameters

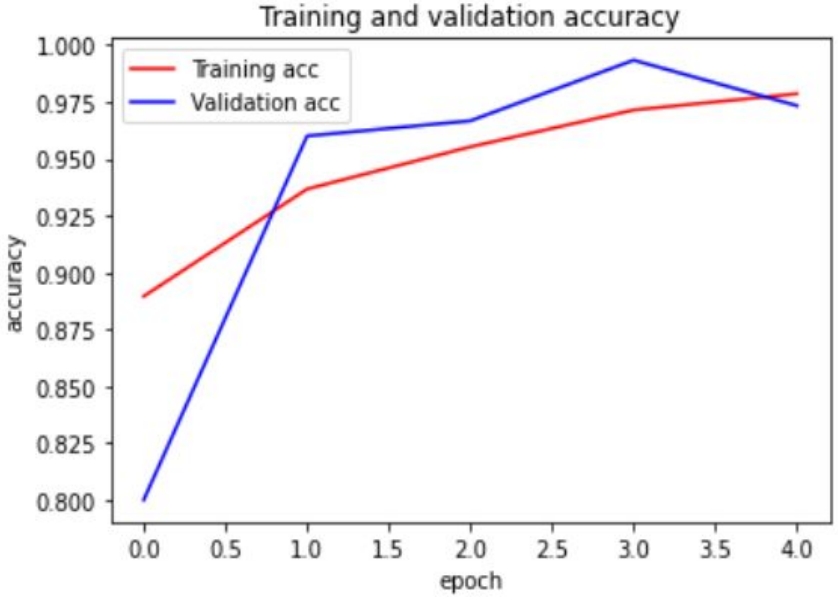
2500 total images

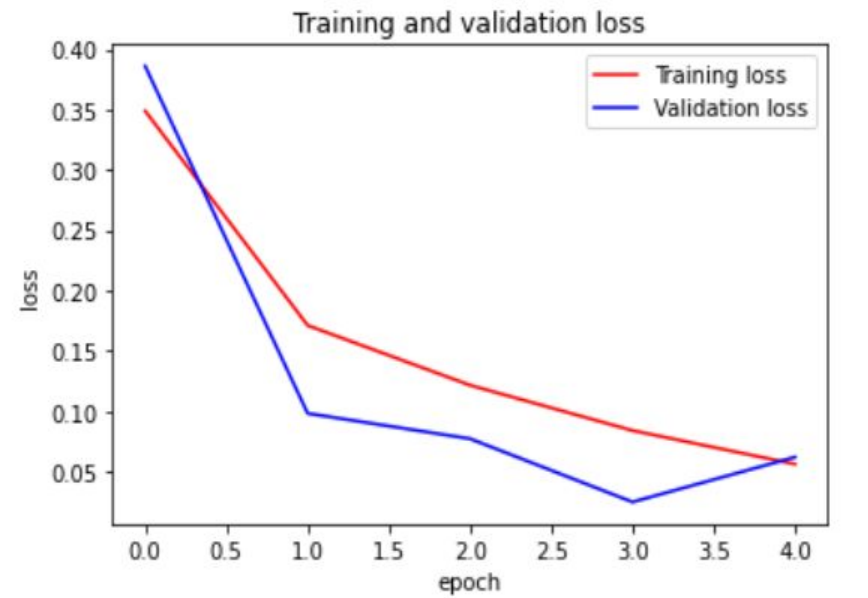
Cross entropy loss function

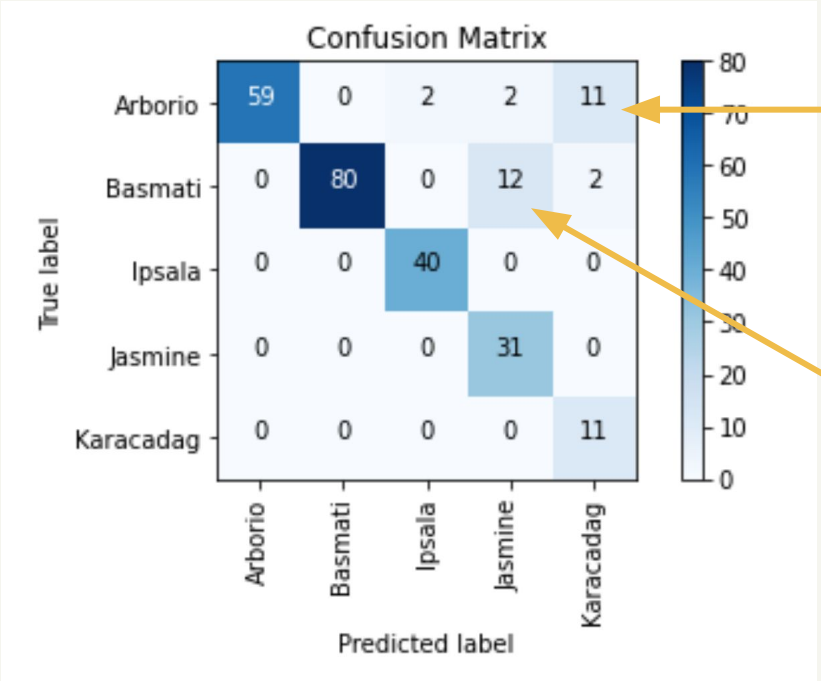
Test Accuracy: ***0.976000***

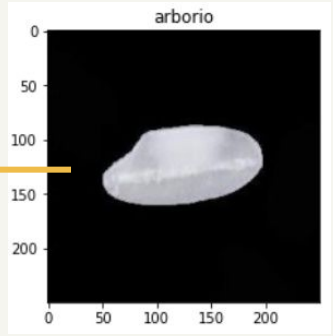
Test Loss: ***0.054742***



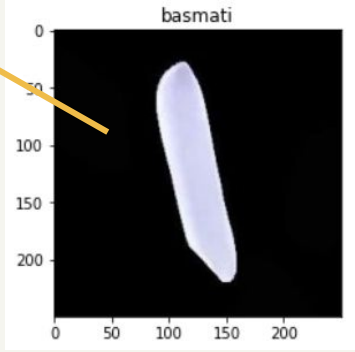

Training and validation accuracy
Training acc
Validation acc
accuracy
epoch
1.000
0.975
0.950
0.925
0.900
0.875
0.850
0.825
0.800
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0

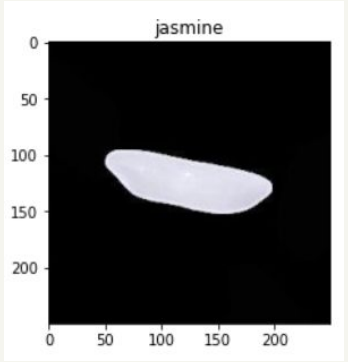

Training and validation loss
Training loss
Validation loss
loss
epoch
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0


Confusion Matrix
Arborio
Basmati
Ipsala
Jasmine
Karacadag
True label
Predicted label
59 0 2 2 11
0 80 0 12 2
0 0 40 0 0
0 0 0 31 0
0 0 0 0 11
80
70
60
50
40
30
20
10
0


arborio


karacadag


basmati


jasmine

model comparison

Quantitative vs Visual Classification

k-NN classification

Less data needed

Very quick implementation

Accurate Results

Accuracy: **0.9791**

Impractical Usage

VS

visual classification (cnn)

A lot more data is needed

Extremely long implementation

Accurate Results

Accuracy: **0.9760**

Practical Usage

model comparison

To Other Existing Classification Models

Rice Image NN Classification

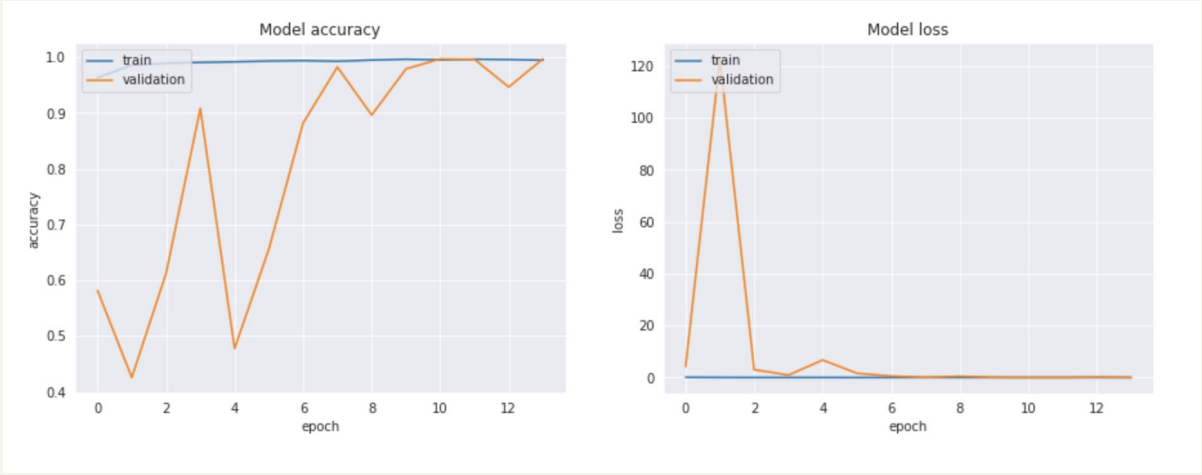
By Karl Adrian de Guzman (kaggle)

- 99.9% accuracy
- Utilizes transfer learning
- Utilizes dropout
    - Decrease overfitting
    - Maximize efficiency
- 4932 iterations across 12 epochs
- Batch size - 128 (larger than ours)

accuracy and **loss**

PRE

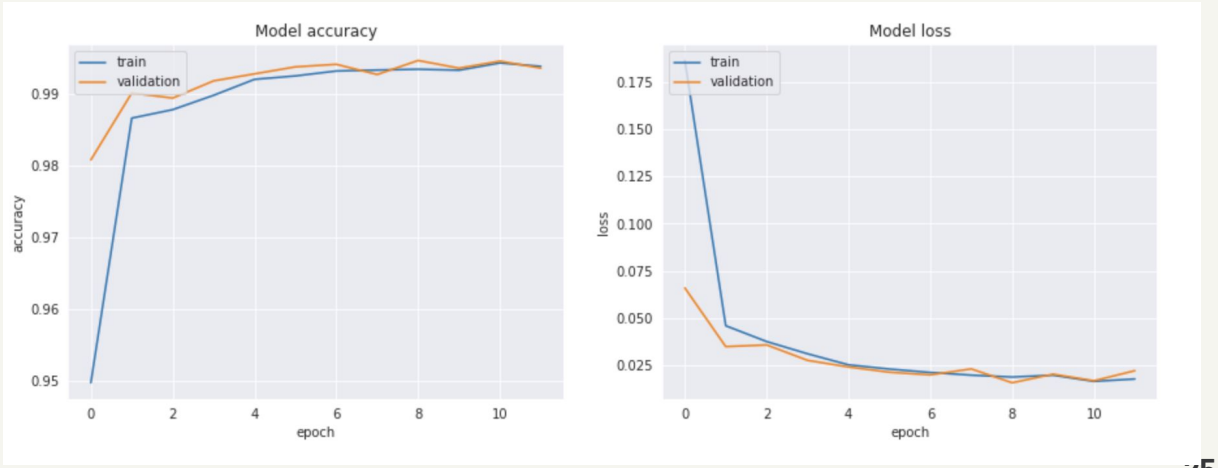
transfer learning and model evaluation



accuracy and **loss**

POST

transfer learning and model evaluation



CONCLUSIONS

- Best performing feature model : kNN
  - Accuracy: ***97.91%***
- Best performance image classification model: CNN
  - Accuracy: ***97.60%***
- However, CNN beats kNN in **practicality**
  - Accessibility to images is greater than a rice metric file
- Allows for **other visual detection**, not only classification
  - Color, unwanted substances, blemishes, etc

Thank you!

Questions?

References

de Guzman, Karl Adrian. “(99.9%) Rice Image Classification Using CNN in TF.” *Kaggle*, Kaggle, 25 Apr. 2022, https://www.kaggle.com/code/karladriandeguzman/99-9-rice-image-classification-using-cnn-in-tf.