NIVA Geotagged photos – Danish AI Challenge

Jappe Franke, Arun Pratihast, Rob Knapen

Wageningen Environmental Research

August 2022

Introduction

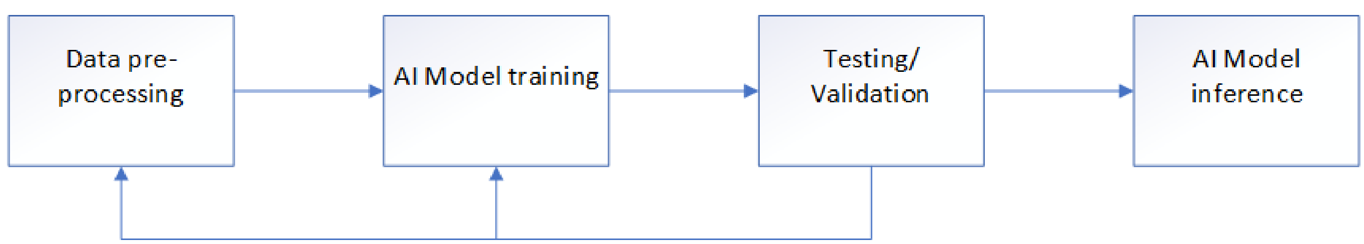
As part of the NIVA project, a platform for AI powered auto review and pre-validation of geo-tagged agriculture images for IACS (Integrated Administration and Control System) will be developed. This platform will be hosted at the Walton Institute on behalf of NIVA. The platform is split into two main aspects, a general facility for image uploading and tagging, and a facility for reviewing by a human or auto-reviewing by an AI model.

For the targeted initial three auto-review (AI) tasks, suitable auto-review challenges have been proposed by Paying Agencies. Including the Danish challenge for which the model training and validation is described in this document.

In Denmark it is already possible for farmers to use images from a geo-tagged app as prove the activity demand on livestock grazing on grasslands. However, this is resulting in many images being sent in and needing manual inspection for approval. A possible pre-selection by an AI model is helpful. The initial challenge for the AI model is straight-forward, a classification task of approval or non-approval of the grazing activity. Such a model might be relevant for other countries in Europe, especially those with a high percentage of grasslands.

Workflow

The functional features of the workflow include: 1) Data pre-processing; 2) AI Model training; 3) Testing/Validation; and 4) AI Model inference for the proper operationalization. The individual modules are shown below:

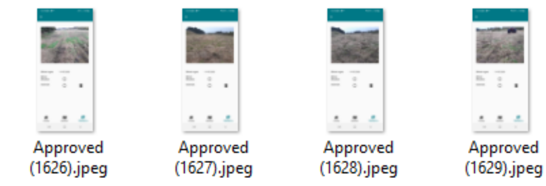


Data pre-processing

After obtaining the dataset from Denmark, the dataset of 5500 images of grazed (Approved) fields was prepared for training a Convolutional Neural Network (CNN) deep learning AI model.

Manual images clean-up

The dataset contained ~50 images that were not particularly useful for training. These images contained mainly images that were not of fields; these were removed. A few examples:



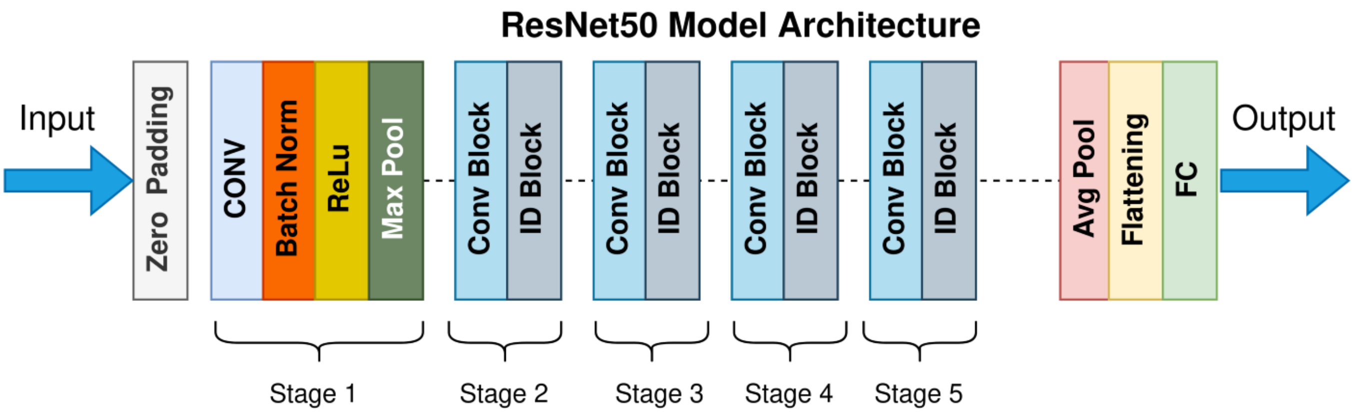
During examination of the available dataset, it was noted that it only contained about 38 photos from the 'Not Approved' category. The clear cause of this is that the farmers do not take such photos and send them in for validation, so they simply do not exist. However, this leaves far too few images in this category to expect any successful model training. As an alternative it was decided to use (1800) images from the Irish Clean Permanent Pasture (not grazed grassland) dataset instead.

Automated image cropping

All images selected for use were automatically cropped to a center-bottom image of 500 x 500 pixels, so that some noise was removed at the borders of the landscape and portrait images in the dataset. The reasoning behind it was that this would force the model to focus on the (expected) region of the photo that typically contains most of the 'grass' area, where it could learn the most usable patterns.

Model selection

Considering that the challenge is a computer vision task, the well-known ResNet50 convolutional neural network (CNN) model was selected for this initial experiment. The ResNet architecture (2015) is based on a series of convolutional and max pooling steps, and the first model to introduce skip connections. In general, it has good performance on classification tasks with a limited number of trainable parameters. Instead of training from scratch it is also possible to load ImageNet (https://www.image-net.org) based pre-trained parameters, which gives the model some initial knowledge about features visible in real-world photos.



The PyTorch (https://pytorch.org) open source machine learning framework has been used to train the model on the available images, and for inference (predictions). The Albumentations (https://albumentations.ai) Python package has been used as well for data augmentation.

Training

A ResNet50 model was trained on a GeForce RTX 2080 GPU (Graphics Processing Unit) for 20 epochs (iterations). The training took about 5 hours

Image augmentation (by adding additional variation in orientation, hue, saturation, brightness) was done on all the images, but augmentation did not improve the classification; another indication the dataset is very noisy.

After training the results were as follows;
The training dataset achieved an internal accuracy of 94.8%

Table 1 Model training confusion matrix

Training	**Approved**	**Clean Permanent Pasture**
Approved	3731	37
Clean Permanent Pasture	224	1008

Table 2 Model training performance metrics

Metrics	**Precision**	**Recall**	**F1-score**	**Support**
Approved	0.943	0.990	0.966	3768
Clean Permanent Pasture	0.965	0.818	0.885	1232
Total Accuracy			0.948	5000
Macro Average	0.954	0.904	0.926	5000
Weighted Average	0.949	0.948	0.946	5000

Validation

The validation dataset achieved an even higher accuracy: 95.4%

Table 3 Model validation confusion matrix

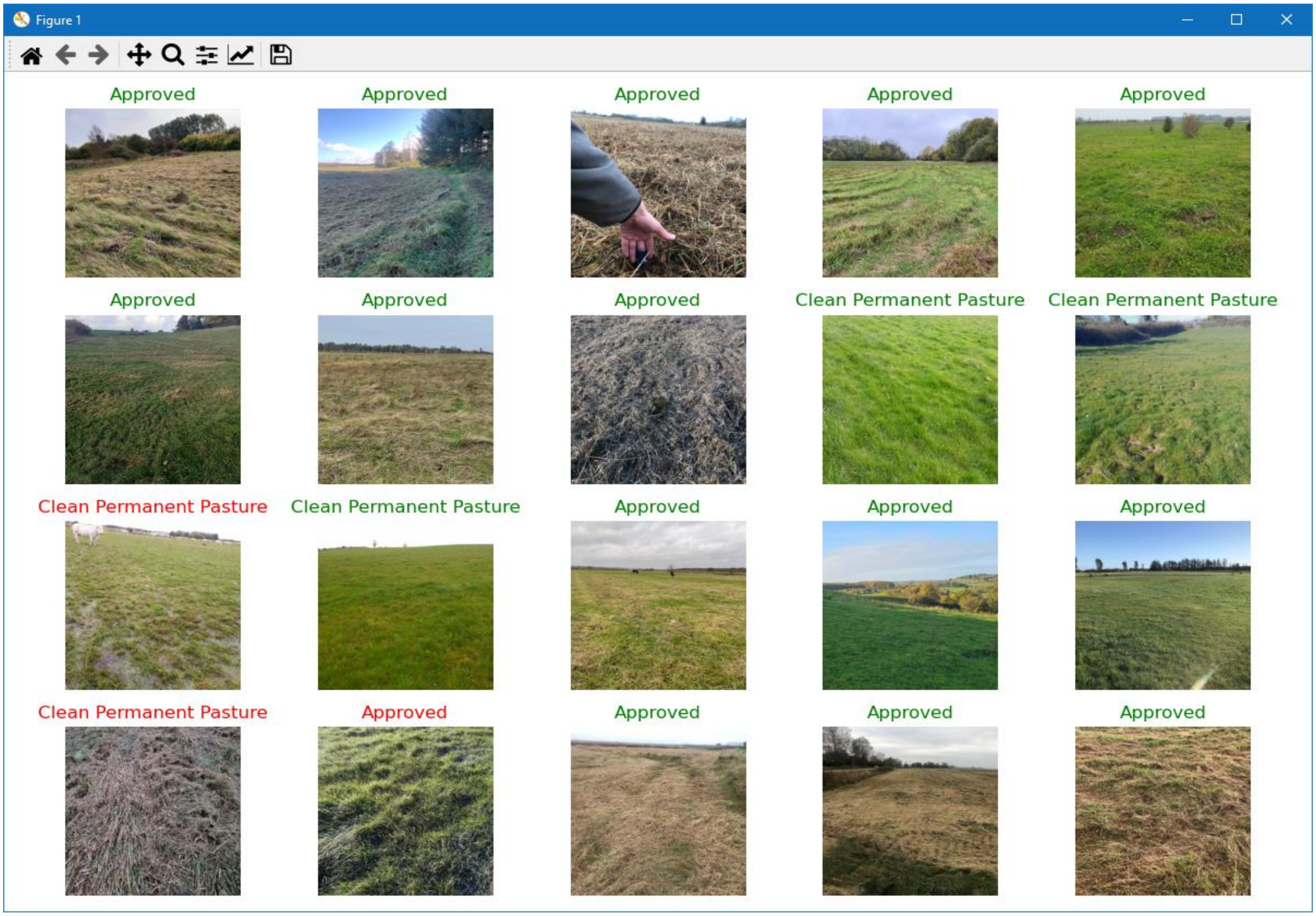
Validation	**Approved**	**Clean Permanent Pasture**
Approved	1745	7
Clean Permanent Pasture	99	435

Table 4 Model validation performance metrics

Metrics	**Precision**	**Recall**	**F1-score**	**Support**
Approved	0.946	0.996	0.971	1752
Clean Permanent Pasture	0.984	0.815	0.891	534
Total Accuracy			0.954	2286
Macro Average	0.965	0.905	0.931	2286
Weighted Average	0.955	0.954	0.952	2286

If this is the case the model is struggling to fit the training data properly. It is overfitting and probably needs clearer images of grazed and ungrazed fields to increase consistency.

Below are a few example images; red labels are falsely predicted images



Testing with local grassland images

Using a few images obtained locally, the model performed at only 34.3% accuracy:

```
local fields\Approved (Local – Approved in table below)
local fields\Clean Permanent Pasture (Local – Not Approved in table below)
```

Table 5 Confusion matrix for model predictions on local grassland images

Validation	**Local - Approved**	**Local – Not Approved**
Local - Approved	32	4
Local – Not Approved	61	2

Table 6 Model prediction performance on local grassland images

Metrics	**Precision**	**Recall**	**F1-score**	**Support**
Local - Approved	0.344	0.889	0.496	36
Local – Not Approved	0.333	0.032	0.058	63
Total Accuracy			0.343	99
Macro Average	0.339	0.460	0.277	99
Weighted Average	0.337	0.343	0.217	99

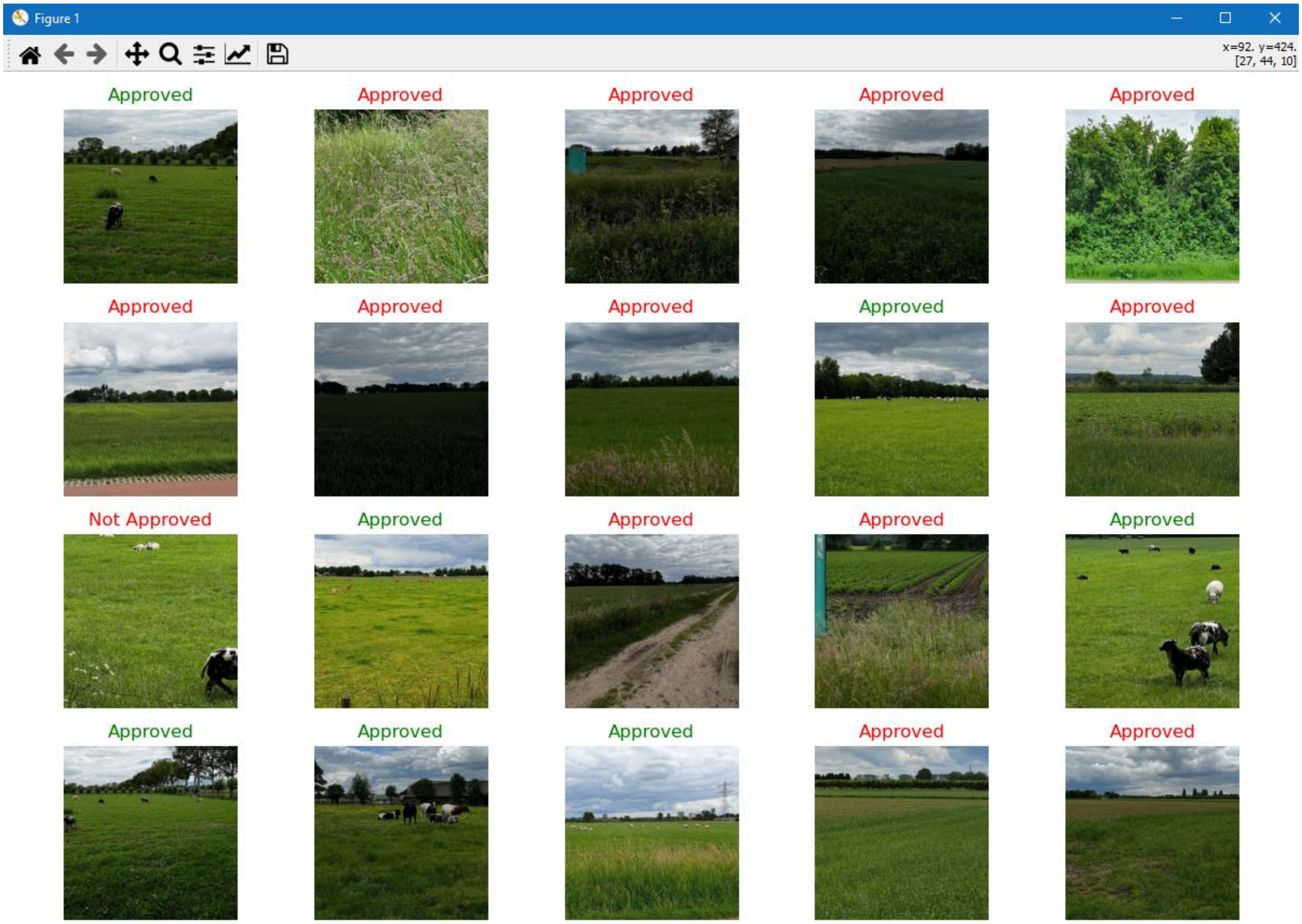


Figure 1
x=92. y=424.
[27, 44, 10]
Approved
Approved
Approved
Approved
Approved
Approved
Approved
Approved
Approved
Approved
Not Approved
Approved
Approved
Approved
Approved
Approved
Approved
Approved
Approved
Approved

Testing “Not Approved” Danish images

Testing against the 38 “Not Approved” Danish images showed 67.6% accuracy. The model was also trained on these images, so results are positively skewed.

```
67 Approved Danish images from original training set\Approved (Danish – Approved in table below)
38 Not Approved (Clean Permanent Pasture) Danish images from original training set\Not Approved (Danish – Not Approved in table below)
```

Table 7 Confusion matrix for model predictions on Danish Approved - Not Approved images

Validation	**Danish - Approved**	**Danish – Not Approved**
Danish - Approved	66	1
Danish – Not Approved	33	5

Table 8 Model prediction performance on Danish Approved - Not Approved images

Metrics	**Precision**	**Recall**	**F1-score**	**Support**
Danish - Approved	0.667	0.985	0.795	67
Danish – Not Approved	0.833	0.132	0.227	38
Total Accuracy			0.676	105
Macro Average	0.750	0.558	0.511	105
Weighted Average	0.727	0.676	0.590	105

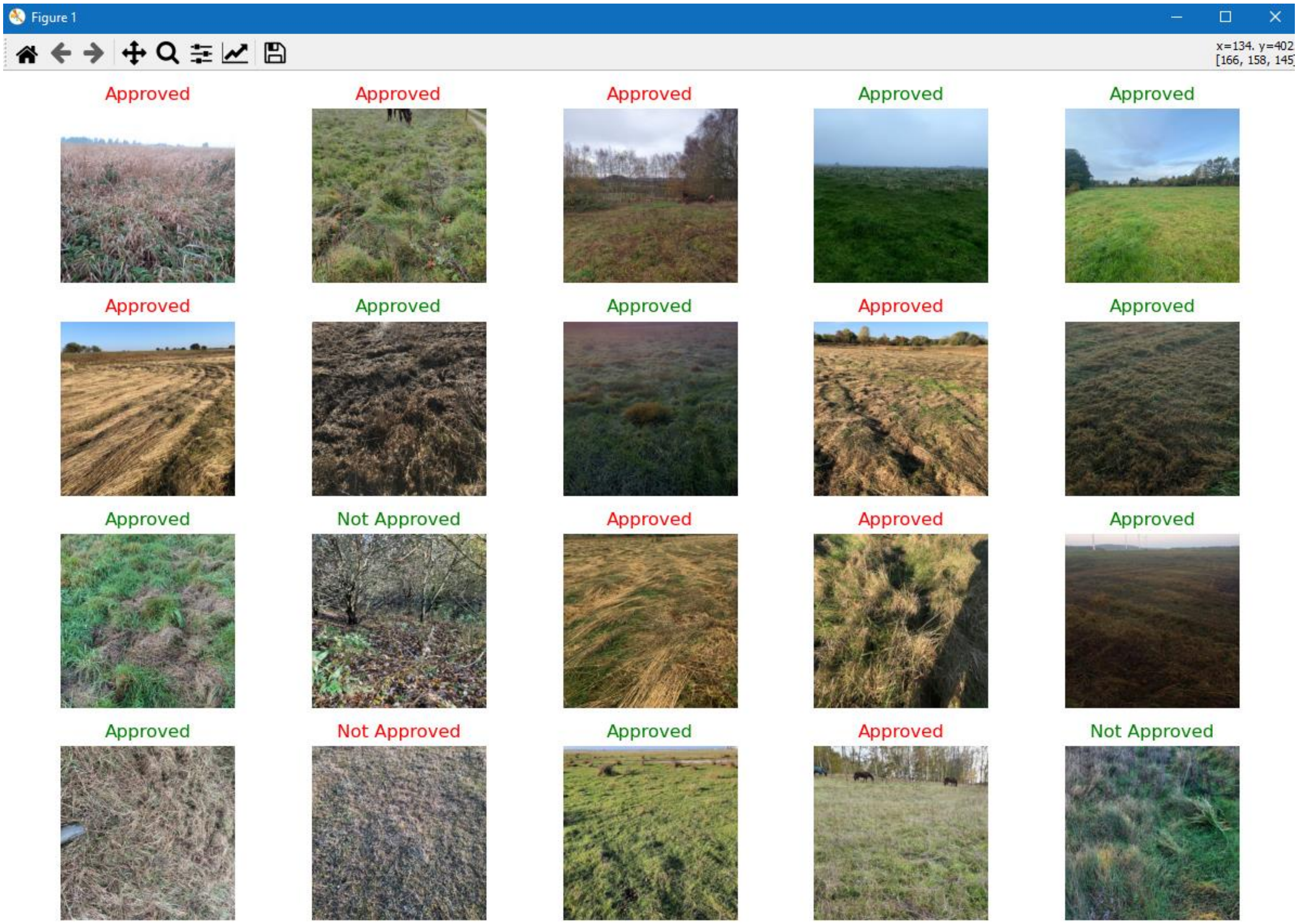


Figure 1
x=134. y=402
[166, 158, 145
Approved
Approved
Approved
Approved
Approved
Approved
Approved
Approved
Approved
Approved
Approved
Not Approved
Approved
Approved
Approved
Approved
Not Approved
Approved
Approved
Not Approved

Conclusions and lessons learned

The AI model was developed and trained on 5000 geo-tagged field images from the Danish Challenge, the classification of "Approved" and "Not Approved" grazing activity was achieved. The results show that the AI model (using a ResNet 50 CNN (Convolutional Neural Network) architecture) trained well on the available Danish Approved and Irish Clean Permanent Pasture images with a training accuracy of 94.8%, weighted F1 0.946, and a validation accuracy of 95.4%, weighted F1 0.952.

However, this model's performance drops significantly when used to predict the Danish Not Approved images (accuracy 67.7%, weighted F1 0.590). These scores are still relatively high because the dataset used was a mix of the Approved and Not Approved images, and the Approved images were also used in the model training. So, the model probably overfits and may not be learning the features we would want it to learn.

The expectation is that these results can be improved substantially by a more tailored training dataset. E.g., the model must be told what aspects in the image make sense to learn about, so that it will be able to make the expected distinction between Approved and Not Approved situations. Currently it looks at a lot of noisy / "landscapy" photos. Pre-processing images were already zoomed in on a 500x500 pixel center-bottom crop of the image, but that does not help enough.

Another AI related negative aspect is that there is possibly an information leak between training and validation since the Approved images can contain multiple photos of the same fields and some of those might end up in the validation dataset (when randomly split). This might (partly) explain why the validation metrics are slightly higher than the training metrics.

Looking at the high initial training accuracy and the overall poor performance on other field images, it is apparent the model is overfitting and does not do well on unseen field images. More clear and higher quality field images are needed of both grazed and ungrazed fields. Currently the visual characteristics for "Approved" grazed fields and "Not Approved" ungrazed fields are very noisy.