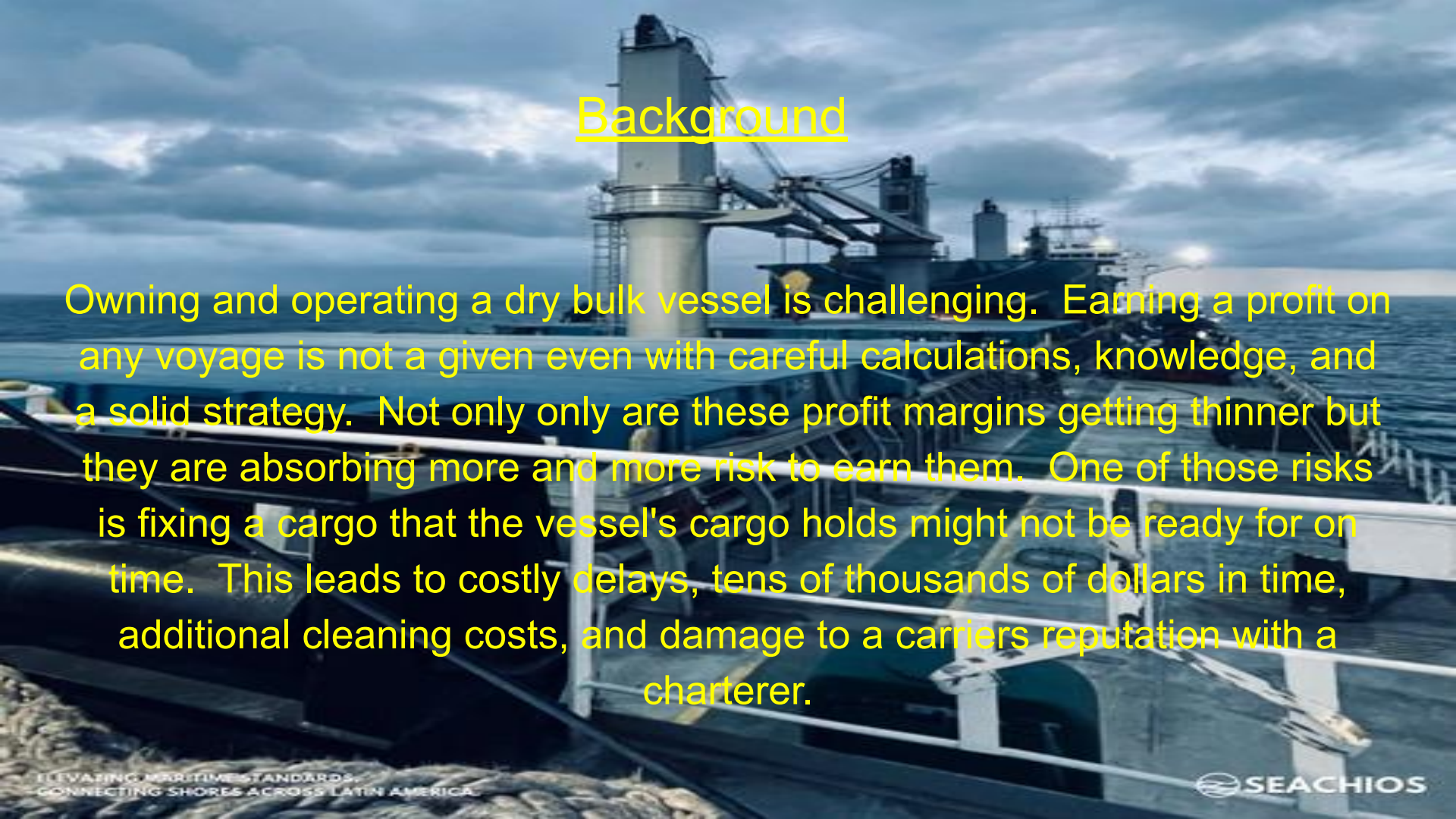
Clean or Dirty?

An Image Classification Project

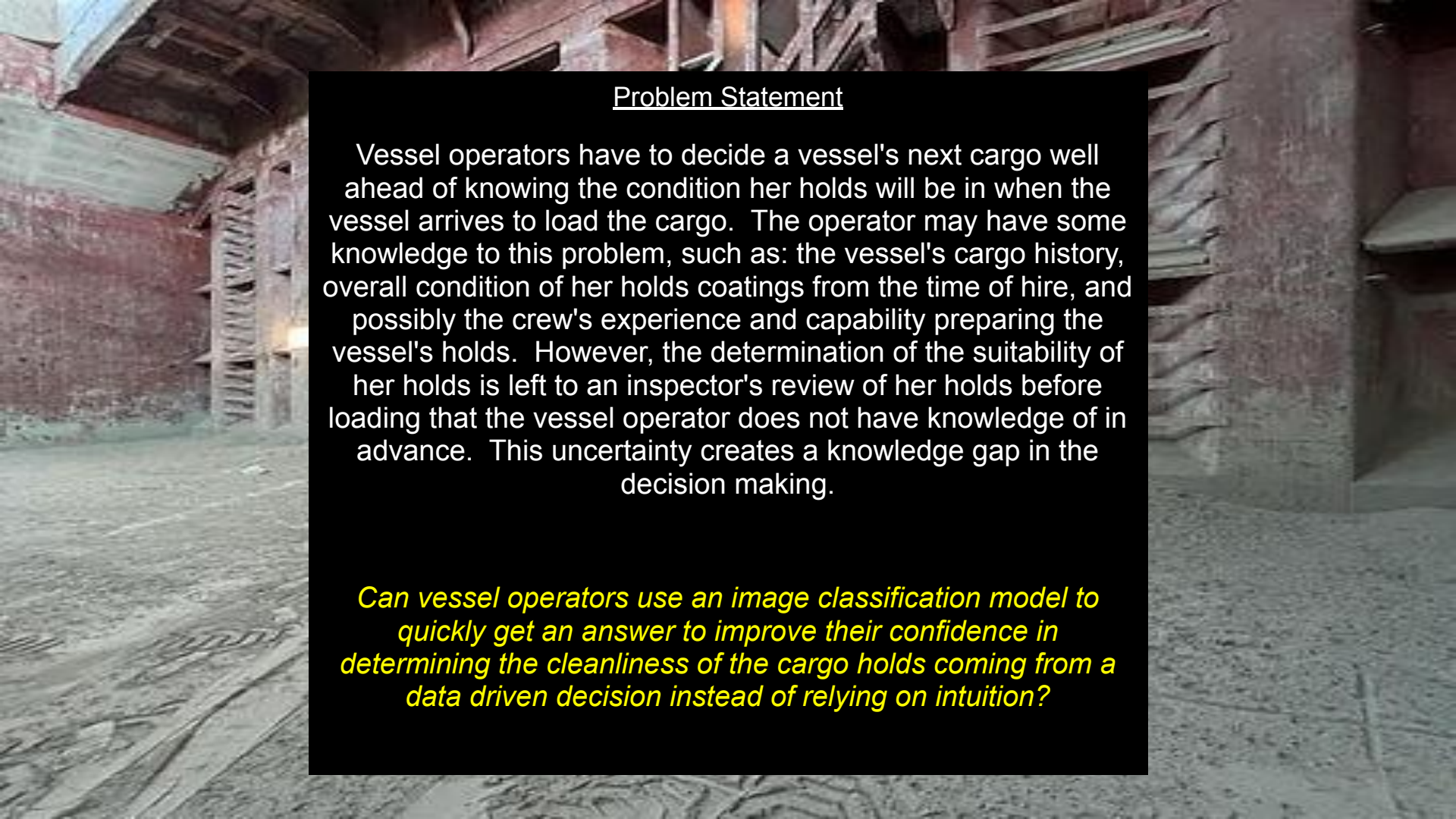
Acknowledgements

- Three D’s Marine Inc
- Seachios Marine Services
- Nippon Paint Marine



Background

Owning and operating a dry bulk vessel is challenging. Earning a profit on any voyage is not a given even with careful calculations, knowledge, and a solid strategy. Not only only are these profit margins getting thinner but they are absorbing more and more risk to earn them. One of those risks is fixing a cargo that the vessel's cargo holds might not be ready for on time. This leads to costly delays, tens of thousands of dollars in time, additional cleaning costs, and damage to a carriers reputation with a charterer.



Problem Statement

Vessel operators have to decide a vessel's next cargo well ahead of knowing the condition her holds will be in when the vessel arrives to load the cargo. The operator may have some knowledge to this problem, such as: the vessel's cargo history, overall condition of her holds coatings from the time of hire, and possibly the crew's experience and capability preparing the vessel's holds. However, the determination of the suitability of her holds is left to an inspector's review of her holds before loading that the vessel operator does not have knowledge of in advance. This uncertainty creates a knowledge gap in the decision making.

Can vessel operators use an image classification model to quickly get an answer to improve their confidence in determining the cleanliness of the cargo holds coming from a data driven decision instead of relying on intuition?

!!! SOLUTION !!!

YES! This model shows the potential to differentiate between clean and dirty cargo holds. The CNN model trained on hundreds of images of clean and dirty holds. The model shows promise to perform this task well and possibly segment different levels of cleanliness. This will be achieved by retraining the model on pictures of vessels that have passed inspections of interest. With this tool the vessel operator can quickly determine the likelihood, not a guarantee, that the vessel's holds will be accepted for the intended cargo.

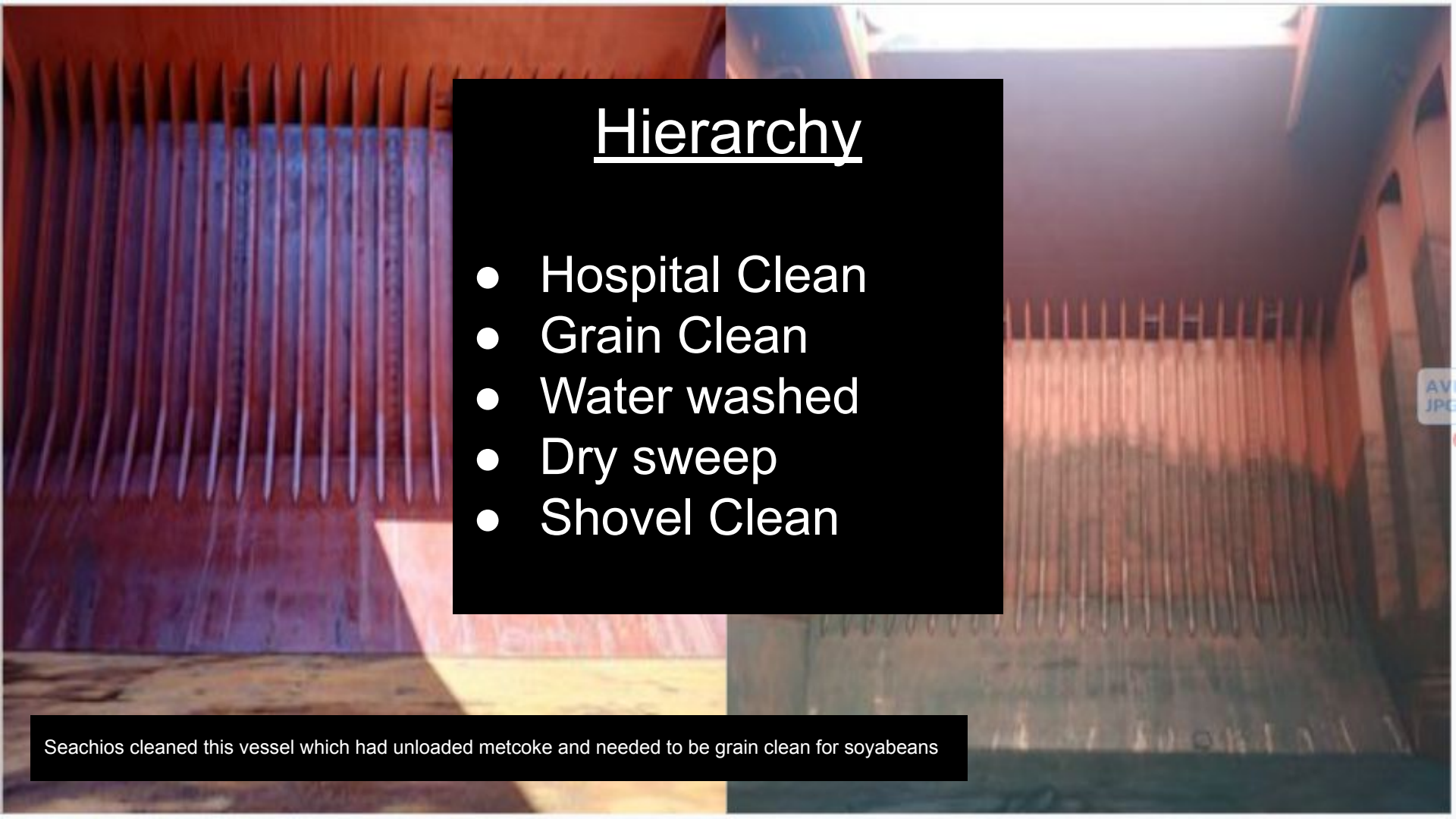
Clean or Dirty?


Seachios® Marine Services
03 November 2024 2:42 pm

Hospital Clean

$100,000


Clean or Dirty?
Both!

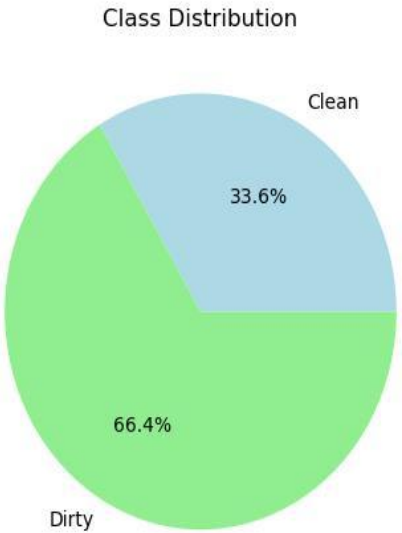


Hierarchy

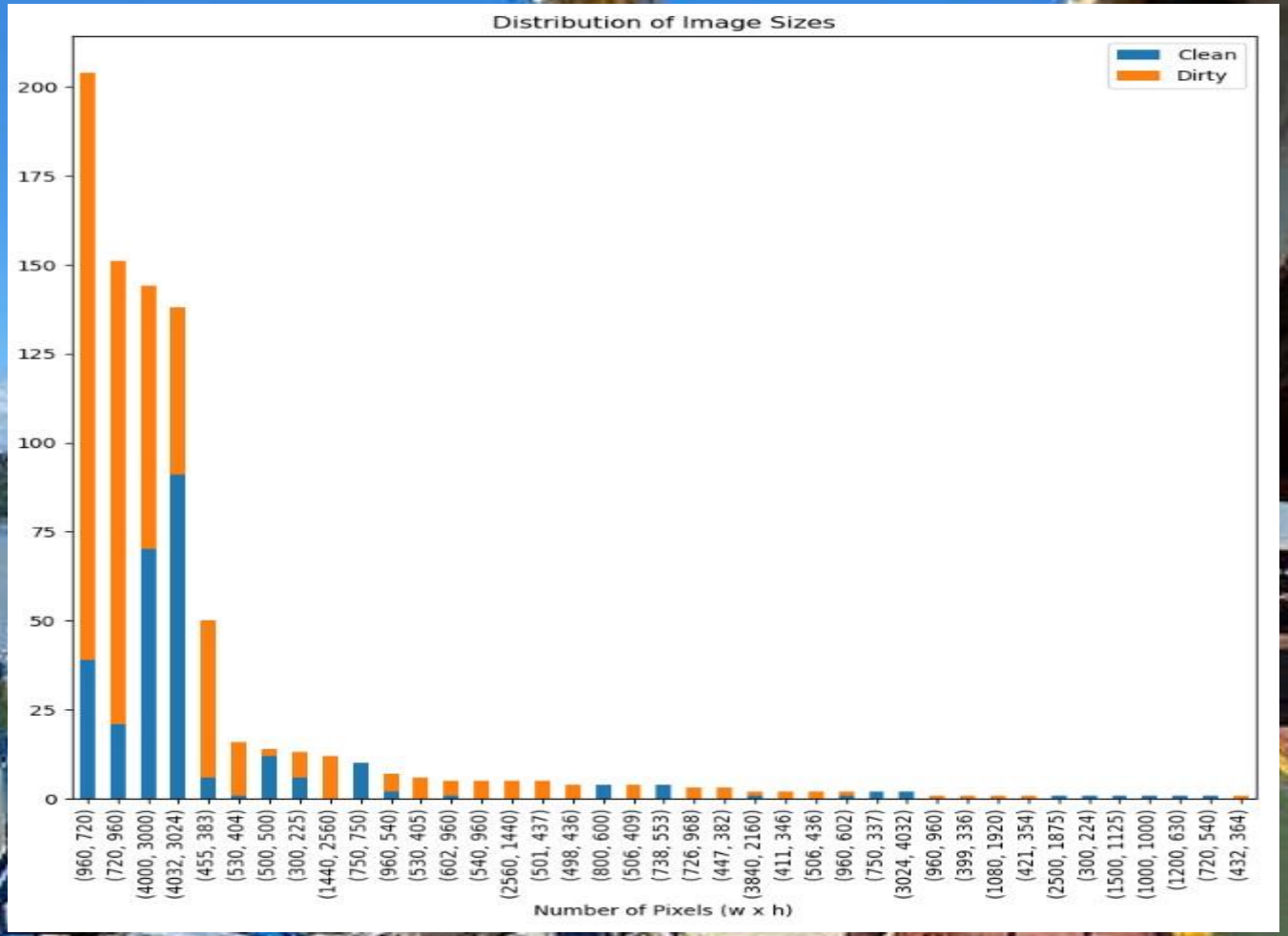
- Hospital Clean
- Grain Clean
- Water washed
- Dry sweep
- Shovel Clean

Seachios cleaned this vessel which had unloaded metcoke and needed to be grain clean for soyabeans

Image Data Size

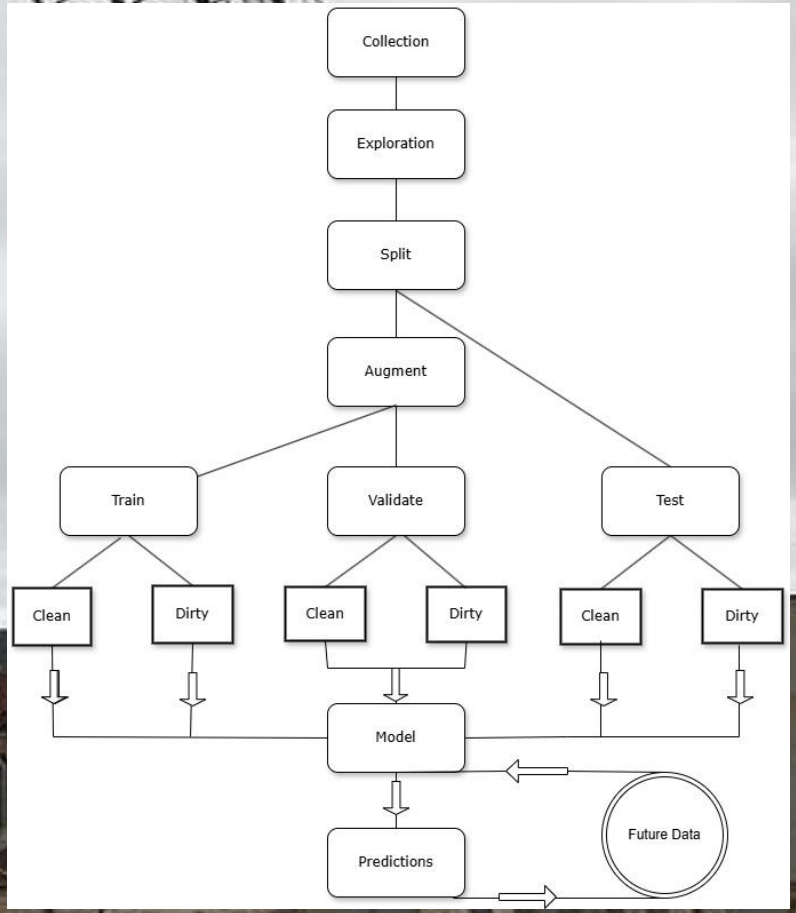


Total:	834
Clean:	279
Dirty:	551
Test:	110
Train:	1040
Validation:	110


Distribution of Image Sizes
Clean
Dirty
200
175
150
125
100
75
50
25
0
(960, 720)
(720, 960)
(4000, 3000)
(4032, 3024)
(455, 383)
(530, 404)
(500, 500)
(300, 225)
(1440, 2560)
(750, 750)
(960, 540)
(530, 405)
(602, 960)
(540, 960)
(2560, 1440)
(501, 437)
(498, 436)
(800, 600)
(506, 409)
(738, 553)
(726, 968)
(447, 382)
(3840, 2160)
(411, 346)
(506, 436)
(960, 602)
(750, 337)
(3024, 4032)
(960, 960)
(399, 336)
(1080, 1920)
(421, 354)
(2500, 1875)
(300, 224)
(1500, 1125)
(1000, 1000)
(1200, 630)
(720, 540)
(432, 364)
Number of Pixels (w x h)

After collecting, exploring, classifying, and splitting the data its time to construct a model to generate predictions.

For this project, a CNN model was developed and trained.



CNN Model Architecture

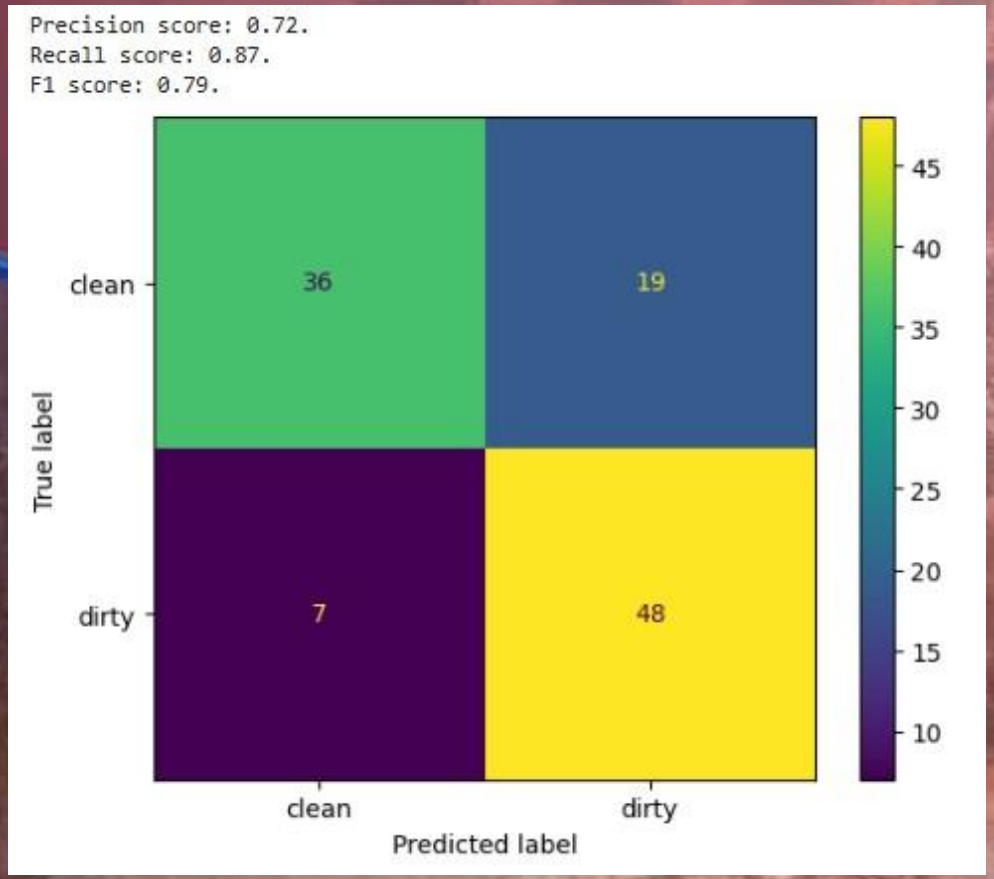
Model: "sequential"

Layer (type)	Output Shape	Param #
conv2d (Conv2D)	(None, 222, 222, 32)	896
max_pooling2d (MaxPooling2D)	(None, 111, 111, 32)	0
batch_normalization (BatchNormalization)	(None, 111, 111, 32)	128
dropout (Dropout)	(None, 111, 111, 32)	0
flatten (Flatten)	(None, 394272)	0
dense (Dense)	(None, 64)	25233472
batch_normalization_1 (BatchNormalization)	(None, 64)	256
dropout_1 (Dropout)	(None, 64)	0
dense_1 (Dense)	(None, 1)	65

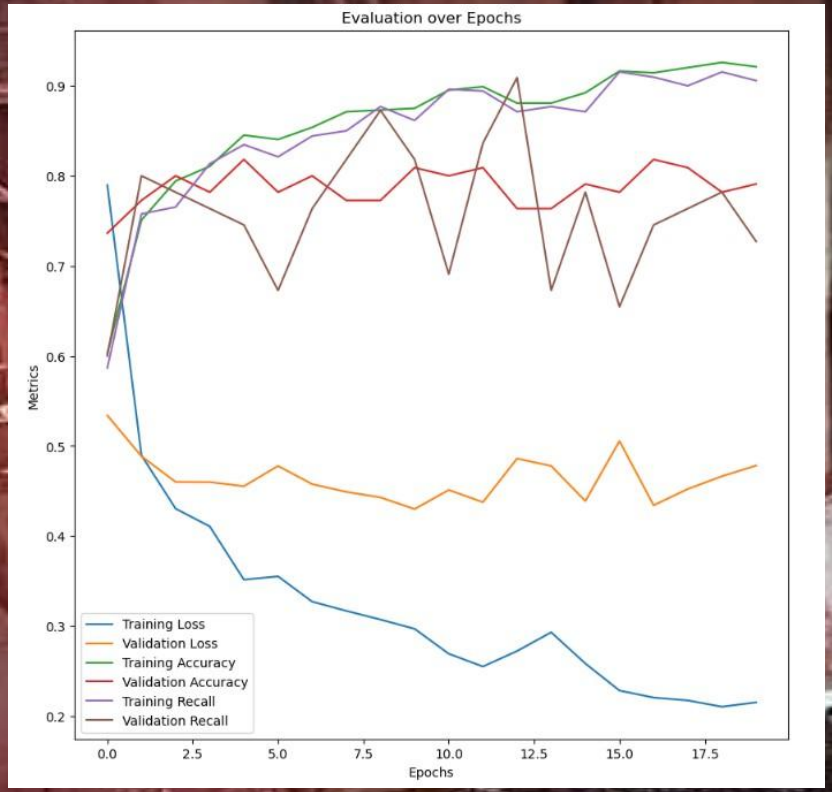
Total params: 25,234,817
Trainable params: 25,234,625
Non-trainable params: 192


Evaluation over Epochs
Training Loss
Validation Loss
Training Accuracy
Validation Accuracy
Training Recall
Validation Recall
Metrics
Epochs
1.2
1.0
0.8
0.6
0.4
0.2
0
5
10
15
20
25
Training Metrics

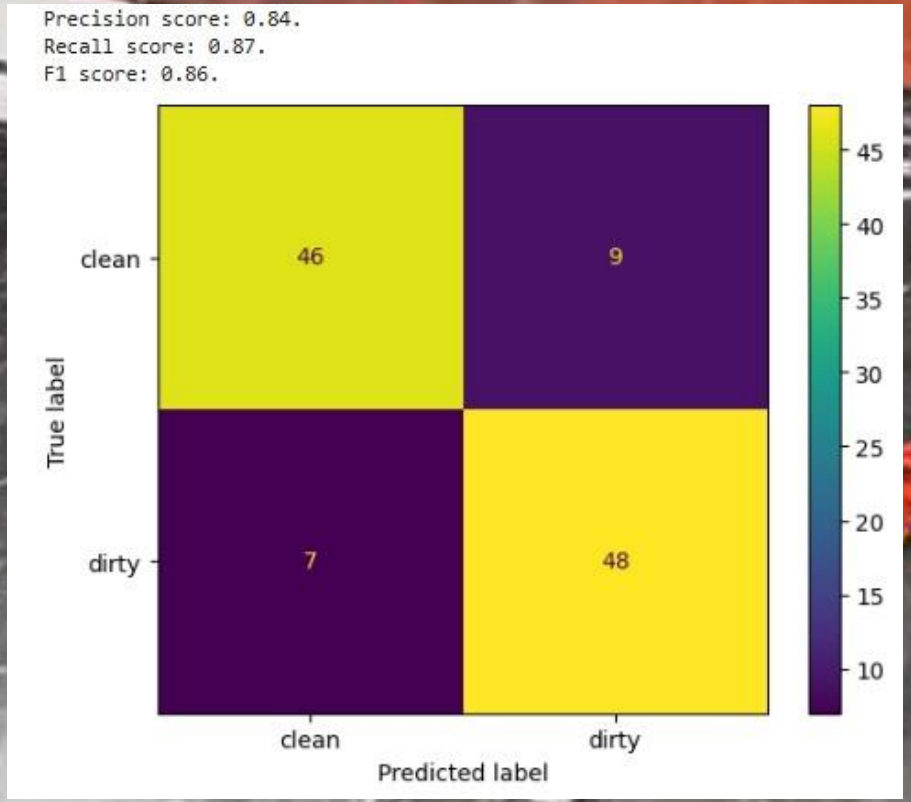
CNN Confusion Matrix



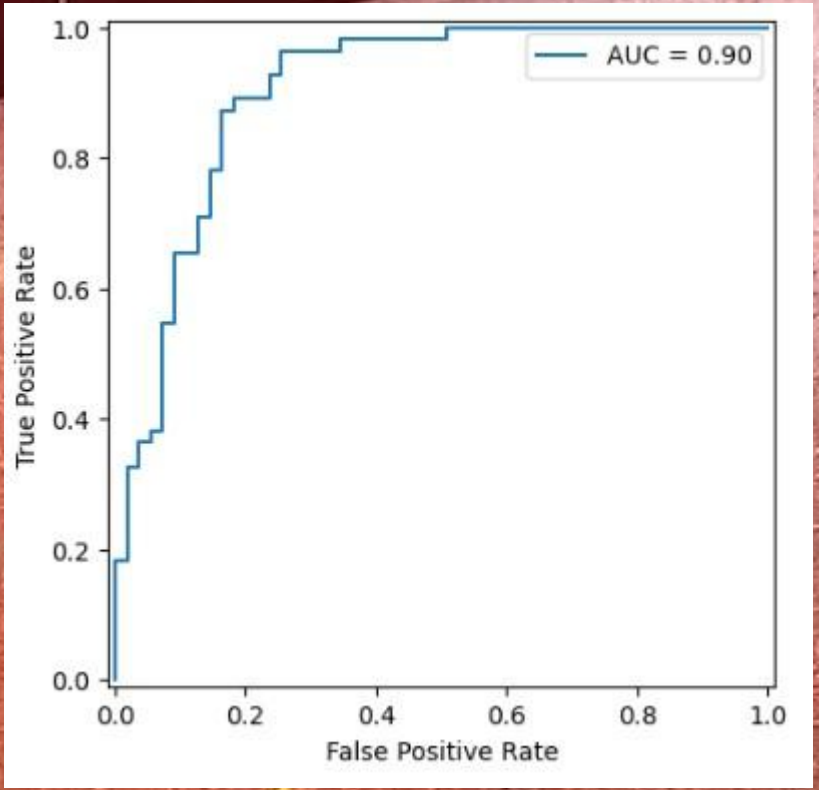
VGG16 Training Metrics



VGG 16 Confusion Matrix



VGG16 ROC AUC



Conclusion: ONWARD!

There is evidence that the dataset is small, but not too small. A complex and deep CNN model was overfit to the data while a simple CNN model was not showing signs of overfitting. Further this model was obtaining matching results on the test set as the training and validation sets.


Thank you very much!
QUESTIONS