# Model Evaluation – The Effect of Noisy Data

## Explainable Machine Learning - Deep Learning Life Cycle


Jonas Amling  Baptiste Patrice Francis Bony  Benedikt Markus Marsiske

January 19, 2023

University of Bamberg

# Table of contents

# Research Question

# Research Question and Introduction

Goals for evaluating a trained model:

- How well does the model perform?
- Is there a difference between the classes?
- How robust is our model?

# Research Question and Introduction

Goals for evaluating a trained model:

- How well does the model perform?
- Is there a difference between the classes?
- How robust is our model?

Specific Research Question: **How does the model perform on distorted data? Does the usage of distorted test data lead to a worse model performance compared to the same test data without distortion?**

# Basic Model Evaluation

# Model Performance - Training Data

How well did our model perform on our training data during the training?

```
Epoch 10 loss: 0.4450182304324874
Accuracy of the network on the test set: 83.34380892520427%
Accuracy of rock : 84.93408662900188%
Accuracy of paper : 82.02676864244742%
Accuracy of scissors : 83.05400372439479%
```

**Figure 1:** accuracy during training

# Model Performance - Test Data

How well did our model perform on our test data?

- same kind of data as our training data

# Model Performance - Test Data

How well did our model perform on our test data?

- same kind of data as our training data
- model performance:

```
print('testing against the test dataset:')
test_accuracy(testloader)
```

✓ 2.2s

```
testing against the test dataset:
Accuracy of the network on the test set: 86.09077598828696%
Accuracy of rock : 87.71929824561404%
Accuracy of paper : 85.26785714285714%
Accuracy of scissors : 85.28138528138528%
```

**Figure 2:** accuracy on test dataset

# Confusion Matrix - Testset

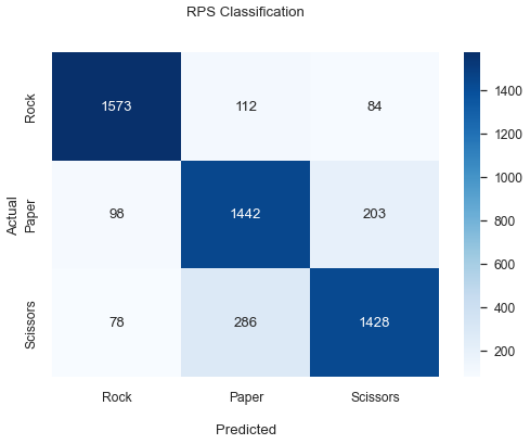


**Figure 3:** Numeric CM of testset

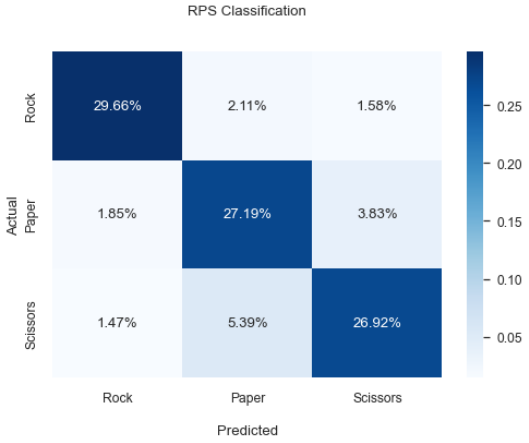


**Figure 4:** CM of testset (in %)

# Model Performance - Unseen Dataset

How well did our model perform on the provided testset?

- different kind of images:
  - only custom made data
  - no images of big datasets (big part of training data)

# Model Performance - Unseen Dataset

How well did our model perform on the provided testset?

- different kind of images:
  - only custom made data
  - no images of big datasets (big part of training data)
- model performance:

```
testing against the unknown dataset:
Accuracy of the network on the test set: 67.79661016949153%
Accuracy of rock : 94.73684210526316%
Accuracy of paper : 60.0%
Accuracy of scissors : 50.0%
```

**Figure 5:** accuracy on unseen test dataset

# Confusion Matrix - Unseen Testset

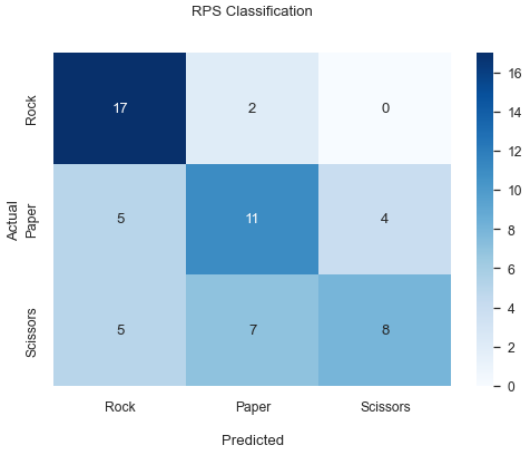


**Figure 6:** Numeric CM of unseen testset

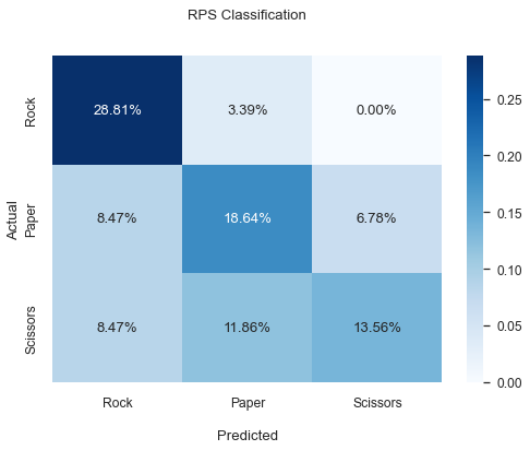


**Figure 7:** CM of unseen testset (in %)

# Evaluating On Distorted Images

# Image Distortion - Random Distortion

Each pixel has a chance to be removed (25%):

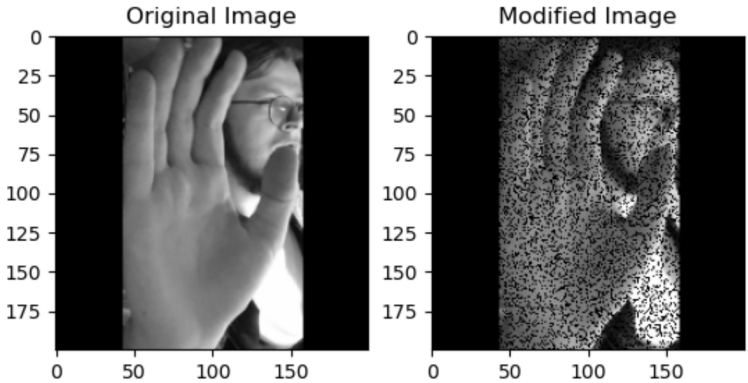


**Figure 8:** random distortion with a pixel elimination probability of 0.25

# Image Distortion - Gaussian Distortion

Gaussian Filter:

- follows normal distribution
- parameter: standard deviation (25 in our case)

# Image Distortion - Gaussian Distortion

Gaussian Filter:

- follows normal distribution
- parameter: standard deviation (25 in our case)

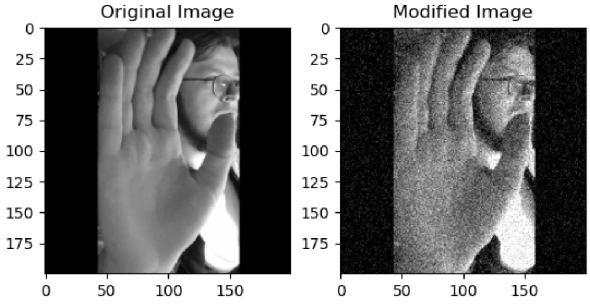


**Figure 9:** distortion using a Gaussian filter with SD = 25

# Model Performance on distorted data

Random Distortion:

```
testing against the randomly distorted dataset:
Accuracy of the network on the test set: 42.3728813559322%
Accuracy of rock : 5.2631578947368425%
Accuracy of paper : 60.0%
Accuracy of scissors : 60.0%
```

**Figure 10:** accuracy on randomly distorted testset

Gaussian Distortion:

```
testing against the dataset with gaussian distortion:
Accuracy of the network on the test set: 55.932203389830505%
Accuracy of rock : 31.57894736842105%
Accuracy of paper : 65.0%
Accuracy of scissors : 70.0%
```

**Figure 11:** accuracy on testset distorted with a Gaussian filter

# Gaussian Distortion - Confusion Matrix

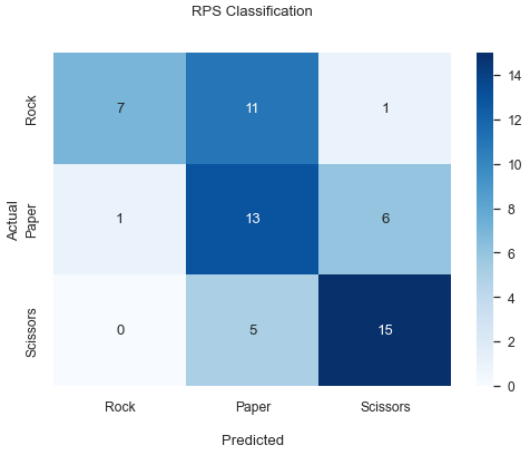


**Figure 12:** Numeric CM of distorted testset

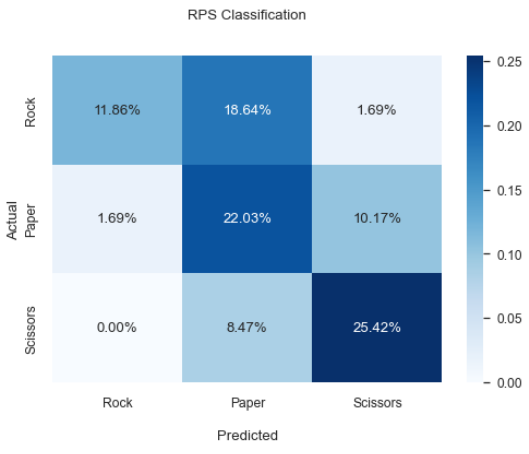


**Figure 13:** CM of distorted testset (in %)

# Future Considerations

# Future Considerations

Things we will have to consider to finalize our project:

- improve model performance:
  - train more epochs
  - use deeper net (VGG-16 inspired)
  - adjust position of dropout layers
  - train with noisy data to increase robustness
  - fine tuning

# Thank you!