Model Evaluation – The Effect of Noisy Data

Explainable Machine Learning - Deep Learning Life Cycle

Jonas Amling Baptiste Patrice Francis Bony Benedikt Markus Marsiske

February 1, 2023

University of Bamberg

Table of contents

Research Question

Research Question and Introduction

Focus for evaluating our trained model:

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

Research Question and Introduction

Focus for evaluating our 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?

```
testing against the training dataset
Accuracy of the network on the test set: 98.91792029559251%
Accuracy of rock : 99.0506329113924%
Accuracy of paper : 98.99598393574297%
Accuracy of scissors : 98.7109375%
```

Figure 1: Accuracy at the end of training

Confusion Matrix - Training Data

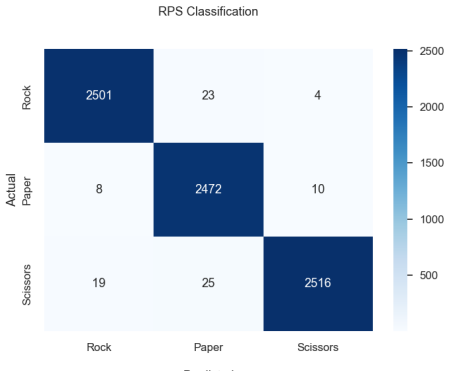

Figure 2: Numeric CM of Training Data

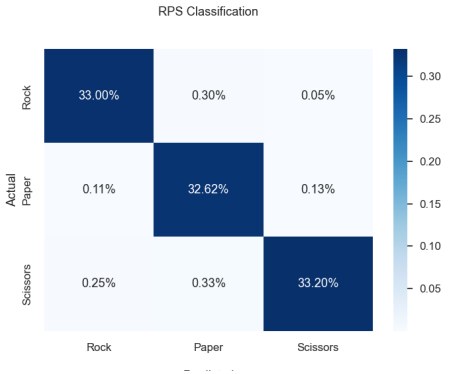

Figure 3: CM of Training Data (in %)

Model Performance - Validation Data

How well did our model perform on the provided validation data?

- custom made data
- no images of big datasets (significant portion of training data)
- incorporated in our training

Model Performance - Validation Data

How well did our model perform on the provided validation data?

- custom made data
- no images of big datasets (significant portion of training data)
- incorporated in our training
- model performance:

```
testing against the validation dataset
Accuracy of the network on the test set: 86.44067796610169%
Accuracy of rock : 84.21052631578948%
Accuracy of paper : 85.0%
Accuracy of scissors : 90.0%
```

Figure 4: Accuracy on validation dataset

Confusion Matrix - Validation Set

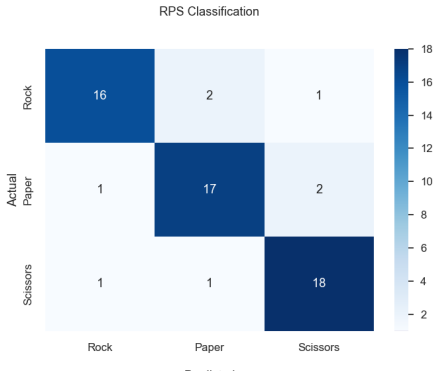

Figure 5: Numeric CM of testset

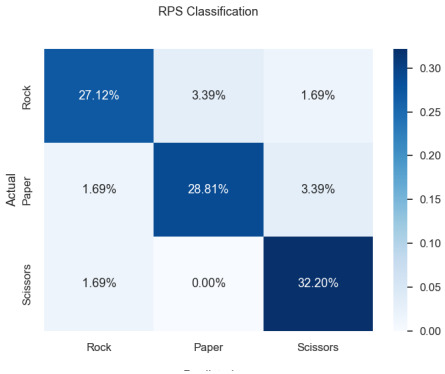

Figure 6: CM of testset (in %)

Model Performance - Test Data

How well did our model perform on the provided test data?

- more custom made data
- no images of big datasets (significant portion of training data)
- unseen

Model Performance - Test Data

How well did our model perform on the provided test data?

- more custom made data
- no images of big datasets (significant portion of training data)
- unseen
- model performance:

```
testing against the testing dataset
Accuracy of the network on the test set: 83.54430379746836%
Accuracy of rock : 80.76923076923077%
Accuracy of paper : 98.14814814814815%
Accuracy of scissors : 71.15384615384616%
```

Figure 7: accuracy on test dataset

Confusion Matrix - Unseen Testset

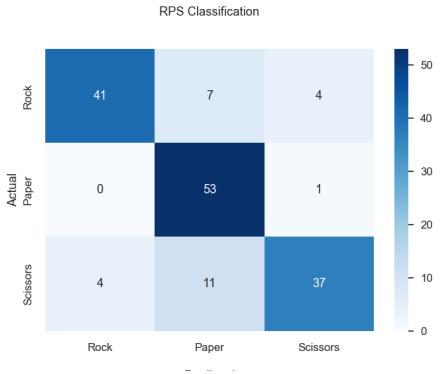

Figure 8: Numeric CM of unseen testset

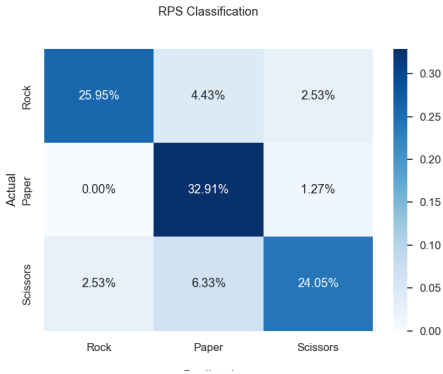

Figure 9: CM of unseen testset (in %)

Evaluating On Distorted Images

Image Distortion - Random Distortion

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

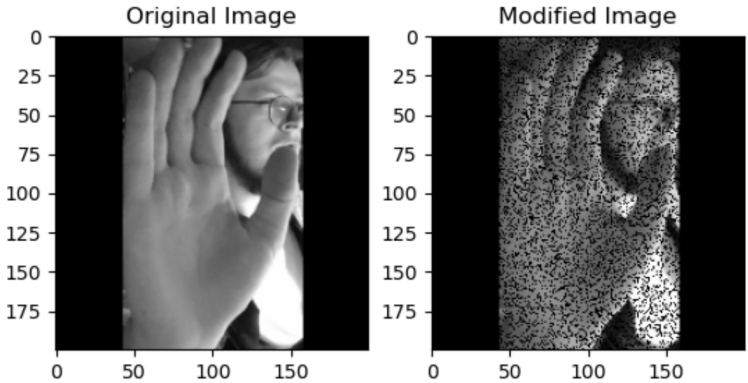

Figure 10: random distortion with a pixel elimination probability of 25%

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)

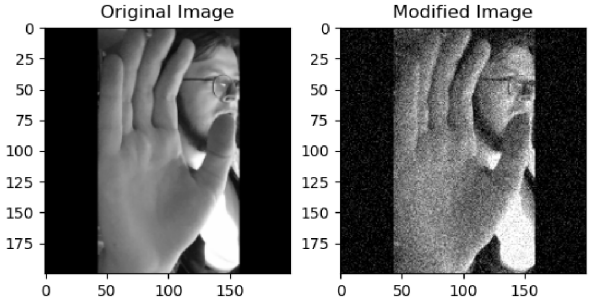

Figure 11: distortion using a Gaussian filter with SD = 25

Model Performance on distorted data

Random Distortion:

```
testing final_model_D_True_B_False against the testing dataset with random noise:
Accuracy of the network on the test set: 44.30379746835443%
Accuracy of rock : 67.3076923076923%
Accuracy of paper : 53.7037037037037%
Accuracy of scissors : 11.538461538461538%
```

Figure 12: Accuracy on randomly distorted testset

Gaussian Distortion:

```
testing final_model_D_True_B_False against the testing dataset with gaussian noise:
Accuracy of the network on the test set: 47.46835443037975%
Accuracy of rock : 28.846153846153847%
Accuracy of paper : 87.03703703703704%
Accuracy of scissors : 25.0%
```

Figure 13: Accuracy on testset distorted with a Gaussian filter

Random Distortion - Confusion Matrix

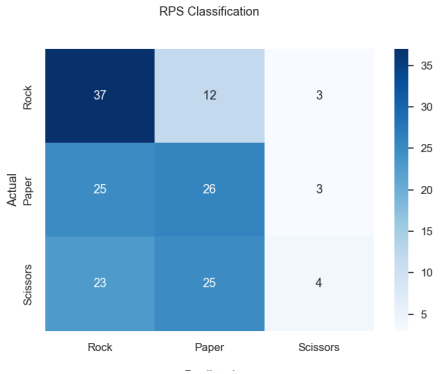

Figure 14: Numeric CM of distorted testset

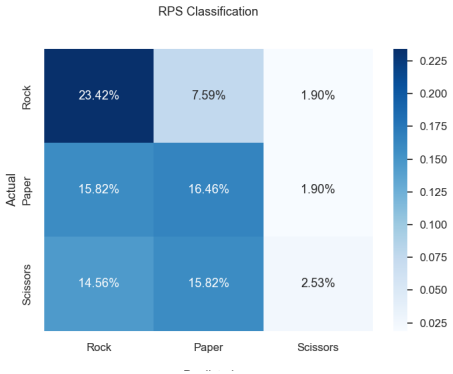

Figure 15: CM of distorted testset (in %)

Gaussian Distortion - Confusion Matrix

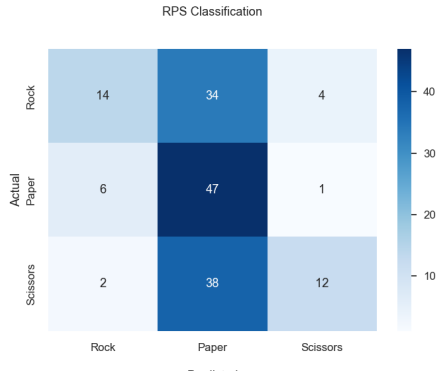

Figure 16: Numeric CM of distorted testset

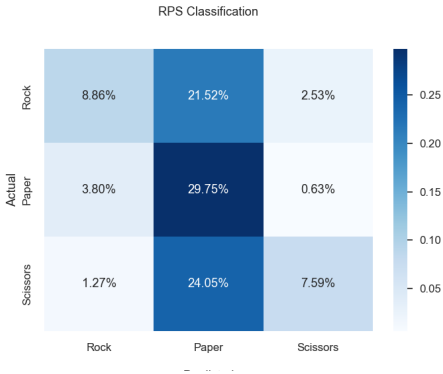

Figure 17: CM of distorted testset (in %)

Performances on Different Test Data

	Old	Current	No Regularization	With Noisy Data
Undistorted Testset	67.8%	83.5%	80.4%	61.4%
Random Distortion	42.4%	44.3%	38.0%	64.4%
Gaussian Distortion	55.9%	47.5%	58.2%	63.3%

Table 1: Performance comparison of various models on different test data

Conclusion

Conclusion

Things we learned evaluating our model:

- better performing model $\neq$ more robust model
- big impact of noisy data
- robustness can be improved by training with noisy data

Conclusion

Things we learned evaluating our model:

- better performing model $\neq$ more robust model
- big impact of noisy data
- robustness can be improved by training with noisy data

Answering our research question:

The usage of distorted test data does lead to a worse model performance compared to the same test data without distortion.

Thank you!