Classifying recaptured identity documents using the biomedical Meijering and Sato algorithms

John Magee| john.magee@tudublin.ie
Stephen Sheridan| stephen.sheridan@tudublin.ie
Christina Thorpe| christina.thorpe@tudublin.ie

School of Informatics and Cyber Security
Faculty of Computing, Digital and Data
Technological University Dublin
Dublin, Ireland

1. INTRODUCTION

Recaptured identity documents are a low-cost, high-risk threat to modern eKYC systems. Bad actors can easily manipulate images and print them. Existing solutions typically demand manual review of remotely captured identity documents, this is expensive and does not scale. In 2022, the UK National Crime Agency estimated fraud cost business hundreds of billion pounds per year and document forgery is an area of investigation by Europol.

2. OBJECTIVE

Develop techniques that are:

- Successfully detect 95% of recaptured documents.
- Techniques that are secure.
- Easy to understand and explain (GDPR).
- Document agnostic.
- Cost effective implementation.

3. METHOD

- Empirical research to compare different types of recaptured documents.
- Meijering[2] and Sato[3] biomedical imaging filters applied to recaptured identity documents.
- Machine learning algorithms, Support Vector Machine and Decision Tree, are trained to classify recaptured images utilising their histograms has input features.

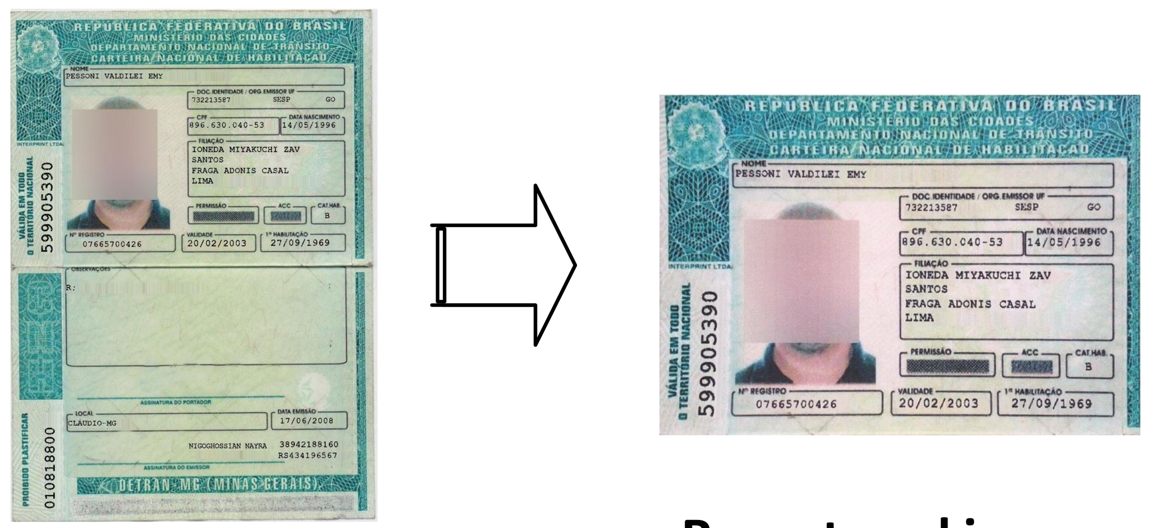
4. DATA

The Brazilian Identity Document dataset published by Soares et al [1] is the data set used in this research.

5. Data Capture And Processing

These documents were reproduced using

- 2 printers
- 2 mobile phones
- 1 BID image -> 10 recaptured images
- 1040 recaptured images



Recaptured image

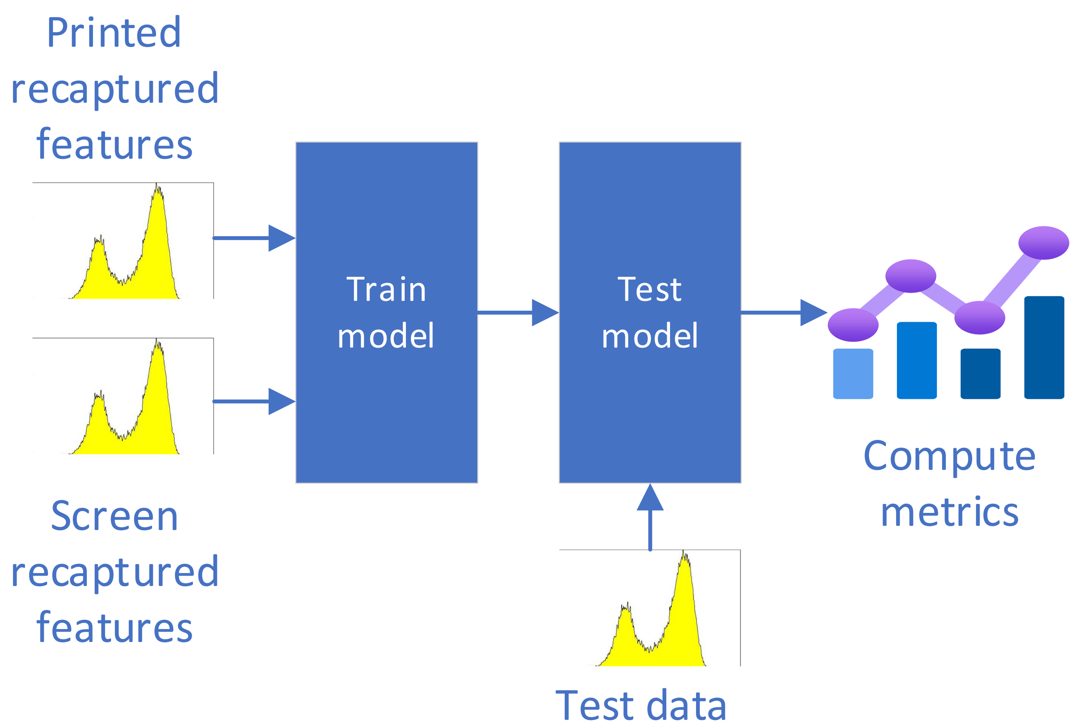
Source image

Simple transformation applied to ensure the process is easy to understand:

- Load identity document image
- Apply Meijering filter
- Apply Sato filter
- Generate histograms and record intensity value
- Repeat for different number of histogram bins

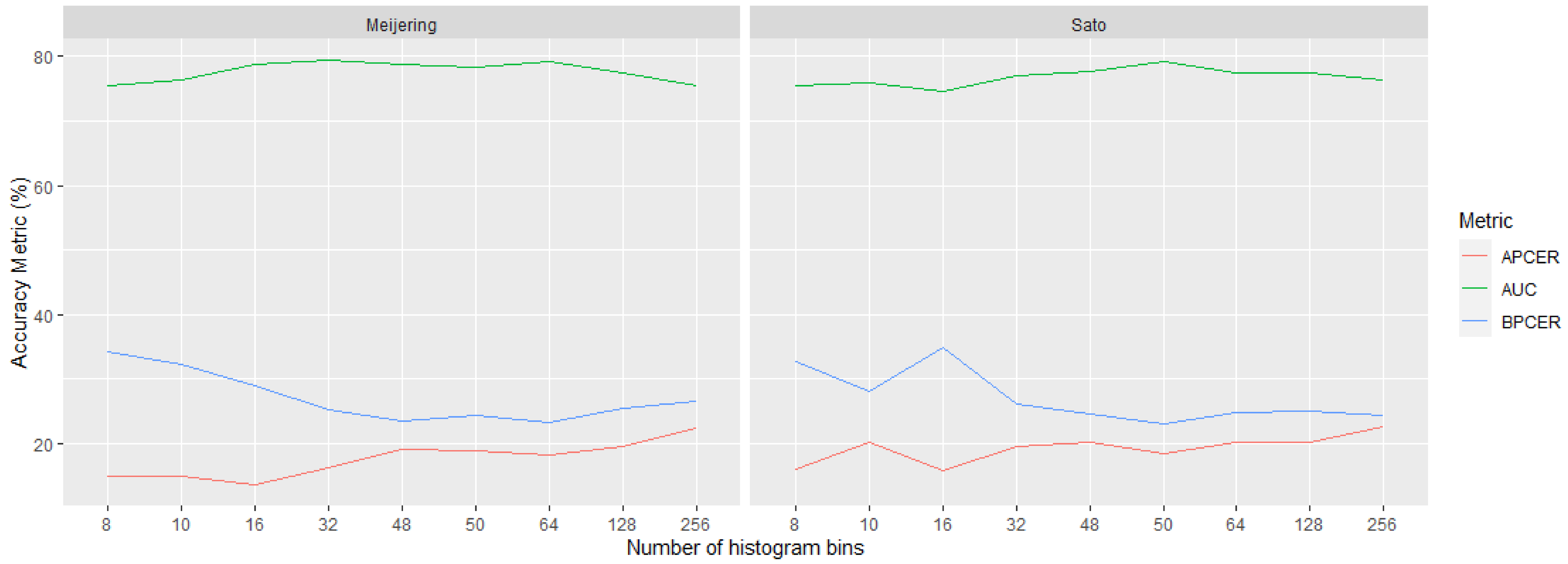
6. Model Generation

Support Vector Machine and Decision Tree models are trained and tested using histogram intensity values.



7. RESULTS

The chart shows the APCER, BPCER and AUC results of the Support Vector Machine models when trained using histogram intensity values from Meijering and Sato filtered images.



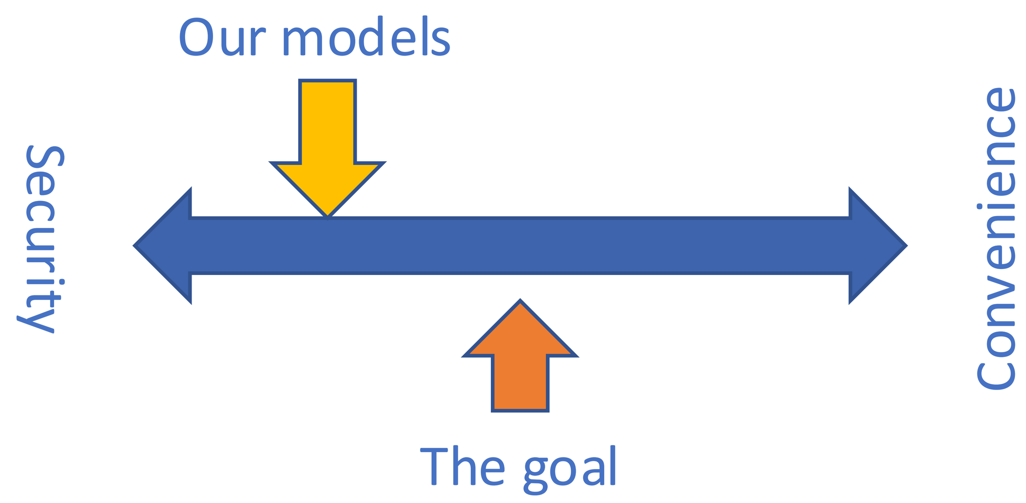
The chart shows a snapshot the results obtained using the Decision Tree algorithm trained using the histogram intensity values from the Meijering and Sato filtered images. Different histogram bin numbers and split criteria are used.

Metric	Filter	Split	Bins								
			8	10	16	32	48	50	64	128	256
iPhone8											
APCER	**Meijering**	Gini	14.55	15.66	15.59	15.60	15.33	14.31	14.15	15.91	14.83
		Entropy	14.20	14.81	14.66	14.82	15.13	13.98	14.97	14.82	15.31
		log loss	14.05	14.88	14.78	14.67	14.98	13.85	14.98	15.00	15.28
	Sato	Gini	20.70	19.16	19.42	20.88	21.86	20.62	22.20	23.89	21.20
		Entropy	19.33	19.82	19.17	20.25	20.80	20.51	20.35	23.22	21.06
		log loss	19.36	19.55	18.86	20.09	20.87	20.60	20.62	23.01	21.03

- Models achieve low APCER but high BPCER.
- Meijering filter typically outperformed the Sato filter.
- Decision Tree results were invariant to the number of histogram bins used
- Entropy splitting provided the most accurate Decision Tree results

9. CONCLUSIONS

Our models are too strict, rejecting too many legitimate documents. However, we have demonstrated that the biomedical filters can be used to help identify recaptured identity documents and more research into this area is warranted.



REFERENCES

[1] Soares, D. S., Das Neves Junior, R. B., Bezerra, B. L. D., BID dataset: a challenge dataset for document processing tasks, in: Anais Estendidos da Conference on Graphics, Patterns and Images.

[2] Meijering, E., Jacob, M. , Sarria, J.-C. , Steiner, P. , Hirling, H. , Unser, M., Design and validation of a tool for neurite tracing and analysis in fluorescence microscopy images, Cytometry Part A 58A (2004) 167–176.

[3] Sato, Y., Nakajima, S., Shiraga, N., Atsumi, H., Yoshida, S., Koller, T., Gerig, G., Kikinis, R., Three dimensional multi-scale line filter for segmentation and visualization of curvilinear structures in medical images 2 (1998) 143–168.