BANCA D'ITALIA
EUROSISTEMA

Mercati, infrastrutture, sistemi di pagamento

(Markets, Infrastructures, Payment Systems)

Siamese Networks for AI-powered Automated Banknote Quality Control

by Salvatore Gentile, Andrea Luciani, Sabina Marchetti, Domenico Pansini and Marco Viticoli

June 2026

Number 82

BANCA D'ITALIA
EUROSISTEMA

Mercati, infrastrutture, sistemi di pagamento

(Markets, Infrastructures, Payment Systems)

Siamese Networks for AI-powered Automated Banknote Quality Control

by Salvatore Gentile, Andrea Luciani, Sabina Marchetti, Domenico Pansini and Marco Viticoli

Number 82 – June 2026

The papers published in the 'Markets, Infrastructures, Payment Systems' series provide information and analysis on aspects regarding the institutional duties of the Bank of Italy in relation to the monitoring of financial markets and payment systems and the development and management of the corresponding infrastructures in order to foster a better understanding of these issues and stimulate discussion among institutions, economic actors and citizens.

The views expressed in the papers are those of the authors and do not necessarily reflect those of Banca d'Italia.

The series is available online at www.bancaditalia.it.

Printed copies can be requested from the Paolo Baffi Library: richieste.pubblicazioni@bancaditalia.it.

ISSN 2724-6418 (online)
ISSN 2724-640X (print)

Banca d'Italia
Via Nazionale, 91 - 00184 Rome - Italy
+39 06 47921

Designed and printing by the Printing and Publishing Division of Banca d'Italia

SIAMESE NETWORKS FOR AI-POWERED AUTOMATED BANKNOTE QUALITY CONTROL

by Salvatore Gentile[**], Andrea Luciani[*], Sabina Marchetti[*], Domenico Pansini[**] and Marco Viticoli[**]

Abstract

We apply Siamese Networks, a class of artificial intelligence (AI) neural network models, to the field of banknote production. The criticality of quality control in banknote production requires human oversight throughout the process. We rely on few-shot learning to develop an AI-powered tool that could support quality control by highly trained human experts at the end of the production pipeline, by spotting potential defects on banknote images. We show our approach succeeds in such a complex application, where the enumeration of all possible defect occurrences – in terms of type, shape, severity class and location – is unfeasible. Our proposal achieves high accuracy and proves especially reliable in ensuring that no defects occur, i.e. classifying a banknote as fit. As is the case with most neural network models, our tool is accurate but inherently difficult to read in its underlying logics. To address this potential issue while enhancing transparency, we complement our proposal with an image segmentation-based explainability approach to support the application of our tool.

JEL Classification: : C45, L15, L69, O39.

Keywords: banknote production, artificial intelligence, artificial neural networks, one-shot learning, quality control.

Sintesi

Il lavoro considera l'applicazione di modelli di Intelligenza Artificiale (IA), basati sulle Reti Siamesi, per supportare gli operatori altamente specializzati incaricati del controllo di qualità nel processo di produzione delle banconote. Attraverso l'utilizzo del c.d. *few-shot learning* si sviluppa una rete neurale utile a supportare l'operatore nella analisi delle immagini delle banconote e nell'individuazione di potenziali errori di stampa. Anche in un contesto in cui non può esistere una enumerazione esaustiva di tutti i potenziali difetti presenti su una banconota – in termini di tipo, forma, severità e posizione – il modello sviluppato risulta efficace. Viene infine proposto un approccio in grado di rendere conto delle indicazioni generate dallo strumento ("spiegabilità") e di supportare quindi l'operatore nell'identificazione di eventuali difetti e nella successiva decisione.

[*] Banca d'Italia, Directorate General for Economics, Statistics and Research.
[**] Banca d'Italia, Directorate General for Human Resources, Information and Banknote production.

Contents

1. Introduction 7

2. Data and methods 7

2.1 Related works 7

2.2 Data 8

2.3 Modeling 9

3. Results 10

4. Conclusions 15

References 17

Appendix 19

1. Introduction[1]

The production of banknotes is a highly articulated process, regulated by comprehensive quality control measures to ensure the integrity and security of the final product. While many stages of the production and quality controls have been effectively automated over time or supported by visual inspection systems, final assessment of finished banknotes is carried out through manual and visual inspection by specialized operators. It is a complex task, where operators are responsible for identifying the number and type of printing defects, which constitute a crucial discriminating parameter for determining the conformity of each production batch.

Although quality controls are carried out by highly skilled and trained personnel that relies on shared guidelines, the assessment of printing defects is very time consuming. Furthermore, it could be exposed to potential subjectivity bias, reflecting the experience and sensitivity of the operators.

This work aims to contribute to the literature on artificial intelligence (AI) applications in industrial quality control, and specifically to the niche but critical area of banknote production. It extends and improves previous research from Boria *et al.* (2023), which explores the potential of so-called Siamese networks, a class of artificial neural networks, to discern and classify printing irregularities on banknote images. First, we widen the defect class to encompass all possible defects - in terms of type, size and location and severity - from a condition entailing the reject of the entire batch, to a minimal defect that does not necessitate any intervention unless associated with other defects. Furthermore, we extend the analysis to both sides of the banknote, called *front* and *back*. We also develop a visualization strategy, to support localization of defects by human experts, given a model output.

The paper is structured as follows. Section 2 provides an overview of data; it establishes the theoretical underpinnings of Siamese networks and their suitability for the task at hand, i.e., to deliver a robust model that can withstand the rigors of a production environment. Section 3 outlines main aggregate results and provides insights on visual tools to support defect detection by analysts and validate models' results. We conclude our analysis in Section 4.

2. Data and methods

Quality assessment is crucial to reducing waste during the banknote production process. Our application leverages Siamese networks, a class of artificial neural networks, i.e., machine-learning models inspired by interconnected neurons that learn patterns from data, to support and enhance quality controls. This section provides an overview of the elements that underpin our application. It is organized into three subsections: Section 2.1 presents an essential review of related work in the literature; Section 2.2 discusses the data utilized in our study; Section 2.3 outlines our modeling approach, which employs few-shot learning with Siamese networks, explaining how this technique is leveraged to address the challenges specific to our task.

2.1 Related works

In the context of banknote images, the collection of a comprehensive dataset that represents all possible defects is often impractical or even impossible to carry out. This limitation makes it challenging for traditional learning approaches to generalize to certain defect instances, as they may not have been adequately represented in the training data.

[1] We are grateful to Valentina Michelangeli, Angelamaria Fiori, Alessio De Vincenzo, Emilia Bonaccorsi Di Patti, Giuseppe Bruno, for their valuable comments and suggestions.

One-shot learning (Fei et al., 2006) represents a viable option to overcome challenges related to representation and variability issues in banknote images. One-shot learning mimics human ability to associate unfamiliar entities based on similarity. This makes it suitable to learn a strategy to assign an input image to a given class by comparing common and differing features with each class representatives. A more general approach to one-shot learning is few-shot, or k-shot, learning (Palatucci et al., 2009; Fink, 2004), which uses multiple examples per class. Both one- and k-shot learning approaches often rely on the class of artificial neural networks called *Siamese networks*, first introduced by Koch et al. (2015). Siamese networks offer a robust and reliable solution for defect detection tasks, particularly when dealing with highly variable data.

The use of Siamese networks in industrial manufacturing applications beyond banknote production is documented in Deshpande et al. (2020) and Mohr et al. (2021). In Boria et al. (2023), we explored the use of Siamese networks to automate and improve quality control in banknote production. Therein, the analysis narrows the scope of defect detection to a very specific type confined to a particular location: missing ink spots (so-called *bite*) on the flag element located in the upper left part of the front side of the 50 Euro banknote. The defect, characterized by high severity, is also targeted by Ke et al. (2016), which rely on established convolutional neural network architectures and traditional supervised learning for classification. Further applications for processing banknote images include recognition - e.g., Jadhav et al. (2019), Sultana et al. (2024) and Nazir et al. (2024) - and counterfeit detection - e.g., Sawant et al. (2022), Parnika et al. (2023) and Venu (2023). For an updated review of the literature, see Sadyk et al. (2024).

2.2 Data

To train our neural network models, i.e., to estimate parameters on labelled examples, we use images of fit (not defective) banknotes and contrast them with sets of unfit (defective) banknote images. Our dataset is representative of fit instances as well as unfit ones, that can exhibit one or more defects. Defects include, among others, missing ink, misaligned printing, ink smears, and damage to the banknote's physical integrity.

Defects can be *severe, mild* or *light*.[2] Severe defects are often identifiable by the human eye and, if detected, lead to the disposal of the entire production batch, due to their potential to compromise the fulfillment of product quality standards. In contrast, mild and light defects, though undesirable, may not require such stringent measures unless they are compounded by other defects or specific conditions that could collectively compromise product quality.

We started with severe and mild defects and gradually incorporated lighter defects to improve variability and detection performance.

We collected 446 images, with around 48 percent *fit* instances (Table 1). Images are acquired through a less refined process of digital scanning, compared to Boria et al. (2023) and they are not restricted to represent any specific class of defect nor region or side of banknotes. As a result, available training instances comprise 21 different defect types occurring on both sides of banknotes and characterized by different severity levels.

[2] This classification adheres to standard practices, with procedures governed by established ECB guidelines and relevant International Organization for Standardization (ISO) norms. See also Boria et al. (2023).

Table 1 – Number of available samples for back and front-side images, according to defect class. Training and test sets result from stratified random sampling based on the binary target variable taking value 1 for fit banknotes, 0 otherwise.

		Total	Severe	Mild	Light	Fit
Back	Train	130	4	42	15	69
	Test	57	6	16	5	30
Front	Train	181	13	39	39	90
	Test	78	7	24	20	27

Since defects are not uniformly distributed in our sample, available instances cannot be considered fully representative of the entire spectrum of potential manifestations, as they are impossible to exhaustively enumerate. Certain defects being unevenly represented by severity class may partly reflect their characteristics; other defects may be particularly visible and therefore associated with the severe class, or *vice versa*. For instance, over 90 percent of defects of type *bite* and all available *setoff* defects, i.e., shadow or distorted version of elements from the other banknote side resulting from excessive pressure, belong to the *mild* class. Similarly, some defects may occur during a specific printing stage and thus materialize more frequently on a specific side of the banknote. Finally, *wiping* defects, i.e., ink smears transferring from one printed area to another, are only available for front side images, where they represent 35 percent of severe defects, 11 percent of mild defects and 12 percent of light defects.

2.3 Modeling

To handle varying defects, we follow a hybrid few-shot learning approach and train a Siamese network for each banknote side, each targeting the classification of *fit* banknotes. Our training strategy is based on images annotated as either *fit* or *unfit*. Since the *unfit* class encompasses a diverse and non-exhaustive range of defects, by severity, type, shape, size, location and so forth, the learning process relies on a limited number of examples while being characterized by high variability. With few-shot learning, models are trained to classify unseen defect types, following exposure to a few training instances representing the *unfit* class. Such approach is inspired by human cognition, which can recognize new objects or concepts based on their similarity to known categories, i.e., by leveraging knowledge about the relationships between classes (*fit* and *unfit*, in our setting).

Siamese networks are neural network architectures well suited to carry out few-shot learning tasks, as highlighted in Mul and Zuidema (2019). They are composed of two identical subnetworks sharing the same weights and parameters, to learn a common representation for input data. A Siamese network receives as input a pair of instances, each processed by a subnetwork, that it flags as either belonging to the same class or to different ones. When presented with a pair of input images, each subnetwork processes its assigned image independently, extracting relevant features. Notably, the pairwise approach of Siamese network fundamentally changes the volume of training data. Unlike traditional classifiers, that learn from N individual samples, Siamese architectures learn from the relationship characterizing pairs of observations; this shifts the learning target from individual instances to pairs, providing a significant advantage in data-constrained environments.[3]

[3] Given N observations, the number of unique pairs is determined by the so-called binomial coefficient $\binom{N}{2} = N(N-1)/2$.

In our application, the output of each subnetwork is mapped into a dense vector layer[4] before it reaches the output layer. This layer, whose elements correspond to the Euclidean distance between each feature vector resulting from the convolutional sequence, is then processed by the final binary classification layer to assess membership of the two images to the same class. By training the sub-networks on pairs of banknote images, the model learns to distinguish between similar and dissimilar instances, to effectively identify defects and anomalies even in the presence of significant variability. Our approach focuses on the classification of *fit* banknote images, i.e., during evaluation, we feed the Siamese network input pairs containing a testing image of an unknown class and a validation image that always belongs to the *fit* class.

The choice of the number of layers and other architectural hyper-parameters results from a calibration stage, based on previous analysis described in Boria et al. (2023). In our application, the two twin networks contain three convolutional layers each. A convolutional layer slides so-called filters across an input and performs a mathematical operation on groups of units, e.g., pixels in an image. This process allows the network to effectively capture and summarize visual information while preserving the spatial relationship between neighboring elements. We set the filter size shape to decrease with increasing layer depth, allowing the network to learn increasingly complex features.
Convolutional layers are alternated with pooling layers, aimed at extracting information from the filter maps resulting from convolution, while reducing their dimensionality. Pooling layers down-sample the data processed by convolutional layers by replacing a cluster of neighboring values with a single summary statistic, such as the maximum or average. These layers serve the purpose to retain the most important features while discarding redundant details, making the model more computationally efficient and less sensitive to small shifts in the input images.
The outputs from the two subnetworks are combined into a dense vector layer using the Euclidean distance. This is successively fed into a final binary output layer, reflecting the probability that the input pair belongs to the *fit* class.
Our training routine optimizes the binary cross-entropy loss function, and iteratively checks model performance on hold-out classification pairs from a random validation sample. Throughout the training, we apply the dropout regularization technique, with a dropout rate equal to 0.5, to improve the generalization abilities of our network.
To further avoid overfitting, i.e. the model learns to map noise and outliers in training data resulting in high in-sample accuracy and poor out-of-sample performance, we also employ early stopping, i.e. halting the training in case the model performance starts to deteriorate on the validation set. To avoid reliance of the learning process on training data's variability, we ex ante admit up to 75 training epochs and employ an adaptive learning rate, to let the model to converge to a reasonable solution more quickly.[5]

3. Results

To evaluate the performance of our proposed Siamese network architecture, we conduct an experimental analysis on two test datasets, comprising 57 instances for back-side and 78 instances for front-side images of banknotes, capturing a diverse range of defects and variations (see Table 1).

The datasets are curated to ensure a balance between fit and unfit banknotes, as well as a representative distribution of different defect types and severities. This section outlines the

[4] Dense (or fully connected) layers are key components of neural network architectures; they perform a linear transformation where every input is connected to every output node. Nodes, in turn, are fundamental computational units in a neural network that receive one or more inputs, process them, and transmit an output to the subsequent layer.
[5] See Figure A.1 in the Appendix on the convergence of the learning process.

quantitative and qualitative results of our experiments, focusing on the model's ability to accurately classify banknotes and identify potential defects.

Classification errors can arise from both false positives (i.e., the incorrect classification of an *unfit* instance as *fit*) and false negatives (i.e., the incorrect classification of a *fit* instance as *unfit*). In our setting, these errors entail different consequences and, *ceteris paribus*, a false positive instance is more costly than a false negative. False positives can lead to the disposal of a production batch after distribution, resulting in significant financial and reputational costs. False negatives, on the other hand, may require additional quality control measures or, in fully automated settings, could lead to the unnecessary disposal of batches. This can undermine the efficiency gains sought through automation.

In our framework, *fit* banknotes represent positive instances whereas *unfit* banknotes represent negative instances. To evaluate our models' performance, we focus on three main metrics: the ratio of correctly predicted fit instances (*precision)*, the ratio of correctly predicted fit instances to all fit observations in the class (*recall)* and the ratio of correctly predicted unfit instances to all unfit observations (*specificity)*.[5] All three are widely used metrics in machine learning, to evaluate the accuracy of a classifier, based on the number of true positive, true negative, false positive and false negative instances. Given the higher costs associated with false positives in our application, we prioritize models that optimize precision while maintaining an acceptable level of overall classification accuracy.

For each banknote side - *front* and *back* - we train our Siamese networks on two incremental sets of *unfit* images. We first use images with severe and mild defects, whose detection would imply production batch disposal, to represent the *unfit* class, and annotate images with light defects as fit.[6] We successively include light defect banknotes in the *unfit* set, and thus reduce size of the fit training sample from 129 to 90 and from 84 to 59 instances, respectively, for *front* and *back* side. In principle, extending the unfit set to lighter defects could increase their variability and representativeness (in terms of size, intensity, shape, location, *etc.*), to improve detection performance. At the same time, the high similarity between *fit* and *light* defect instances might reduce the recall ability of the model, which would tend to be more conservative in its *fit* scores.

Table 2 presents the performance metrics for detecting defects on both the *front* and *back* sides of the samples. Overall, the metrics suggest that the detection system performs better when considering all defects, achieving higher precision, recall, and specificity values compared to when only severe and mild defects are considered.

The networks trained on more visible defects only, i.e. severe and mild, tend to systematically misclassify fit banknotes as unfit. In detail, precision is equal to 0.5 and 0.8, respectively for front and back sides; recall is especially low, equal to 0.074 for front side and to 0.267 for back; and finally, the specificity is 0.961 and 0.926, for front and back sides.[7]

When considering the Siamese network trained on all defects, the precision on the considered test set is perfect, at 1.0, and very high, at 0.95, for *front* and *back* sides, respectively. The recall also improves significantly: to 0.926 for front side, 0.633 for back side. Finally, training on all defect severities allows Siamese networks to achieve a strong ability to correctly identify non-defective samples as their specificity increases to 1. and 0.963, for front and back side.

[6] See Section A.1 of the Appendix for details.
[7] Values are robust to the case of light defect images labelled as *fit*.

Table 2 - Performance metrics for fit banknote detection on front and back sides. The table shows precision, recall, and specificity values for detecting fit banknotes given a training set comprising severe and mild defects only and all defects, on both the front and back sides of the samples.

	Training	Precision	Recall	Specificity
Front	Severe and mild defects	0.5	0.074	0.961
	All defects	**1.0**	**0.926**	**1.0**
Back	Severe and mild defects	0.8	0.267	0.926
	All defects	**0.95**	**0.633**	**0.963**

Given the superior performance of the Siamese networks trained on all defects, we further assess their characteristics to gain deeper insights into their strengths and limitations.

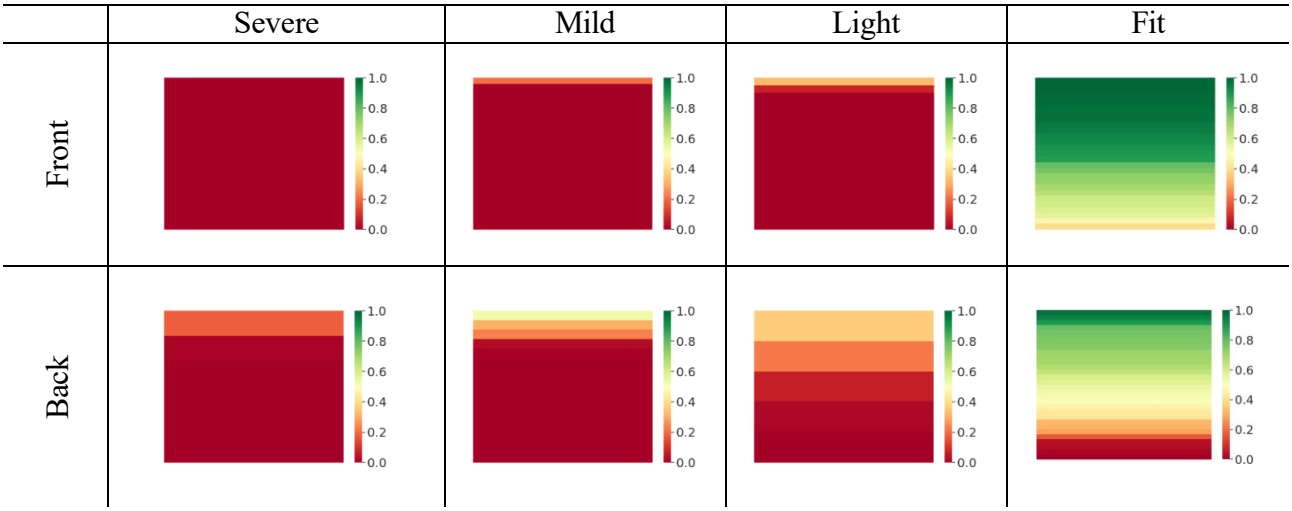
Being primarily designed to identify *fit* banknotes, our models learn the characteristics of instances from such class and flag any significant deviations from such norm. This implies our models are not explicitly trained to recognize defects by type, nor to be sensitive to their location or severity. However, we observe a positive correlation between the sensitivity of the networks to defects' severity, due to the associated impact on the overall appearance of the banknote (Figure 1). The plots illustrate the models' performance in classifying images based on the defect severity, with each column representing a different severity class, in decreasing order from severe to fit. The two rows correspond to the two sides of the images: *front* and *back*. The plots provide insights on the strong correlation between the defect severity and the scores assigned by the models: more severe defects tend to receive lower scores whereas fit banknotes receive high scores.

Empirical results suggest both models are effectively distinguishing between different levels of defect severity, despite their being not explicitly requested to, because of the larger size or higher intensity of severe defects compared to light ones. By setting a 0.5 threshold value, we obtain two false negatives, i.e. two fit banknotes labelled as unfit, among front-side instances, albeit with relatively high values, and none among the back-side ones. It is noteworthy that we get one single false positive for a mild defect on back-side images, with score 0.538.[8] The substantial lack of false positive is a desirable property for our application: it suggests that our modeling approach is conservative in its defect identification, prioritizing precision.[9]

[8] See also Fig. A.2 in the Appendix.

[9] If we considered a conservative 0.75 threshold, we would achieve maximum precision among back-side images but would halve the recall rate to around 0.3. Conversely, if we lowered the threshold to 0.25, to further reduce likelihood of false negatives, we would get 0.833 recall score for the back-side set but would reduce precision and specificity by misclassifying as positives three instances, two with mild and one with light defects.

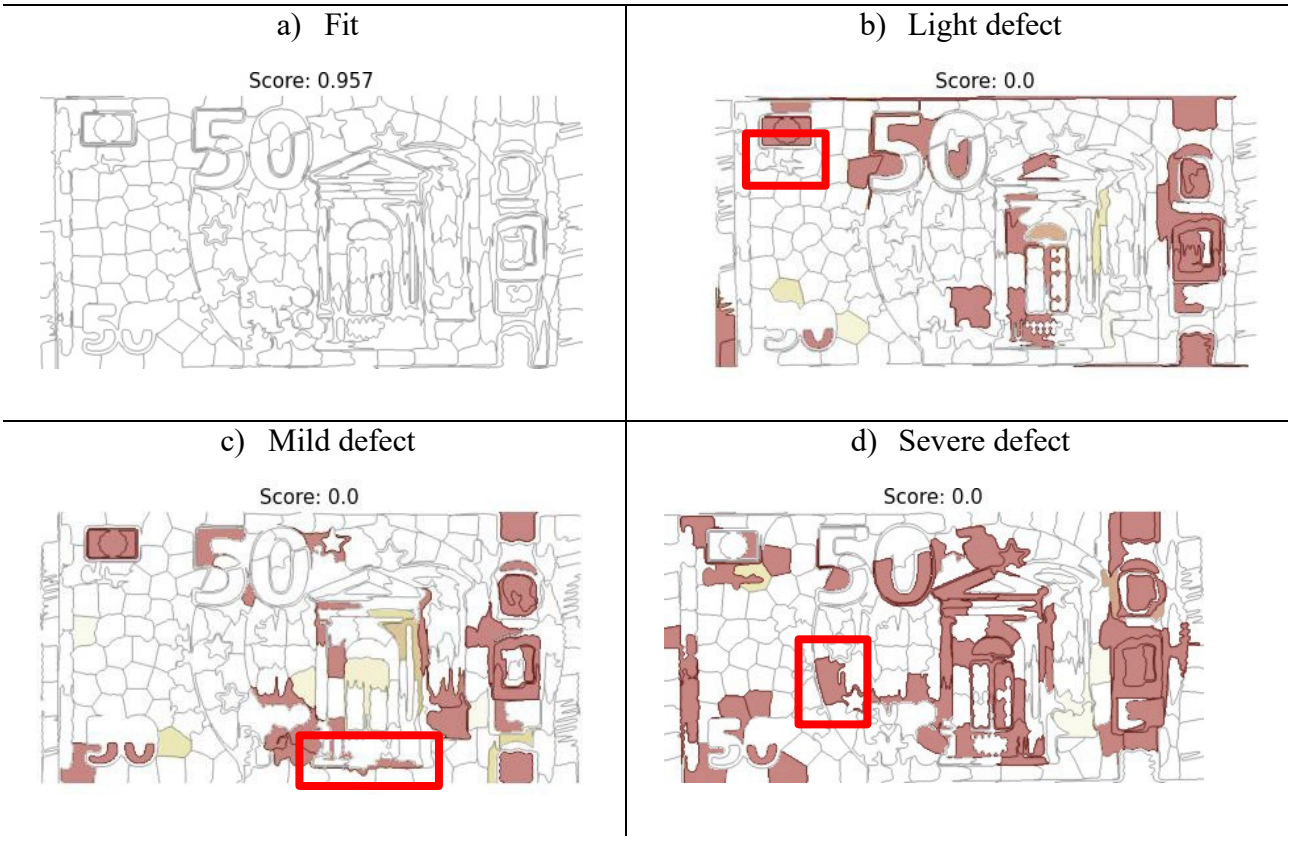
Figure 1 - Classification scores by defect severity and image side. The plots display the Siamese networks' classification scores for images based on defect severity, with each column representing a different severity class. The two rows correspond to the two sides of the images, and associated Siamese network model.



To gain deeper insights into the *fit* score assigned by our Siamese networks, we employ a segmentation-based approach, building upon Fedele et al. (2024). We leverage Simple Linear Iterative Clustering (SLIC) segmentation to identify homogeneous segments from the test image and apply them to the most similar fit image as well. SLIC is a segmentation algorithm introduced by Achanta et al. (2010). SLIC segmentation groups adjacent pixels into so-called *superpixels* to partition an image into meaningful regions, thereby reducing computational complexity for subsequent image-processing tasks. Through the iterative comparison of corresponding segments from the two images, we identify regions where the similarity between the images improves significantly as differences are cancelled out by replacing the test image's segment with the fit one's.

We combine the quantitative information from the Siamese network with qualitative insights provided by the segmentation-based visualization. Our proposal aims at supporting human experts' understanding of the network's output and pinpointing specific areas where the test image deviates from the fit image (Figures 2 and 3). Changes in similarity between the images, after each segment replacement, are visualized using color-coded intensities. More intense red-orange colors indicate regions where the similarity changed significantly, suggesting potential defects within or adjacent to the segments. As a technical remark, the intensity of the color coding is also influenced by a normalization process, which helps to emphasize the most significant deviations. We point out that the extent to which similarity can be increased through segment replacement is also influenced by the initial similarity between the test and fit images: when the initial similarity is very high, i.e. the Siamese network assigns the test image to the *fit* class, the score is already close to its maximum value of 1 and there is limited room for improvement through segment swapping. Consequently, fewer regions are highlighted (see panel *a* of Figures 2 and 3). Conversely, when the initial similarity is lower, there is more potential for increasing the score through segment replacement, leading to a larger number of highlighted regions (see panel *d* of Figures 2 and 3).

Figure 2 – Examples of scoring decisions for a fit banknote (panel a) and for banknotes with light, mild and severe defects (panels b, c and d, respectively) by the Siamese network model trained on front side images. The red squares identify the area where the defect is located.

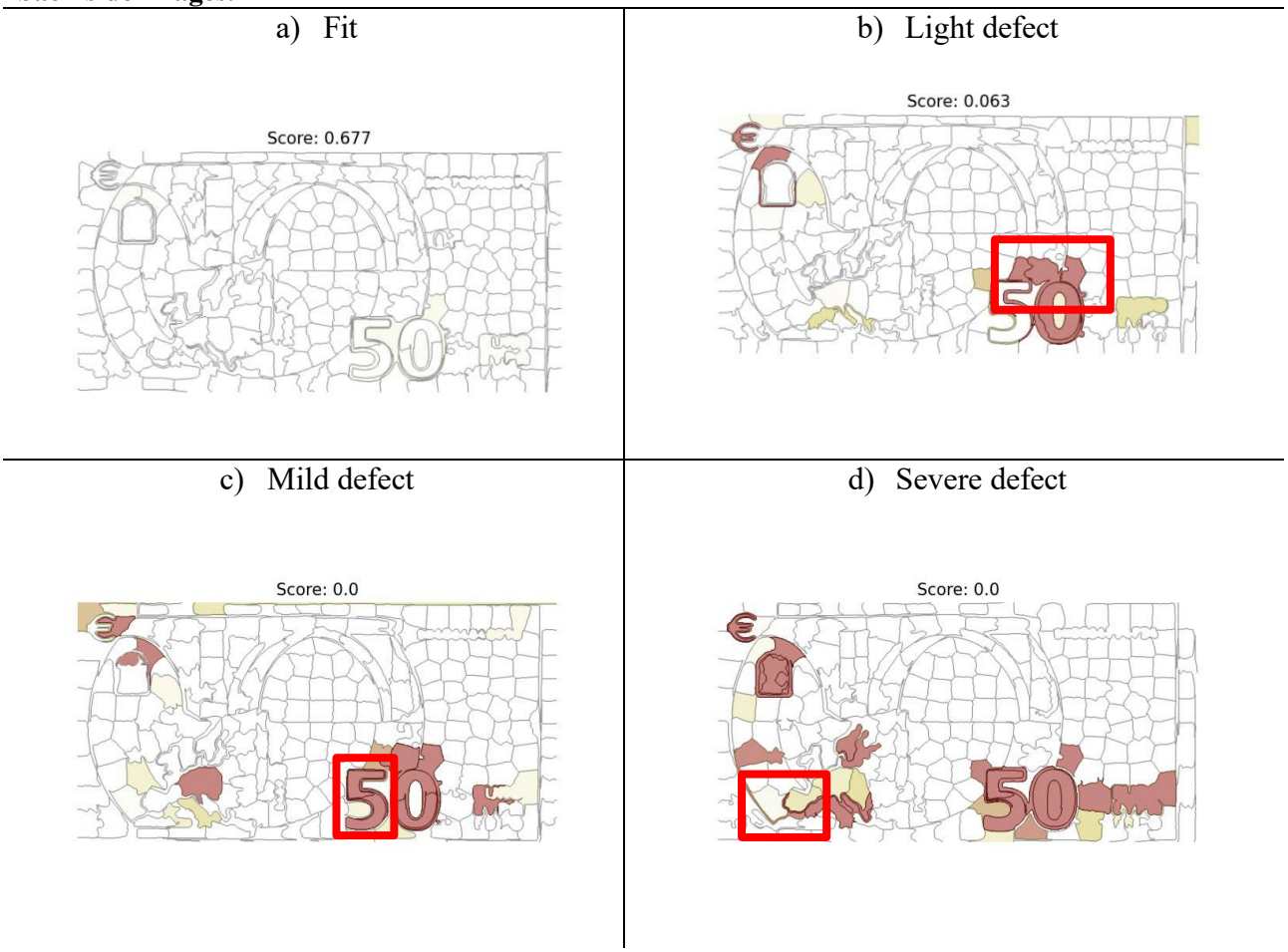


This segmentation-based approach provides valuable and transparent insights. However, it leverages structurally different assumptions compared to Siamese network models. While designed specifically for Siamese networks, this approach operates by considering the input shape rather than the internal parameters of the model, i.e. it is *de facto* inherently agnostic to the specific architecture. This means that, *ceteris paribus*, the segmentation-based investigation could be applied to any Siamese network without entailing modifications to the segmentation process. Furthermore, segmenting the image into fixed regions underlies a strong assumption on the reasoning paths of a given Siamese network. Since Siamese network architectures process input information to derive a vector layer, they allow for the integration of pixels from multiple regions of each image. In other words, elements of the vector produced by the Siamese network could reflect information from spatially distant regions, potentially misaligning with any segmentation-based approach.

Keeping in mind the differences characterising the Siamese networks' functioning and the segmentation-based explainability approach, evaluation of the latter should consider two main aspects. The first is the alignment of signals: in high-score scenarios, i.e., the Siamese network assigns a high score to a test image, we expect the absence of darker regions in the segmentation visualization; conversely, we expect low-score images to flag segments, to indicate a strong correlation between the network's decision and the explainability approach. The second aspect pertains to the localization of defects on specific regions. Given a low score, human experts should prioritize the evaluation of darker regions and their immediate surroundings to pinpoint potential defects. However, as explained above, human experts may face more than a single area being highlighted. Due to segment-swapping, regions flagged by the explainability approach may not be the ones hosting the defect nor the ones surrounding it. This might be due to subtle variations in color intensity, texture, or other visual cues

not easily perceptible to the human eye, that may lead the Siamese network to flag additional regions. Although we aim to minimize differences between the two images by picking the most similar fit one, consistent level of noise may persist. For instance, discrepancies often arise in the regions where the hologram is located (see panels *b*, *c* and *d* of Figure 2) or at color-shifting ink patches (see panels *b*, *c* and *d* of Figure 3).[10]

Figure 3 – Examples of scoring decisions for a fit banknote (panel *a*) and for banknotes with light, mild and severe defects (panels *b, c* and *d*, respectively) by the Siamese network model trained on back side images.



4. Conclusions

In this paper, we used Siamese networks for automated banknote quality control. We leveraged few-shot learning to address the inherent challenges that characterize the banknote production process, particularly the impossibility of exhaustively enumerating all potential defect occurrences.

While few-shot learning is well-suited to the task at hand, data constraints may still limit classification accuracy. Future experimental stages of the application could contribute to the collection of additional samples, further validating the model's performance and potential for real-world deployment. As the model's performance may already be near its theoretical maximum, increasing the dataset size may be especially useful for optimizing the classification threshold.

[10] See Figure A.2 in the Appendix on incorrectly classified instances.

The explainability approach we propose represents a novel contribution to the field, differing from previous work such as (Fedele et al., 2024) in its reliance on region swapping rather than perturbation or replacement. The explainability method leverages segmentation, which might not be inherently aligned with the operational principles of Siamese networks in their mapping information from an image's region to the signal vector. Nevertheless, it offers a valuable tool for understanding model decisions. The development of a model-specific explainability method, tailored to the unique architecture of Siamese networks, could further support human analysis.

References

Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., & Süsstrunk, S. (2010). *Slic superpixels.*

Boria, K., Luciani, A., Marchetti, S., & Viticoli, M. (2023). *Siamese neural networks for detecting banknote printing defects (No. 34)*. Bank of Italy, Directorate General for Markets and Payment System.

Deshpande, A. M., Minai, A. A., & Kumar, M. (2020). *One-shot recognition of manufacturing defects in steel surfaces*. Procedia Manufacturing, 48, 1064-1071.

Fedele, A., Guidotti, R., & Pedreschi, D. (2024). *Explaining Siamese networks in few-shot learning.* Machine Learning, 1-38.

Fei-Fei, L., Fergus, R., & Perona, P. (2006). *One-shot learning of object categories.* IEEE transactions on pattern analysis and machine intelligence, 28(4), 594-611.

Fink, M. (2004*). Object classification from a single example utilizing class relevance metrics.* Advances in neural information processing systems, 17.

Jadhav, M., Kumar Sharma, Y., & Bhandari, G. M. (2019). *Currency identification and forged banknote detection using deep learning*. 2019 International conference on innovative trends and advances in engineering and technology (ICITAET) (pp. 178-183). IEEE.

Ke, W., Huiqin, W., Yue, S., Li, M., & Fengyan, Q. (2016). *Banknote image defect recognition method based on convolution neural network*. International Journal of Security and Its Applications, 10(6), 269-280.

Koch, G., Zemel, R., & Salakhutdinov, R. (2015). *Siamese neural networks for one-shot image recognition*. ICML deep learning workshop (Vol. 2, No. 1, pp. 1-30).

Mohr, J., Breidenbach, F., & Frochte, J. (2021). *An Approach to One-shot Identification with Neural Networks*. IJCCI (pp. 344-351).

Mul, M., & Zuidema, W. (2019). *Siamese recurrent networks learn first-order logic reasoning and exhibit zero-shot compositional generalization*. arXiv preprint arXiv:1906.00180.

Nazir, B., Imran, M., & Jehangir, B. (2024). *Banknote Verification using Image Processing Techniques*. Journal of Computing & Biomedical Informatics, 7(01), 125-143.

Palatucci, M., Pomerleau, D., Hinton, G. E., & Mitchell, T. M. (2009). Zero-shot learning with semantic output codes. *Advances in neural information processing systems*, *22*.

Parnika, J., Sen, S., Divya, R., Deepika, S., & Dharshini, V. (2023). *An Explainable Machine Learning Model to Analyze and Detect Fake Currency*. 2023 International Conference on Communication, Circuits, and Systems (IC3S) (pp. 1-6). IEEE.

Sadyk, U., Baimukashev, R., & Turan, C. (2024*). State-of-the-Art Review of Deep Learning Methods in Fake Banknote Recognition Problem*. International Journal of Advanced Computer Science & Applications, 15(1).

Sawant, V. M., Tupe, R. D., Tawade, A. S., & Bhalerao, S. M. (2022). *Fake Currency Identification System Using CNN*. International Journal of Wireless Network Security, 8(1), 18-23.

Sultana, S. S., Bhargavi, M., Devi, R., Niharika, B., Rishitha, C., & Seswitha, A. (2024). *Deep Learning-based Banknote Classification: Harnessing Artificial Neural Networks*. 2024 3rd International Conference on Applied Artificial Intelligence and Computing (ICAAIC) (pp. 21-27). IEEE.

Venu, D. N. (2023). *An Automatic recognition system of fake Indian currency notes detection using Image processing analysis*. European Chemical Bulletin, 12(9), 280-307.

Appendix

A.1 Evaluation metrics

Precision measures the share of positive predictions (*fit* banknotes) made by a model that is actually correct. Recall and specificity measure the proportion of actual positive (*fit* banknotes) and negative instances (*unfit* banknotes) that were correctly identified by the model, i.e., they are nothing but the true positive and true negative rates, respectively. For a fixed accuracy level, a classifier with higher precision makes fewer false positive errors, whereas a classifier with higher specificity makes fewer false negative errors.

In detail, precision is defined as:

$$Precision = \frac{n.\,true\ positives}{n.\,true\ positives + n.\,false\ positives}$$

recall is defined as:

$$Recall = \frac{n.\,true\ positives}{n.\,true\ positives + n.\,false\ negatives}$$

and specificity is defined as:

$$Specificity = \frac{n.\,true\ negatives}{n.\,true\ negatives + n.\,false\ positives}$$

A.2 Additional figures

Figure A.1 – Loss function values during training.

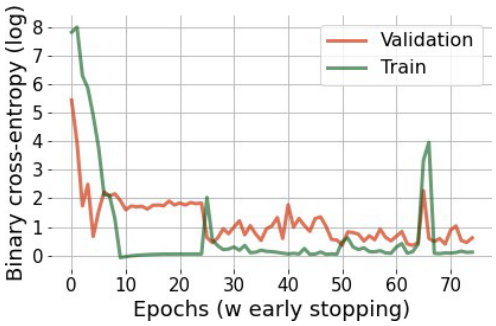
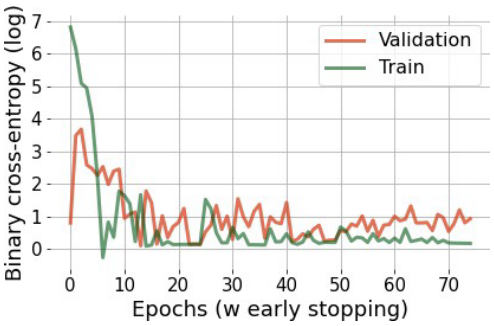
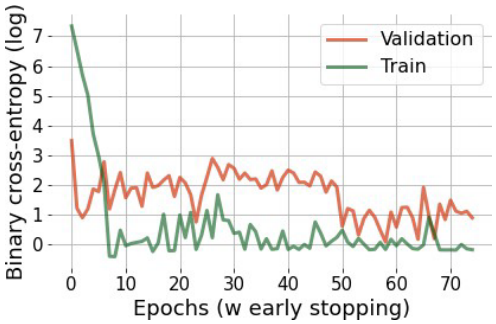
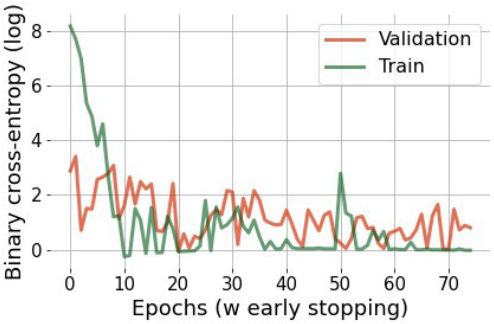
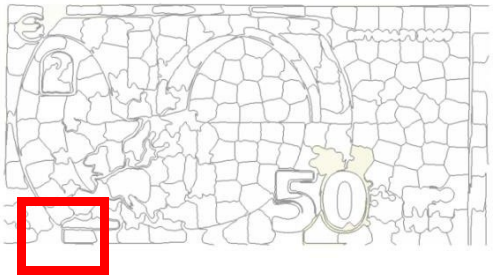
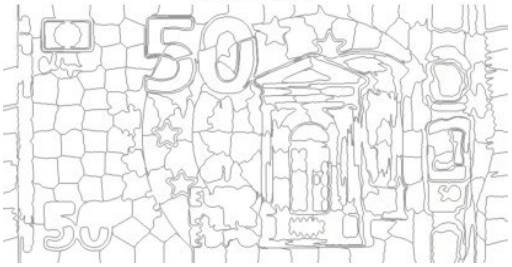
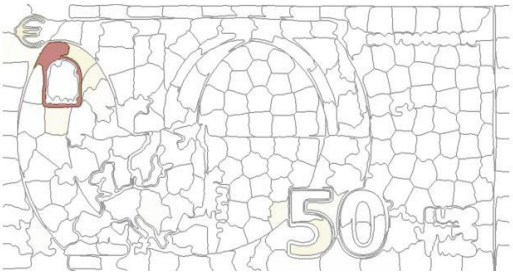
	Severe and mild defects	All defects
Back		
Front		

Figure A.2 – Examples of classification errors

<table>
<tr><td colspan="2">False positive(s)</td></tr>
<tr><td>a) Front

NA</td><td>b) Back
Score: 0.538
  </td></tr>
<tr><td colspan="2">False negatives</td></tr>
<tr><td>c) Front
Score: 0.469
  </td><td>d) Back
Score: 0.285
  </td></tr>
</table>

Recently published papers in the 'Markets, Infrastructures, Payment Systems' series

n. 45 Sustainability at shareholder meetings in France, Germany and Italy, *by Tiziana De Stefano, Giuseppe Buscemi and Marco Fanari* (in Italian)

n. 46 Money market rate stabilization systems over the last 20 years: the role of the minimum reserve requirement, *by Patrizia Ceccacci, Barbara Mazzetta, Stefano Nobili, Filippo Perazzoli and Mattia Persico*

n. 47 Technology providers in the payment sector: market and regulatory developments, *by Emanuela Cerrato, Enrica Detto, Daniele Natalizi, Federico Semorile and Fabio Zuffranieri*

n. 48 The fundamental role of the repo market and central clearing, *by Cristina Di Luigi, Antonio Perrella and Alessio Ruggieri*

n. 49 From Public to Internal Capital Markets: The Effects of Affiliated IPOs on Group Firms, *by Luana Zaccaria, Simone Narizzano, Francesco Savino and Antonio Scalia*

n. 50 Byzantine Fault Tolerant consensus with confidential quorum certificate for a Central Bank DLT, *by Marco Benedetti, Francesco De Sclavis, Marco Favorito, Giuseppe Galano, Sara Giammusso, Antonio Muci and Matteo Nardelli*

n. 51 Environmental data and scores: lost in translation, *by Enrico Bernardini, Marco Fanari, Enrico Foscolo and Francesco Ruggiero*

n. 52 How important are ESG factors for banks' cost of debt? An empirical investigation, *by Stefano Nobili, Mattia Persico and Rosario Romeo*

n. 53 The Bank of Italy's statistical model for the credit assessment of non-financial firms, *by Simone Narizzano, Marco Orlandi and Antonio Scalia*

n. 54 The revision of PSD2 and the interplay with MiCAR in the rules governing payment services: evolution or revolution?, *by Mattia Suardi*

n. 55 Rating the Raters. A Central Bank Perspective, *by Francesco Columba, Federica Orsini and Stefano Tranquillo*

n. 56 A general framework to assess the smooth implementation of monetary policy: an application to the introduction of the digital euro, *by Annalisa De Nicola and Michelina Lo Russo*

n. 57 The German and Italian Government Bond Markets: The Role of Banks versus Non-Banks. A joint study by Banca d'Italia and Bundesbank, *by Puriya Abbassi, Michele Leonardo Bianchi, Daniela Della Gatta, Raffaele Gallo, Hanna Gohlke, Daniel Krause, Arianna Miglietta, Luca Moller, Jens Orben, Onofrio Panzarino, Dario Ruzzi, Willy Scherrieble and Michael Schmidt*

n. 58 Chat Bankman-Fried? An Exploration of LLM Alignment in Finance, *by Claudia Biancotti, Carolina Camassa, Andrea Coletta, Oliver Giudice and Aldo Glielmo*

n. 59 Modelling transition risk-adjusted probability of default, *by Manuel Cugliari, Alessandra Iannamorelli and Federica Vassalli*

n. 60 The use of Banca d'Italia's credit assessment system for Italian non-financial firms within the Eurosystem's collateral framework, *by Stefano Di Virgilio, Alessandra Iannamorelli, Francesco Monterisi and Simone Narizzano*

n. 61 Fintech Classification Methodology, *by Alessandro Lentini, Daniela Elena Munteanu and Fabrizio Zennaro*

n. 62 The Rise of Climate Risks: Evidence from Expected Default Frequencies for Firms, *by Matilde Faralli and Francesco Ruggiero*

n. 63 Exploratory survey of the Italian market for cybersecurity testing services, *by Anna Barcheri, Luca Bastianelli, Tommaso Curcio, Luca De Angelis, Paolo De Joannon, Gianluca Ralli and Diego Ruggeri*

n. 64 A practical implementation of a quantum-safe PKI in a payment systems environment, *by Luca Buccella and Stefano Massi*

n. 65 Stewardship Policies. A Survey of the Main Issues, *by Marco Fanari, Enrico Bernardini, Elisabetta Cecchet, Francesco Columba, Johnny Di Giampaolo, Gabriele Fraboni, Donatella La Licata, Simone Letta, Gianluca Mango and Roberta Occhilupo*

n. 66 Is there an equity greenium in the euro area?, *by Marco Fanari, Marianna Caccavaio, Davide Di Zio, Simone Letta and Ciriaco Milano*

n. 67 Open Banking in Italy: A Comprehensive Report, *by Carlo Cafarotti and Ravenio Parrini*

n. 68 Report on the payment attitudes of consumers in Italy: results from ECB SPACE 2024 survey, *by Gabriele Coletti, Marialucia Longo, Laura Painelli, Emanuele Pimpini and Giorgia Rocco*

n. 69 A solution for cross-border and cross-currency interoperability of instant payment systems, *by Domenico Di Giulio, Vitangelo Lasorella, Pietro Tiberi*

n. 70 Do firms care about climate change risks? Survey evidence from Italy, *by Francesca Colletti, Francesco Columba, Manuel Cugliari, Alessandra Iannamorelli, Paolo Parlamento and Laura Tozzi*

n. 71 Demand and supply of Italian government bonds during the exit from expansionary monetary policy, *by Fabio Capasso, Francesco Musto, Michele Pagano, Onofrio Panzarino, Alfonso Puorro and Vittorio Siracusa*

n. 72 Statistics on tokenized financial instruments: A challenge for central banks, *by Riccardo Colantonio, Massimo Coletta, Riccardo Renzi*

n. 73 Credit Risk Assessment with Stacked Machine Learning, *by Francesco Columba, Manuel Cugliari, Stefano Di Virgilio*

n. 74 What if Ether Goes to Zero? How Market Risk Becomes Infrastructure Risk in Crypto, *by Claudia Biancotti*

n. 75 The Cyber Risk of Non-Financial Firms, *by Francesco Columba, Manuel Cugliari, Marco Orlandi, Federica Vassalli*

n. 76 Sustainability and financial innovation: The emerging role of Fintech for Good (F4G), *by Alessandro Lentini and Daniela Elena Munteanu*

n. 77 Hydrogeological and credit risk: the Italian firms' physical risk-adjusted probability of default, *by Manuel Cugliari, Simone Narizzano and Federica Vassalli*

n. 78 Liquidity Optimization in Gross Settlement Systems with Quantum Reordering: Application to TARGET2, *by Valerio Astuti, Adriano Baldeschi, Luca Bastianelli, Giuseppe Bruno, Ajit Desai, Danica Marsden and Riccardo Russo*

n. 79 The expert assessment within Banca d'Italia's in-house credit assessment system, *by Lorenzo Esposito, Massimo Guglielmi, Francesco Monterisi, Simone Narizzano and Marco Orlandi*

n. 80 Digital payments and economic performance: evidence from Italy, *by Guerino Ardizzi, Niccolò Lippi Boncambi, Cristina Demma and Alberto Leorati*

n. 81 The euro-area repo market: structure, participants and interlinkages, *by Lorenzo Caverni, Annalisa De Nicola and Mattia Persico*