Preprints.org

Article

Enhancing Fingerprint Forensics: A Comprehensive Study of Gender Classification Based on Advanced Data-Centric AI Approaches and Multi-Database Analysis

Assaf Spanier [*] , Dor Steiner , Navon Sahalo , Yoel Abecassis , Dan Ziv , Ido Hefetz , Shimon Kimchi

Article

Enhancing Fingerprint Forensics: A Comprehensive Study of Gender Classification Based on Advanced Data-Centric AI Approaches and Multi-Database Analysis

Assaf B. Spanier [1],*, Dor Steiner [1], Navon Sahalo [1], Yoel Abecassis [1], Dan Ziv [2], Ido Hefetz [3] and Shimon Kimchi [3]

[1] School of Software Engineering and Computer Science, Azrieli College of Engineering, Jerusalem, Israel
[2] Latent Fingermark Development Laboratory, Division of Identification and Forensic Sciences, Israel Police, Jerusalem, Israel
[3] Fingerprint Database Laboratory, Division of Identification and Forensic Sciences, Israel Police, Jerusalem, Israel
* Correspondence: shpanier@jce.ac.il

Abstract: Use of fingerprints found at a crime scene is a common practice for identifying suspects in criminal investigations. Over the past two decades, attempts have been made to obtain additional information from fingerprints, beyond locating suspects as part of an investigation. This includes gender, age and nationality. Researchers demonstrated 75%-90% accuracy in gender classification based on fingerprint images. Nonetheless, despite promising results, these studies have several significant shortcomings with respect to their practical feasibility. First, they ignore the low quality and quantity of fingerprints collected from the crime scene since typically the scene of a crime has only one fingerprint collected, and the fingerprint might be partially or of poor quality as well. Second, as most results are based on a single database, public or private, it is difficult to generalize the most suitable method. Third, studies miss the untapped potential of Data-Centric AI (DCAI) approaches for improving results. The aim of this study was to compare, for the first time, the gender classification from a fingerprint using several datasets and with varying fingerprint image quality. The results from four databases are compared, three public and one internal private database. In addition, we utilize the latest Data-Centric AI (DCAI) approaches for improving classification results. The results demonstrate that a conservative Convolutional Neural Networks (CNN) such as VGG is sufficient for this task. Classification accuracy ranges from 80% to 95% depending on the quality of the fingerprint, with DCAI approaches adding 1%-4% improvement. For partially or low-quality fingerprint images, the periphery of a fingerprint is the most significant area for determining gender. The source code is also provided here for practical application.

Keywords: fingerprint analysis; gender classification; Data-Centric AI (DCAI); forensic science; partial or low-quality fingerprint images

Introduction

A fingerprint found at a crime scene is important evidence in a criminal investigation. Fingerprints are collected and compared against an existing database of fingerprint images in the police database. Comparison is made by finding minutiae features [1,2] and comparing them to those recorded in the police database. If a match is found, the suspect's identity can be established. While minutiae marking conventionally undergoes automated procedures, occasions may arise necessitating manual interventions, specifically involving the removal and modification of minutiae, owing to the presence of artifacts. Beyond the fact that extraction of minutiae features manually is tedious work and prone to errors, the main disadvantages of this technique are: (1) fingerprints that are not of sufficient value are excluded, and (2) attempts are not made to extract additional

information beyond comparison with the database, for example gender, age and other information [3] . Fingerprints as a gender classification method is an exciting research frontier in forensic science. Research on this topic conducted by Acree in 1999 [4] focused on analysis of ridge density and ridge width distributions to differentiate between male and female fingerprints. Acree posited that fingerprint ridge density is a reliable parameter for gender determination and observed higher ridge density in females then males. Over 20 years of research conducted and covered in the latest literature review from 2021 [5] concluded that the number of ridges present on a fingerprint per unit area varies between individuals and shows differences based on gender. However, the mean density of fingerprint ridges varies

between populations, as demonstrated by the different datasets investigated in various studies. Moreover, a study titled "What can a single tell about gender?", published in [6], employed image processing techniques, specifically focusing on minutiae detection and extracting image features centered around this point of interest. However, with 65% accuracy in classifying gender, this study demonstrates a moderately favorable outcome

Alternative approaches to gender classification based on fingerprint ridges use advanced machine learning and deep learning methods based on the whole fingerprint image. In [7] the authors introduce an approach that leverages Fast Fourier Transform (FFT), Principal Component Analysis (PCA) features, and min-max normalization. The models are trained utilizing a Support Vector Machine (SVM) classifier and incorporate sampling techniques such as SOMAT to address dataset imbalance. In this study the right ring finger emerged as the most suitable point of interest for gender identification. It achieves an accuracy rate of 75% and 91% for males and females, respectively. The study refrained from employing deep learning methodologies and relied on a single dataset. Another article [8] presents 99% accuracy, however, the article does not provide references or specifics regarding the precision of the train-test division when considering the occurrence of two impressions for each fingerprint and the gender balance, resulting in uncertainty regarding those results.

One of the most common publicly available databases for gender classification is SOCOfing. In [9] gender classification performance was evaluated using single finger-based classification, resulting in 77% average accuracy. Remarkably, when employing a weighted approach that considered multiple fingers simultaneously, accuracy significantly improved, achieving a 90% accuracy rate. However, a gender classification based on multiple fingerprints is theoretical. According to our knowledge, it has no practical application in crime scene investigation. This is primarily due to the rarity of encountering more than one fingerprint at a crime scene, as supported by prior research [10]. In this study the authors employed only an elementary 5-layer CNN [11] architecture. Moreover, it is noteworthy that the network was trained "from scratch," without the use of any beneficial fine-tuning techniques. Another study conducted on the SOCOfing dataset [12], reported an accuracy rate of 90% on a restricted subset comprising 2000 fingerprints out of the total 6000 fingerprints available in the SOCOfing database. Notably, there is limited clarity regarding the specific criteria employed to select the 2000 fingerprints from the larger pool of 6000. In summary, research articles focusing on the SOCOfing database have reported gender classification outcomes of approximately 75% when employing neural network-based approaches. When multiple fingerprints are used (though not practically applicable), accuracy rates as high as 90% were observed. Hence, it is evident that a baseline accuracy of around 75% is commonly encountered in gender classification tasks conducted on the SOCOfing dataset. This is indicated by references such as [13] and others.

In the context of studies utilizing advanced techniques on private fingerprint datasets, a recent comprehensive publication [14] conducted an extensive comparative analysis. This analysis encompassed nine widely employed classifiers, including CNN, Support Vector Machine (SVM) with three distinct kernels, k-Nearest Neighbors (kNN), Adaboost, J48, ID3, and Linear Discriminant Analysis (LDA). Their methodology involved gender classification based on various fingerprint features, among them ridge-density, ridges, minutiae, and fingertip-size (FTS). The results revealed that CNN achieved the highest success rate for gender classification in this context. The results showed an accuracy of 95%. However, it is important to note that these results were based on private, high-quality fingerprint images collected in controlled laboratory settings (e.g., 500 dpi resolution

with each finger scanned three times to ensure image quality), which may not be representative of real-world crime scene scenarios. Moreover, a somewhat redundant methodology that uses CNN for classification was utilized when it is essentially a method for extracting features.

The Data-Centric Artificial Intelligence (DCAI) approach shifts the AI paradigm by placing more emphasis on data rather than algorithms or models [15,16]. Instead of solely concentrating on the model or algorithm itself, this approach analyzes data to identify which parameters instances have the most significant impact on classification. It then re-trains models based on these influential instances. In recent years, this method has proven to offer distinct advantages over traditional approaches. DCAI represents an alternative paradigm in artificial intelligence that prioritizes data over algorithms. By implementing a set of margin-based criterion, we effectively filter out uncertain classification, and as mentioned, enabling the model to learn from more reliable and informative examples. This approach has demonstrated the efficacy of the margin of confidence clean method in improving the performance and generalization capabilities of deep learning models.

To conclude, this study's primary contributions encompass several key aspects:

1. Assess the performance of a singular CNN methodology across a diverse range of datasets, public and private, free of prior assumptions or preprocessing steps.
2. Classification in scenarios of partially or low-quality fingerprint images by identification and delineation of the Region Of Interest (ROI) that most contributes to gender classification.
3. Improved classification results using DCAI approaches.

Datasets

In this paper we evaluate four datasets as detailed in Table 1. Three of these datasets are publicly accessible, namely: (1) NIST-DB4 [17], (2) SOCOfing [18], and (3) NIST-302 [19]. Additionally, we incorporated a privately obtained dataset from the Israeli police, henceforth referred to as IsrPoliceDB. To the best of our knowledge, these datasets represent the only publicly available resources that include gender information in conjunction with fingerprint images. The NIST-DB4 dataset [20] comprises 4000 plain fingerprints, captured at a resolution of 500 dpi, sourced from 2000 individuals (380 females and 1620 males), where each subject contributed two impressions of their index finger for this dataset. The SOCOfing database [18], established in 2007, consists of 6000 fingerprints collected from 600 African subjects (123 females and 477 males). Each participant provided a single impression of all ten fingers. The images in this dataset measuring 96 x 103 pixels and having a resolution of 500 dpi. Furthermore, a relatively recent dataset, NIST-302 [21], published by the NIST agency in 2019, includes fingerprint impressions from 200 Americans (132 females and 68 males). Each individual contributed 20 impressions of all ten fingers. The the images in this dataset are at a resolution of 500 dpi, with dimensions of 256 x 360 pixels

Table 1. Summary of Fingerprint Databases Utilized in the Study. The databases considered/used in the study include IsrPoliceDB, NIST-DB4, SOCOfing, and NIST-302. The 'X' and 'V' symbols signify the absence and presence, respectively, of specific characteristics within each database.

Database Name	Age	Gender	Ethnic Group	# of impressions	Image Resolution	Type	Size (male / female)
IsrPoliceDB	X	V	Israel	1	JPG 1000dpi 688x975px	Plan	1271 619 (50%) / 652 (50%)
NIST-DB4	X	V	Not mentioned	2	PNG, 96dpi,	Plan (2 impressions	4000 3250 (81%) / 750 (19%)

				512 X 512px 8bit, gray	of each finger)	
SOCOfing	X	V	African	1	BMP, 500dpi, 96x103px	Plan 600 subjects x 10 fingers x 1 impression	6000 4770 (75%) / 1230 (25%)
NIST-302	V	V	American	20+	PNG, 500dpi, 256 X 360px	plain, rolled, 4-finger, 2-thumb, palm	2000 1583 (75%) / 396 (25%)

Methods

Top of Form

Bottom of Form

The methods section is divided into three key parts. The first part discusses the CNN techniques examined, specifically VGG and ResNet. The second part centers on delineating the use cases for evaluating CNN's performance in handling partially or low-quality fingerprint images. In the third part, we discuss the DCAI approaches implemented to enhance our findings' performance.

The input of our method is a single fingerprint image, and the output a gender classification. This study employed supervised learning and CNN methods for model training. In particular, networks, such as VGG [22] and ResNet [23] which have maintained their significance within the realm of deep learning for image classification. The architectures emphasize depth through the recurrent use of small receptive field convolutions and pooling layers, simplifying implementation and comprehension. ResNet [23] introduced shortcut connections known as residual blocks, however VGG's [22] adaptability remains a significant advantage. By incorporating pre-trained weights into both models, they can both be seamlessly integrated into diverse networks and deep learning tasks, emphasizing their versatility. VGG and ResNet should be carefully considered in the context of smaller datasets. The less complex architecture and shallower depth of pre-trained VGG models may be advantageous in reducing overfitting, which is particularly beneficial for smaller databases. In contrast, ResNet's depth enables it to learn intricate patterns. VGG and ResNet selection for smalle datasets hinges on factors such as computational resources, data complexity, and project-specific requirements. In the broader context of image classification, the choice between VGG and ResNet is emblematic of the ongoing evolution of deep learning, with each framework offering distinct advantages based on dataset characteristics and computational constraints.

Our approach involves fine-tuning CNN architectures tailored to the specific image classification objectives. To train our CNN models we employ standard training techniques [11], including data augmentation, normalization, a method for class imbalance [24], stochastic AdaGrad optimization and others. With a learning rate of 0.0001, batch size of 32, the models were trained using the default parameters. The process of choosing the specific CNN architecture (i.e., VGG or ResNet) is selected based on experimentation and validation, ensuring optimal performance in classifying gender fingerprint images. Given the imbalanced dataset evaluating the predictive model's performance poses challenges as traditional accuracy measures may be misleading, as they can mirror the inherent class distribution. Therefore, we opt for the F-score as our evaluation metric, which integrates precision and recall, offering a more robust performance measure for imbalanced datasets. The F-score provides a comprehensive perspective on accuracy, surpassing individual

metrics like precision or recall, enabling clearer understanding and communication of findings, irrespective of dataset class imbalances

Due to partially or low-quality fingerprint images commonly found at a crime scene. We are exploring the VGG network's performance in classifying gender based on fingerprint internal and external cylindrical regions. By examining the specific contributions of the internal and external areas (see Figure 1), we aim to gain a deeper understanding of the nature of fingerprint data and its relevance to gender classification tasks in case of partially or low-quality fingerprints. We evaluated four scenarios: gender classification based on (a) 50% of the inner regions, (b) 60% of the inner regions, (c) 50% of the outer regions, and (d) 40% of the outer regions. The training and testing sets were kept the same in each case.
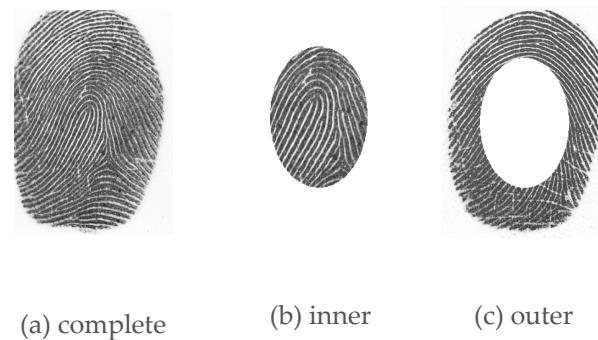

(a) complete   (b) inner   (c) outer

Figure 1. Cases of partially or low-quality fingerprints (a) complete fingerprint, (b) the inner cylindrical region, (c) the outer cylindrical region.

Regarding the DCAI, we employed a comprehensive set of five distinct approaches to identify and rectify issues within the most challenging 5% of data instances. Subsequently, we conducted model retraining using the refined dataset and evaluated the performance of each newly trained model. The five specific DCAI approaches we examined included [15]: Cleanlab Out of Distribution (cleanlab-OOD), FLIP (comprising Easy and Hard variants), and Margin Of Confidence (MOC) (Easy and Hard variants). The following lines will provide a detailed explanation of each of these approaches.

Out of Distribution (OOD): OOD refers to Identifying outliers in test data, such as data samples that do not stem from the distribution of the training data. To evaluate OOD we utilized the Cleanlab framework [15] , which leverages the "Principle of Counting." This principle uses the model's predicted probabilities (the confident joint) to estimate the number of examples in each class. By applying the Principle of Counting to these probabilities, we identified examples that fell outside the expected distribution, designating them as clearnlab Out of Distribution (cleanlab-OOD) instances.

Margin Of Confidence (MOC): MOC is a critical metric for assessing prediction accuracy. It quantifies the margin between two prediction scores. A larger MOC signifies more accurate predictions, while smaller margins imply potentially less reliable and inaccurate predictions. To measure MOC we employed a classification head with two outputs for gender prediction (male and female). This setup allowed us to assess the accuracy of our model's gender predictions based on the given datasets. We distinguished between MOC-Hard and MOC-Easy approaches. The MOC-Hard approach involved removing the top 5% of data instances where the model demonstrated elevated prediction confidence. This strategy allowed us to focus on more challenging data for model optimization. In contrast, the MOC- Easy approach targeted potential noise in the dataset by removing the 5% of data with the lowest prediction confidence scores, enhancing our focus on informative data.

FLIP (Classification Prediction Flips): FLIP measures the number of times a classification prediction changes during training. This metric assesses the robustness and generalizability of models. It provides insights into the effort required to achieve an acceptable level of accuracy on

unseen data. It also helps identify points at which further optimization efforts may not yield improved performance. It also helps identify points where such efforts should be halted due to diminishing model accuracy returns. Similar to MOC, we distinguished between FLIP-Hard and FLIP-Easy approaches. The FLIP-Hard approach involved removing 5% of data with consistent prediction scores, retaining more complex data elements. On the other hand, the FLIP-Easy technique discarded 5% of the most inconsistent prediction data, ensuring a focus on data where the model exhibited persistent prediction clarity.

Results

The results section is divided into three parts. In the first part we outline the process of selecting the optimal CNN for classifying gender using fingerprint images. This part also includes a thorough evaluation of the network's performance across all four databases, ensuring a comprehensive assessment. In the second part, the most influential fingerprint region that contributes significantly to gender classification was identified and evaluated. In the final part we explore several DCAI strategies aimed at enhancing gender classification accuracy. These strategies are designed to refine and optimize our gender classification models, leading to more robust and reliable classifications.

Part I -

Identification of the most suitable network for gender classification: To achieve this, the NITS-302 dataset was used as a basis for evaluating five variants of the two most widely employed CNN architectures (VGG and ResNet). NIST-302 was chosen for this evaluation due to its extensiveness and relatively high image quality, as well as its comprehensive coverage by including images of all 10 fingerprints. We used pre-trained ImageNet networks, at a learning rate of 0.0001, a batch size of 32, and an AdaGrad optimizer to train the selected models.

Table 2 shows ResNet18 exhibited overfitting issues and produced suboptimal results. Compared to ResNet, VGG16 and VGG19 achieved a test accuracy of 0.83, which is notably higher than ResNet's 0.75. It appears that VGG19 optimizes the trade-off between model complexity and generalization of unknown data; this is evident when focusing on the balance between train and test accuracy. Furthermore, the F-score, which represents a consolidated view of model performance based on precision and recall, encourages VGG models. Specifically, VGG19 achieved the highest test accuracy and F-scores of 0.84 and 0.81 respectively.

Table 2. Results of gender classification using fingerprint images employing the NIST-302 dataset across two popular CNN architectures (VGG and ResNet). VGG16, VGG19, ResNet18, ResNet50, ResNet101 were evaluated. The results are presented in terms of accuracy, precision, recall, and F-score.

Network Name	Train Accuracy	Test Accuracy	F-score	Precision	Recall
VGG16	0.92	0.83	0.8	0.8	0.81
VGG19	0.87	**0.84**	**0.81**	0.81	0.81
ResNet18	0.97	0.76	0.77	0.78	0.78
ResNet50	0.85	0.75	0.76	0.78	0.74
ResNet101	0.87	0.76	0.77	0.79	0.75

Next, the selected VGG19 model was evaluated across the four databases we used, one internal and three public.

As can be seen in Table 3, the VGG19 network can be generalized to other datasets, as the classification results range from 70 to 95% with no evidence of overfitting. The results are correlated with the size and quality of the images in the dataset, as indicated in Table 1. IsrPoliceDB yielded the highest classification accuracy rate of 96%, in accordance with the dataset having the largest image size and resolution. A moderately sized image and resolution in the NIST datasets (DB4, 302) achieved an accuracy rate of 80%, while a considerably smaller image achieved an accuracy rate of

68%. The results indicate that the VGG network is capable of capturing positive samples effectively, while minimizing false positives reliably and stable.

Table 3. Results of gender classification using the selected VGG19 model on the four different datasets: our internal IsrPoliceDB and the three public datasets - NIST-DB4, NIST-302 and SOCOfing.

Network Name	Train Accuracy	Test Accuracy	F-score	Precision	Recall	Train and Test size
IsrPoliceDB	0.97	0.96	0.96	0.96	0.96	1016 / 255
NIST-DB4	0.94	0.90	0.82	0.86	0.79	800 / 200
NIST-302	0.88	0.84	0.81	0.81	0.81	1583 / 396
SOCOfing	0.84	0.80	0.68	0.69	0.68	4800 / 1200

Part II-

Next, we will present the outcomes of the second part of our study. This centered on identifying and assessing the critical fingerprint image region essential for gender classification, especially in scenarios involving partially or low-quality fingerprint images. Our approach involves the identification and delineation of the ROI in fingerprint images, aiming to pinpoint the areas that significantly influence gender classification.

As can be seen in Figure 2, a 4% decrease is seen when classification is based on the inner region of the fingerprint. In contrast, when classifying based on the outer region, accuracy is usually not affected, with one exception, in the NIST-DB4, which showed a 1% drop/decrease; all other cases either improved or remained unchanged. The results indicate that the outer region of a fingerprint is more significant for gender classification, which is in accordance with previous literature findings. It is further supported by the absence of "delta" elements and additional features in the outer region that do not interfere with "ridge counting."
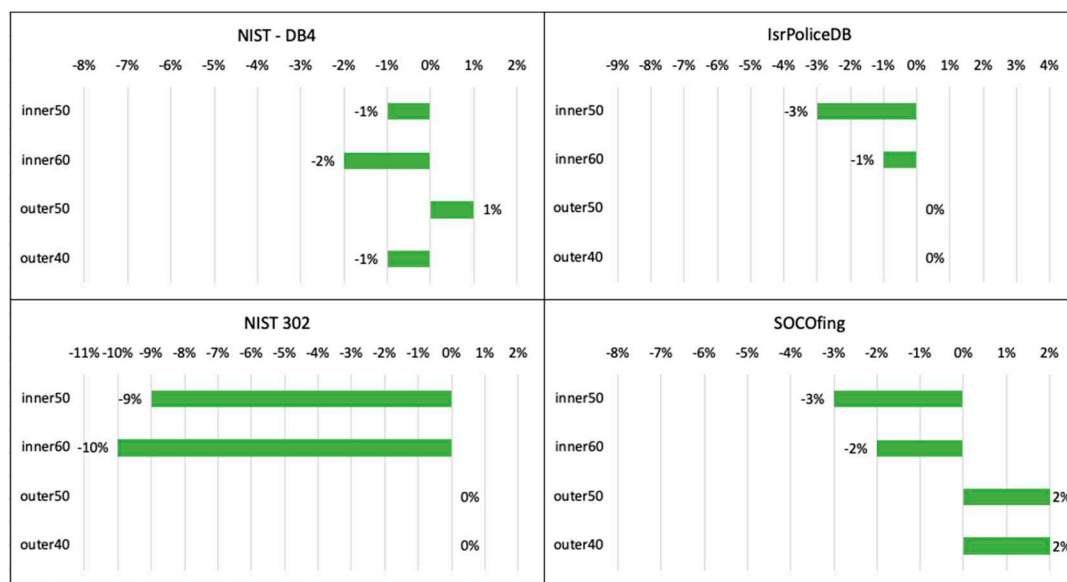

Figure 2. The percentage change in F-score compared to baseline results is shown in Table 3. For each database we examined four cases. Gender classification based on (a) 50% of the inner region, (b) 60% of the inner region, (c) 50% of the outer region, (d) 40% of the outer region.

Part III-

Lastly, we will examine the DCAI strategies to improve the results. As discussed in the methods section, DCAI is a paradigm shift in artificial intelligence that focuses on data as the primary source of information. We examined five DCAI approaches: cleanlab-OOD, FLIP (Easy and Hard), and MOC (Easy and Hard). For each of the five approaches, we scored every instance in the training set and

filtered out the highest 5% of instances in the dataset. Subsequently, we re-trained each model using the resulting clean dataset. The newly trained model was then evaluated on the original test-set.

As evident from Figure 3, Two out of the five approaches, cleanlab-OOD and FLIP-Easy, showed consistent improvements, with an average F-score improvement of 2.5% and 2.75%, respectively. The cleanlab-OOD approach improved the F-score across the four datasets, except for the NIST-302 dataset where the values were equal to the baseline. The most significant improvement was observed for the NIST-DB4 dataset, where the F-score increased from 0.82 to 0.87 (see supplemental table). FLIP-Easy approaches showed noticeable improvements on the NIST-DB4 and NIST-302 datasets. These findings provide compelling evidence of the successful filtering capability of DCAI approaches, effectively eliminating the data that negatively impacted the model's performance.
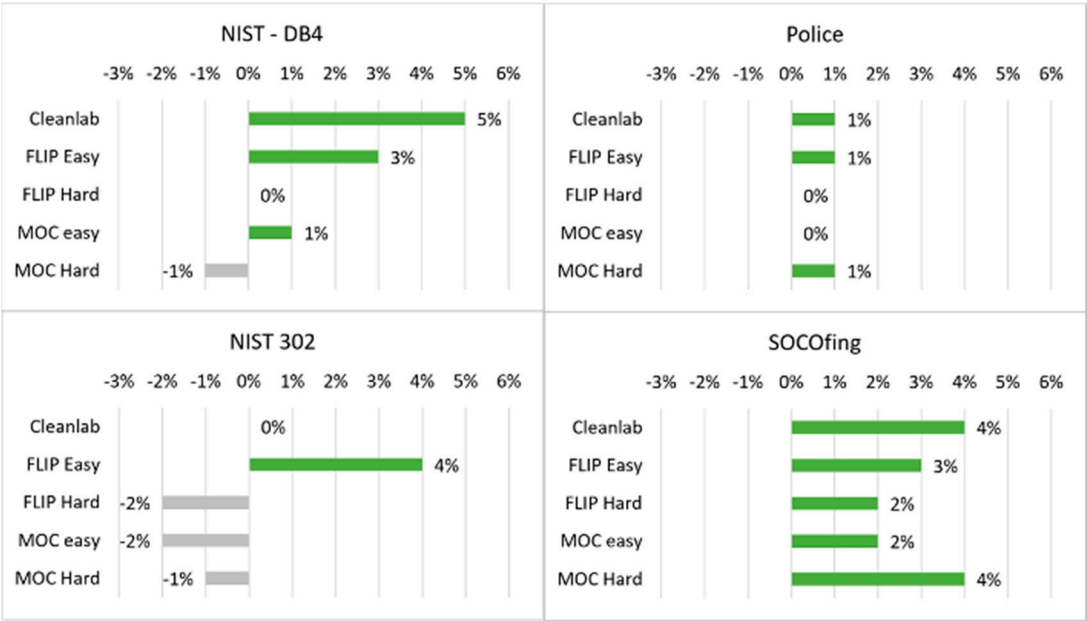

Figure 3. A percentage change in F-score compared to baseline results is shown in Table 3. For each database we examined five DCAI approaches: cleanlab-OOD, FLIP (Easy and Hard), MOC (Easy and Hard).

Discussion

This paper demonstrate a comprehensive evaluation of fingerprint images gender classification using CNN. The study is divided into three main parts: selection of an optimal CNN for gender classification based on the given datasets, identification of critical fingerprint regions for classification, and exploration of Data Cleaning and Augmentation for Improvement (DCAI) strategies to enhance classification accuracy. This discussion section delves into the implications of these findings and their significance for biometric gender classification.

The results indicate that VGG19 outperforms other architectures in terms of accuracy, precision, recall, and F-score. It achieved a test accuracy of 0.84, highlighting its ability to balance model complexity and generalization. These findings emphasize VGG19's superiority in accurately classifying gender based on fingerprint images. VGG19's successful application is further demonstrated when it is tested across different datasets, including the IsrPoliceDB and three public databases (Table 3). The classification results range from 70% to 95%, showing its ability to generalize across various datasets without overfitting. The results are correlated with the size and quality of the images in the dataset. This indicates that VGG19 effectively captures positive samples while minimizing false positives. In order to apply gender classification in a practical manner in a variety of real-world scenarios, this ability to generalize across diverse datasets is vital. The value of fingerprint evidence in detecting crime is worth consideration when analyzing smaller datasets. As

a result of their less complex architecture and shallower depth, pre-trained VGG models may perform better, mitigating overfitting, which is particularly advantageous for smaller databases.

The second part of the study focused on identifying and assessing the critical fingerprint region that significantly influences gender classification (Figure 2). It becomes evident that the outer region holds greater importance in gender classification. This conclusion is in harmony with existing literature and underpinned by the distinct absence of disruptive elements within the outer region. The methodology employed in this study to define the ROI)in fingerprint images yields valuable insights for researchers and practitioners dealing with partially or low-quality fingerprint images.

The third part of the study explored Data-Centric AI (DCAI) strategies aimed at further enhancing classification accuracy (Figure 3). Among the five DCAI approaches examined, cleanlab-OOD and FLIP-Easy consistently demonstrated improvement in F-scores, with an average increase of 2.5% and 2.75%, respectively. FLIP-Easy shoes the optimal approach for several reasons. First, by eliminating 5% of the data with the most inconsistent classifications during training, it ensures that the model focuses primarily on data where the classifications are stable. This enhances the model's generalization. By applying this strategy, overfitting is reduced, model reliability is improved, and independent data can be accurately predicted more accurately. Second, it helps in building a leaner, more resource-efficient model. Data points with fluctuating classifications can slow down the learning process by causing the model to repeatedly adjust its parameters to correct perceived errors. By removing these unstable elements, the model can train faster and use computational resources more efficiently. Lastly, the FLIP-Easy approach encourages a focus on high-quality data. In machine learning algorithms, the quality of the training data is crucial. By removing volatile data, this approach ensures that the model draws its lessons from the most consistent, reliable sources of data. This can result in improved overall performance. To conclude, FLIP-Easy provides the highest performance for constructing a more generalizable, efficient, and higher-performing model.

Conclusion

In conclusion, this study contributes to the field of gender classification using fingerprint images. The selection of VGG19 as the optimal CNN architecture showcases its superior performance, generalization ability, and practicality in real-world applications. The identification of critical fingerprint regions, particularly the outer region, provides valuable insights for improving gender classification in scenarios with partially or low-quality fingerprint images. Additionally, the successful implementation of DCAI approaches demonstrates their potential in enhancing classification accuracy. Overall, this study provides a robust foundation for advancing the accuracy and reliability of gender classification based on fingerprint images. This has implications for various security and forensic applications. Further research in this direction can include a broader analysis of real crime scene fingerprint images beyond the ink fingerprints used in this study. The limitations of this study include highlighting VGG19 as the optimal CNN architecture; its performance might be improved by comparing it with other state-of-the-art architectures. This could provide a more comprehensive understanding of the strengths and weaknesses of the chosen model. Further research could investigate this as well.

Reference

1. P. Schuch, "Survey on features for fingerprint indexing," *IET Biometrics*, vol. 8, no. 1. Institution of Engineering and Technology, pp. 1–13, Jan. 01, 2019. doi: 10.1049/iet-bmt.2017.0279.
2. D. Maltoni, D. Maio, A. K. Jain, and J. Feng, "Handbook of Fingerprint Recognition Third Edition."
3. A. Falohun, O. Fenwa, and F. Ajala, "A Fingerprint-based Age and Gender Detector System using Fingerprint Pattern Analysis," *Int J Comput Appl*, vol. 136, no. 4, pp. 43–48, Feb. 2016, doi: 10.5120/ijca2016908474.
4. M. A. Acree, "Is there a gender difference in fingerprint ridge density?," *Forensic Sci Int*, vol. 102, no. 1, pp. 35–44, May 1999, doi: 10.1016/S0379-0738(99)00037-7.
5. S. Sharma, R. Shrestha, K. Krishan, and T. Kanchan, "Sex estimation from fingerprint ridge density. A review of literature," *Acta Biomedica*, vol. 92, no. 5. Mattioli 1885, 2021. doi: 10.23750/abm.v92i5.11471.
6. P. Terhörst, N. Damer, A. Braun, and A. Kuijper, "What can a single minutia tell about gender?; What can a single minutia tell about gender?," 2018.

7. L. Berriche, "Comparative Study of Fingerprint-Based Gender Identification," *Genet Res (Camb)*, vol. 2022, 2022, doi: 10.1155/2022/1626953.

8. G. Jayakala, L. Sudha, and A. Professor, "Gender Classification Based on Fingerprint Analysis," 2021.

9. O. N. Iloanusi and U. C. Ejiogu, "Gender classification from fused multi-fingerprint types," *Information Security Journal*, vol. 29, no. 5, pp. 209–219, Sep. 2020, doi: 10.1080/19393555.2020.1741742.

10. J. W. Bond, "The value of fingerprint evidence in detecting crime," *International Journal of Police Science & Management*, vol. 11, no. 1, pp. 77–84, 2009.

11. I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016.

12. A. Narayanan and S. K, "Gender Detection and Classification from Fingerprints Using Pixel Count," *SSRN Electronic Journal*, 2019, doi: 10.2139/ssrn.3444032.

13. Y. I. Shehu, A. Ruiz-Garcia, V. Palade, and A. James, "Detailed Identification of Fingerprints Using Convolutional Neural Networks," in *2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA)*, IEEE, Dec. 2018, pp. 1161–1165. doi: 10.1109/ICMLA.2018.00187.

14. Y. Qi, M. Qiu, H. Jiang, and F. Wang, "Extracting Fingerprint Features Using Autoencoder Networks for Gender Classification," *Applied Sciences*, vol. 12, no. 19, p. 10152, Oct. 2022, doi: 10.3390/app121910152.

15. D. Zha *et al.*, "Data-centric Artificial Intelligence: A Survey," Mar. 2023.

16. J. Jakubik, M. Vössing, N. Kühl, J. Walk, and G. Satzger, "Data-Centric Artificial Intelligence," Dec. 2022.

17. C. I. Watson and C. L. Wilson, "NIST Special Database 4 Fingerprint Database," 1992.

18. Y. I. Shehu, A. Ruiz-Garcia, V. Palade, and A. James, "Sokoto Coventry Fingerprint Dataset," vol. 103, pp. 5–7, 2018, [Online]. Available: http://arxiv.org/abs/1807.10609

19. G. Fiumara, B. Woodgate, P. Flanagan, B. Woodgate, and M. Schwarz, "NIST Technical Note 2007 NIST Special Database 302," p. 15, 2019.

20. C. I. Watson and C. L. Wilson, "NIST Special Database 4 Fingerprint Database," 1992.

21. G. Fiumara, M. Schwarz, J. Heising, J. Peterson, P. Flanagan, and K. Marshall, "NIST Special Database 302: Supplemental Release of Latent Annotations," 2019, doi: 10.6028/NIST.TN.2190.

22. K. Simonyan and A. Zisserman, "VERY DEEP CONVOLUTIONAL NETWORKS FOR LARGE-SCALE IMAGE RECOGNITION," 2015, Accessed: Nov. 22, 2023. [Online]. Available: http://www.robots.ox.ac.uk/

23. K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition." Accessed: Dec. 18, 2020. [Online]. Available: http://image-net.org/challenges/LSVRC/2015/

24. J. M. Johnson and T. M. Khoshgoftaar, "Survey on deep learning with class imbalance," *J Big Data*, vol. 6, no. 1, p. 27, Dec. 2019, doi: 10.1186/s40537-019-0192-5.