

Testing the Performance of Face Recognition for People with Down Syndrome

C. Rathgeb¹, M. Ibsen¹, D. Hartmann¹, S. Hradetzky¹ and B. Ólafsdóttir²

¹Hochschule Darmstadt, Darmstadt, Germany

²Danmarks Tekniske Universitet, Lyngby, Denmark

Abstract—The fairness of biometric systems, in particular facial recognition, is often analysed for larger demographic groups, e.g. female vs. male or black vs. white. In contrast to this, minority groups are commonly ignored. This paper investigates the performance of facial recognition algorithms on individuals with Down syndrome, a common chromosomal abnormality that affects approximately one in 1,000 births per year. To do so, a database of 98 individuals with Down syndrome, each represented by at least five facial images, is semi-automatically collected from YouTube. Subsequently, two facial image quality assessment algorithms and five recognition algorithms are evaluated on the newly collected database and on the public facial image databases CelebA and FRGCv2. The results show that the quality scores of facial images for individuals with Down syndrome are comparable to those of individuals without Down syndrome captured under similar conditions. Furthermore, it is observed that face recognition performance decreases significantly for individuals with Down syndrome, which is largely attributed to the increased likelihood of false matches.

I. INTRODUCTION

Nowadays, systems incorporating face recognition technologies have become ubiquitous in personal, commercial, and governmental identity management applications. Recently, however, there has been a wave of public and academic concerns regarding the existence of systemic bias in face recognition. That is, algorithms have often been labelled as “racist” or “unfair” by the media, nongovernmental organizations, and researchers alike [9], [24]. While various methods have been proposed to improve the fairness of biometric technologies, these efforts have mainly been devoted to fairness across large demographic groups mostly related to gender and race [17], [2]. In contrast, only little research has been directed towards analysing the *inclusiveness* of biometric technologies for minorities. With the increasing popularity of biometric systems, there is a growing concern that these technologies might not be inclusive towards minority groups. Many of these concerns have in common that they are often linked directly to worries about exclusion and discrimination as a result of addressing the demands of a larger user group rather than including minority groups.

Like many other automated decision-making systems, face recognition technologies are heavily based on machine learning. These algorithms have been shown to work well on data which is similar to the data they have been trained on. However, minorities such as people with Down syndrome, are often underrepresented in the training data. With one



Fig. 1. Example image pairs of two subjects with Down syndrome. (images licensed under CC taken from [22])

in 1,000 births per year [20], Down syndrome is the most common chromosomal abnormality in humans. The cause of this syndrome is a trisomy of all or part of chromosome 21 in all or some cells of the body [21]. Certain clinical and phenotypic features are shared among affected individuals, such as congenital heart disease, cognitive impairment as well as facial appearance. Typical facial characteristics for people with Down syndrome include slanted eyes, brush marks in the iris, a round face, abnormal external ears, a flat bridge of the nose and a flattened nose, as well as a small chin [3], see Figure 1. While some research in the field of medical image analysis has been directed towards the detection of Down syndrome from face images [1], a comprehensive analysis of face recognition performance for people with Down syndrome has not yet been carried out. In response, this paper makes the following contributions:

- A collection of a database of 98 individuals with Down syndrome; in total, this database consists of 1,058 face images with at least five images per individual, which allows for a reasonable number of mated (i.e., genuine) and non-mated (i.e., impostor) comparisons.
- A comprehensive performance analysis of face image sample quality assessment and face recognition using state-of-the-art algorithms including two open source algorithms for face quality assessment and three open source as well as two commercial-of-the-shelf (COTS) algorithms for face recognition.
- A comparison of the results obtained on the collected Down syndrome database with baselines on two commonly used face image databases, together with a detailed analysis of the causes of the observed performance differentials.

This work is organised as follows: related works are briefly summarised in Sect. II. The conducted data collection and the experimental setup are described in Sect. III and Sect. IV, respectively. Results are presented and discussed in Sect. V.

Finally, conclusions are drawn in Sect. VI.

II. PREVIOUS WORKS

As mentioned before, some previous works have analysed the detection of Down syndrome from face images, i.e., a binary classification based on a single image. While earlier methods aim at classifying features obtained from basic textural [5] or geometrical analysis [6], more recent approaches are based on deep learning [11]. For a survey on proposed approaches the interested reader is referred to [1]. Databases employed in said works usually contain only a single image per subject and no identity labels are available. They only contain information whether the person is affected by Down syndrome or suffers from another gene mutation. In addition, most studies aim at an early diagnosis of Down syndrome and therefore used databases primarily contain images of children or toddlers and infants.

With respect to the inclusiveness of biometric systems for disabled individuals, the accessibility of various security methods within a mobile banking application, including face recognition, was examined in [4]. The experiment involved the participants with multiple disabilities. All participants struggled with presenting their faces in the way required by the face recognition enrolment. Furthermore, there was a noticeable difference in error rates in the tasks between people with disabilities and the control group, which consisted of non-disabled individuals. In [14], the experience of disabled people with biometric systems was analysed, including fingerprint, face, and voice recognition. More than a third of the participants stated that they had experienced face recognition to fail to recognise them or identified them as someone else. In response to such issues, the International Organization for Standardization (ISO) attempted to stipulate requirements for the inclusiveness of biometric systems [10]. In this guide, the most prevalent impairments have been listed, along with the problems these individuals might encounter and some suggestions on how these challenges might be resolved. Note that the aforementioned works mainly investigate the usability of face recognition for people with disabilities. To the best of the authors' knowledge, the performance of automated face recognition for people with disabilities (in particular Down syndrome) has not been investigated in previous works.

III. DATA COLLECTION

For the performance evaluation of face recognition algorithms, a suitable database containing face images of individuals with Down syndrome is required. Face images must fulfill the following requirements:

- The subject's face depicted in an image should be mostly frontal while common variations, i.e., slight variations in pose and facial expressions, are favoured and important to investigate recognition performance.
- For the aspect of investigating recognition performance, unique identity labels are necessary and must therefore either already exist or be created reliably.

- In order to investigate the recognition performance of single individuals, several images of a subject are needed.
- For the purposes of this work, young adults and adults are primarily of interest as the primary users of facial recognition applications.
- Images must have a minimum resolution of 128×128 pixels.
- A fair distribution between gender and race groups is desired.

As already mentioned in the previous section, the existing databases do not meet the listed requirements. This motivated the compilation of a new database according to the listed criteria. To this end, the online video sharing and social media platform YouTube was used, as it has been done in various other scientific works on face recognition, e.g., in [7]. The advantage of extracting face images from videos depicting individuals with Down syndrome is that faces are traceable throughout a video. Thus identity labels can be generated for individuals and it is additionally possible to obtain several images of the same person.

To start the actual recording of the images, a link collection of videos featuring people with Down syndrome was manually created. This was done by searching for people with Down syndrome using additional keywords such as "interview" or "report" to ensure that the videos found contained appropriate facial images. The resulting link collection will be publicly released¹. Using the Python package Selenium², the corresponding videos are automatically opened/started and screenshots are taken in five second intervals to ensure some variance. Then, images are stored in folders according to the video title to generate corresponding ID's in a later step. Before the data are processed further, dlib [16] is used for face detection to ensure that a face is visible. Subsequently, images that did not meet the aforementioned criteria are manually sorted out, including mainly images of individuals who are not affected by Down syndrome but are visible in the corresponding videos.

Overall, face images of 98 individuals with Down syndrome are extracted from 69 videos. The resulting database comprises a total number of 1,058 face images. It should be noted, that the number of images per person varies greatly: there are between five and 46 pictures per subject and on average about ten images.

IV. EXPERIMENTAL SETUP

The performance of commercial and open-source state-of-the-art facial analysis methods are evaluated on the collected database including face image quality assessment and recognition. Regarding quality estimation, the open-source deep learning-based algorithms FaceQnet [12] and MagFace [19] are used. In recognition experiments, AdaFace [15], ArcFace [8], MagFace [19] and two COTS systems (COTS-1

¹<https://dasec.h-da.de/research/biometrics/hda-down-syndrome-face-database/>

²www.selenium.dev/selenium/docs/api/py/api.html

and COTS-2) are employed. The use of the COTS algorithms raises the practical relevance of the presented benchmark. While the COTS systems are closed-source, it is assumed that these are based on deep learning, as this is the case for most of state-of-the-art methods.

Results obtained on the collected database are compared against results on two public databases: a subset of the FRGCv2 database [23], representing a constrained scenario, and the CelebA database [18], representing a less constrained scenario. To the best of the authors' knowledge the latter two databases do not contain images of people with Down syndrome.

Biometric recognition performance is measured based on metrics defined in [13]. Specifically, decision thresholds at false match rates (FMRs) of 1% and 0.1% are reported. Note that corresponding false non-match rates (FNMRs) are not reported, since the images of the collected database generally exhibit a rather low intra-class variation since these have been extracted from single videos (as will be shown in experiments). Therefore, any conclusions drawn from mated comparisons could be misleading. On the contrary, statistics of obtained quality and recognition scores are presented.

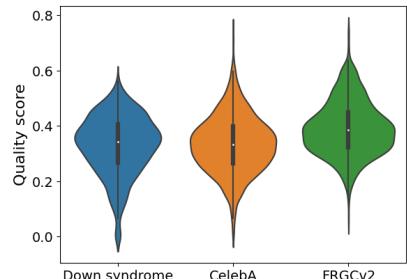
V. RESULTS

Average face image sample quality scores along with corresponding standard deviations for the different datasets are summarised in Table I (higher values indicate better sample quality). Corresponding quality score distributions are visualised in Figure 2. It can be observed that values obtained for the collected Down syndrome databases are similar to those obtained on CelebA and FRGCv2 for both of the used face image quality assessment algorithms. That is, on average the tested algorithms calculate similar quality scores for people with and without Down syndrome.

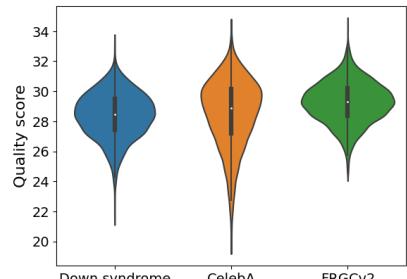
TABLE I
SAMPLE QUALITY STATISTICS (AVERAGE SCORE AND STD. DEV.)
ACROSS DATASETS.

Algorithm	Database	μ	σ
FaceQnet	Down syndrome	0.33	0.11
	CelebA	0.33	0.10
	FRGCv2	0.39	0.10
MagFace	Down syndrome	28.40	1.50
	CelebA	28.53	2.25
	FRGCv2	29.26	1.36

Regarding the recognition experiments, the thresholds required to achieve relevant FMRs of 0.1% and 1% are listed in Table II. It can be seen that for the collected Down syndrome dataset, the estimated thresholds are significantly higher compared to those of the CelebA and FRGCv2 datasets. This indicates that non-mated comparisons between subjects with Down syndrome tend to yield higher similarity scores for all of the employed face recognition systems, which is confirmed by the statistics shown in Table III. Further, non-mated score distributions of the Down syndrome dataset generally exhibit slightly higher standard deviations. A more



(a) FaceQnet



(b) MagFace

Fig. 2. Sample quality score distributions across datasets.

detailed comparison between non-mated score distributions obtained by different face recognition systems is plotted in Figure 3.

Note that mated scores calculated for the Down syndrome dataset are generally higher compared to the ones obtained on the CelebA and FRGCv2 datasets. As mentioned earlier, this is because mated image pairs of the Down syndrome dataset stem from frames of single videos. This means, with respect to the used datasets, the face images contained in the Down syndrome dataset have the lowest intra-class variations followed by the FRGCv2 and the CelebA datasets.

TABLE II
THRESHOLDS REQUIRED FOR RELEVANT FMRs ACROSS DATASETS.

System	Database	Threshold for	
		FMR = 0.1%	FMR = 1%
AdaFace	Down syndrome	0.48	0.40
	CelebA	0.20	0.14
	FRGCv2	0.25	0.17
ArcFace	Down syndrome	0.52	0.44
	CelebA	0.25	0.18
	FRGCv2	0.29	0.21
MagFace	Down syndrome	0.53	0.45
	CelebA	0.27	0.20
	FRGCv2	0.31	0.23
COTS-1	Down syndrome	0.81	0.70
	CelebA	0.34	0.20
	FRGCv2	0.43	0.27
COTS-2	Down syndrome	0.80	0.64
	CelebA	0.30	0.16
	FRGCv2	0.32	0.18

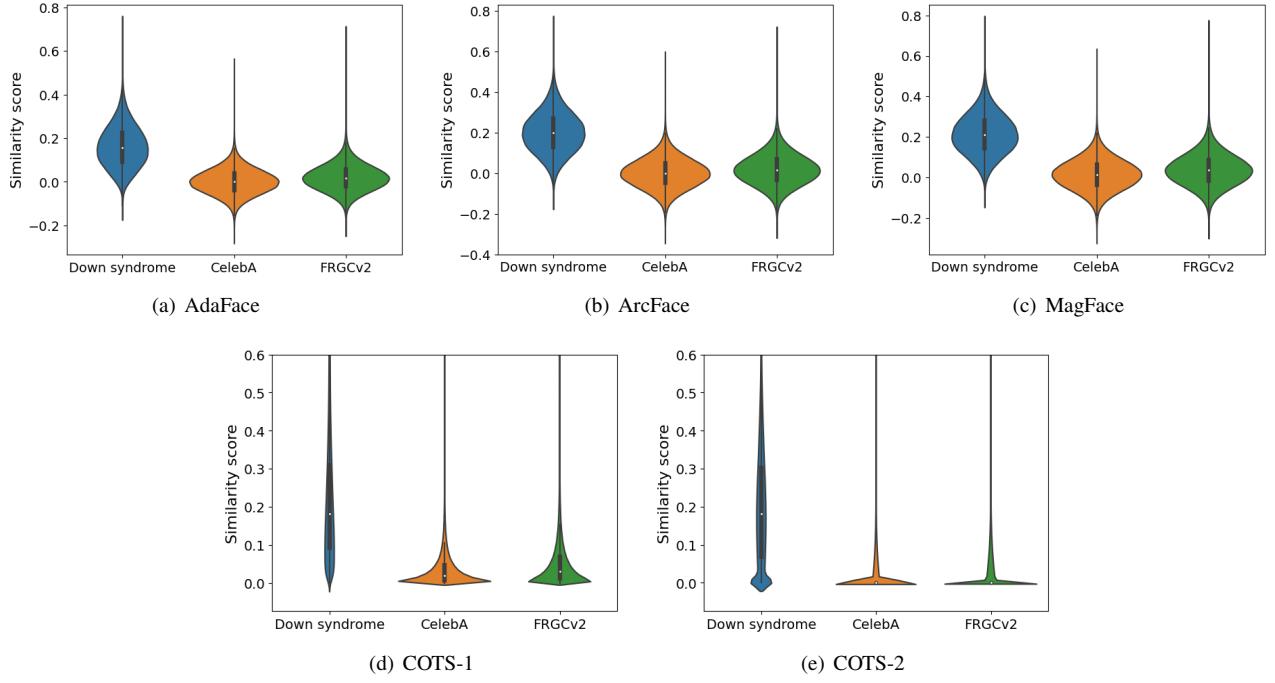


Fig. 3. Non-mated score distributions across datasets.

TABLE III

RECOGNITION STATISTICS FOR MATED AND NON-MATED DISTRIBUTIONS (AVERAGE SCORE AND STD. DEV.) ACROSS DATASETS.

System	Database	Mated		Non-mated	
		μ	σ	μ	σ
AdaFace	Down syndrome	0.84	0.08	0.16	0.10
	CelebA	0.61	0.18	0.00	0.06
	FRGCv2	0.71	0.07	0.02	0.06
ArcFace	Down syndrome	0.85	0.08	0.20	0.10
	CelebA	0.64	0.18	0.00	0.07
	FRGCv2	0.75	0.06	0.02	0.08
MagFace	Down syndrome	0.85	0.08	0.21	0.10
	CelebA	0.66	0.18	0.02	0.08
	FRGCv2	0.76	0.06	0.04	0.08
COTS-1	Down syndrome	0.98	0.02	0.22	0.16
	CelebA	0.87	0.21	0.03	0.04
	FRGCv2	0.96	0.01	0.05	0.06
COTS-2	Down syndrome	0.99	0.04	0.20	0.16
	CelebA	0.91	0.22	0.01	0.03
	FRGCv2	0.99	0.00	0.01	0.04

In summary, the following observations have been made for the tested algorithms:

- Face image sample quality scores for people with Down syndrome are in the same range as those obtained for subjects without Down syndrome. This means, with respect to quality assessment, no disadvantage is observable for people with Down syndrome.
- Albeit being extracted from single videos (which equals a single acquisition session), face images of the collected Down syndrome face database yield high similarity scores for mated comparisons. This indicates

that subjects with Down syndrome are not expected to experience higher rejection rates (in terms of FNMR) when using face recognition technologies.

- Similarity scores resulting from non-mated comparisons are significantly higher for people with Down syndrome. As mentioned earlier, subjects with Down syndrome share certain facial characteristics, which appear to cause this effect. As a result, the probability of false accepts (in terms of FMR) increases which impairs the security of face recognition for individuals with Down syndrome. In case decision thresholds are adjusted accordingly, higher false rejection rates are expected.

VI. CONCLUSION

With the evolution of biometrics, ethical concerns have emerged with respect to the fairness or inclusiveness of these technologies. While many research efforts have been devoted towards investigating race and gender biases, little to no efforts have been put into studying the inclusiveness of biometric systems for disabled individuals.

This work provided a preliminary study on the performance of facial recognition technologies for people with Down syndrome. While the results showed that face image quality assessment algorithms are applicable to this minority group, this was not the case for automated face recognition. Across five different face recognition algorithms it was also found that individuals with Down syndrome suffer from higher probabilities of false matches. To mitigate the observed bias, face images of subjects with Down syndrome should be included in training databases of face recognition technologies. In this context, the database collected in this work can serve as a starting point for future research.

VII. ACKNOWLEDGEMENTS

This research work has been partially funded by the Hessian Ministry of the Interior and Sport in the course of the Bio4ensics project, by German Federal Ministry of Education and Research and the Hessian Ministry of Higher Education, Research, Science and the Arts within their joint support of the National Research Center for Applied Cybersecurity, ATHENE.

REFERENCES

- [1] O. Agbolade, A. Nazri, R. Yaakob, A. A. Ghani, and Y. K. Cheah. Down syndrome face recognition: A review. *Symmetry*, 12(7), 2020.
- [2] V. Albiero, K. K.S., K. Vangara, K. Zhang, M. C. King, and K. W. Bowyer. Analysis of gender inequality in face recognition accuracy. In *IEEE/CVF Winter Conference on Applications of Computer Vision (WACV) Workshops*, 2020.
- [3] S. E. Antonarakis, R. Lyle, E. T. Dermitzakis, A. Reymond, and S. Deutsch. Chromosome 21 and down syndrome: From genomics to pathophysiology. *Nature Reviews Genetics*, 5:725—738, 2004.
- [4] R. Blanco-Gonzalo, C. Lunerti, R. Sanchez-Reillo, and R. Guest. Biometrics: Accessibility challenge or opportunity? *PloS one*, 13(3):e0194111, 2018.
- [5] K. Burçin and N. V. Vasif. Down syndrome recognition using local binary patterns and statistical evaluation of the system. *Expert Systems with Applications*, 38(7):8690–8695, 2011.
- [6] J. Y. R. Cornejo, H. Pedrini, and F. D. L. d. S. N. A. Machado-Lima. Down syndrome detection based on facial features using a geometric descriptor. *Journal of Medical Imaging*, 4, 2017.
- [7] A. Dantcheva, C. Chen, and A. Ross. Can facial cosmetics affect the matching accuracy of face recognition systems? In *International Conference on Biometrics: Theory, Applications and Systems (BTAS)*, pages 391–398, 2012.
- [8] J. Deng, J. Guo, J. Yang, N. Xue, I. Kotsia, and S. Zafeiriou. Arcface: Additive angular margin loss for deep face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(10):5962–5979, 2022.
- [9] P. Drozdowski, C. Rathgeb, A. Dantcheva, N. Damer, and C. Busch. Demographic bias in biometrics: A survey on an emerging challenge. *Trans. on Technology and Society (TTS)*, 1(2):89–103, 2020.
- [10] I. O. for Standardisation. Information technology — biometric performance testing and reporting) – Part 1: Principles and framework. Standard ISO/IEC 19795-1:2021(E), 2021.
- [11] Y. Gurovich, Y. Hanani, O. Bar, N. F. Guy Nadav, D. Gelbman, L. Basel-Salmon, P. M. Krawitz, S. B. Kamphausen, M. Zenker, and L. M. B. A. K. W. Gripp. *Nature Medicine*, page 60–64, 2019.
- [12] J. Hernandez-Ortega, J. Galbally, J. Fierrez, R. Haraksim, and L. Beslay. FaceQnet: Quality assessment for face recognition based on deep learning. In *International Conference on Biometrics (ICB)*, 2019.
- [13] ISO/IEC JTC1 SC37 Biometrics. *ISO/IEC 19795-1:2021. Information Technology – Biometric Performance Testing and Reporting – Part 1: Principles and Framework*, 2021.
- [14] S. K. Kane, A. Guo, and M. R. Morris. Sense and accessibility: Understanding people with physical disabilities' experiences with sensing systems. In *2nd International ACM SIGACCESS Conference on Computers and Accessibility*, pages 1–14, 2020.
- [15] M. Kim, A. K. Jain, and X. Liu. AdaFace: Quality adaptive margin for face recognition. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022.
- [16] D. King. Dlib-ml: A machine learning toolkit. *Journal of Machine Learning Research*, 2009.
- [17] K. S. Krishnapriya, V. Albiero, K. Vangara, M. C. King, and K. W. Bowyer. Issues related to face recognition accuracy varying based on race and skin tone. *IEEE Transactions on Technology and Society*, 1(1):8–20, 2020.
- [18] Z. Liu, P. Luo, X. Wang, and X. Tang. Deep learning face attributes in the wild. In *Proceedings of International Conference on Computer Vision (ICCV)*, 2015.
- [19] Q. Meng, S. Zhao, Z. Huang, and F. Zhou. MagFace: A universal representation for face recognition and quality assessment. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021.
- [20] J. K. Morris, D. E. Mutton, and E. Alberman. Revised estimates of the maternal age specific live birth prevalence of Down's syndrome. *Journal of Medical Screening*, 9(1):2–6, 2020.
- [21] D. Patterson. Molecular genetic analysis of Down syndrome. *Human Genetics*, 126(1):195–214, 2009.
- [22] Pexels Free Stock Photos. <https://www.pexels.com>, 2024.
- [23] J. Phillips, P. Flynn, T. Scruggs, K. Bowyer, J. Chang, et al. Overview of the Face Recognition Grand Challenge. In *Conf. on Computer Vision and Pattern Recognition (CVPR)*, pages 947–954. IEEE, June 2005.
- [24] C. Rathgeb, P. Drozdowski, D. C. Frings, N. Damer, and C. Busch. Demographic fairness in biometric systems: What do the experts say? *IEEE Technology and Society Magazine*, 41:71–82, 2022.