De-anonymizing Facial Recognition Embeddings

István Fábián[1] and Gábor György Gulyás[2]

***Abstract*—Advances of machine learning and hardware getting cheaper resulted in smart cameras equipped with facial recognition becoming unprecedentedly widespread worldwide. Undeniably, this has a great potential for a wide spectrum of uses, it also bears novel risks. In our work, we consider a specific related risk, one related to face embeddings, which are machine learning created metric values describing the face of a person. While embeddings seems arbitrary numbers to the naked eye and are hard to interpret for humans, we argue that some basic demographic attributes can be estimated from them and these values can be then used to look up the original person on social networking sites. We propose an approach for creating synthetic, life-like datasets consisting of embeddings and demographic data of several people. We show over these ground truth datasets that the aforementioned re-identifications attacks do not require expert skills in machine learning in order to be executed. In our experiments, we find that even with simple machine learning models the proportion of successfully re-identified people vary between 6.04% and 28.90%, depending on the population size of the simulation.**

***Index Terms*—facial recognition, de-anonymization, machine learning**

I. INTRODUCTION

We live in times when efficient uses of artificial intelligence and cheap smart technology are exploding. By the spread of smart cameras, applications on facial recognition had become almost ubiquitous in some cities around the world. In some cases we can find the driver reason for this in the security concerns of the public, but face recognition (or FR in short) can be applied to a much broader set of use-cases. Beside identification or authentication of individuals in crowds, it could benefit the society also in criminal detection, searching for lost people, customer behavior analysis, etc. [1].

However, FR technology could be abused and therefore it has the potential to pose risks to individuals, to the society and even to the governmental and business sectors, as well [2]. This puts related ethical issues into the focus. The French data protection authority, the CNIL (French National Commission on Informatics and Liberty) published a recent paper detailing the technical, legal and ethical challenges regarding these applications [3]. The biggest concern probably is how FR is being a part of emerging surveillance technologies [4]. Consequently, several governments made recent attempts in order to regulate the uses of FR technology.

Despite official guidelines for camera surveillance [5], some believe that automated FR breaches GDPR because it fails to meet the requirement for consent by design [6]. The European Commission even considered imposing a temporary ban on using FR in public spaces, which was later discarded [7]. In their white paper released on the 19th February [8], the European Commission rather envisions an approach where companies evaluate their own data processing practices from a risk-based point of view. This is backed up by a recent proposal to conduct an impact assessment analysis when dealing with FR applications [2].

This debate on the ban is also present in the US. While Washington DC just passed facial recognition rules that allow the use of the technology with some restrictions (e.g. government agencies can only use FR software if it's got an application programming interface, and vendors must reveal any reports of bias) [9], San Francisco was the first city to ban FR entirely in public spaces [10]. The unresolved nature of these issues is further confirmed by the Fundamental Rights Agency, who released a paper about the fundamental rights considerations regarding FR [11].

Certain related risks can be associated with the processing and storing of facial imprints. State-of-the-art face imprints are coming from the domain of Deep Metric Learning (DML), in which deep learning techniques are trained to produce descriptive vectors of faces while also considering their similarity [12]. These vectors, or face embeddings, have high similarity when taken from the same person, but have a low similarity score when taken from different people. While these seem as a list of arbitrary numbers to the naked eye, they may contain personal information about the person whose photo was taken. In their recent work, *Mai et al.* showed that the photo itself can be reconstructible from the embedding [13]. In [14] authors argue that it should be an accepted fact that with good accuracy the original sample can be reconstructed from unprotected embeddings. This means that sensitive data could be derived from unprotected templates and other attacks can also be launched based on the reconstruction results. Based on this, it can also be possible to reverse engineer data from face embeddings in order to find out the original identity of the embedding.

In this paper we examine an attack that aims to find out the original identity of face imprints. As the original faces can be partially rebuilt from embeddings, we look at the scenario where the attacker tries to reconstruct demographic data from the embeddings. First, we measure the level of accuracy achievable in predicting age, sex and race from facial embeddings, then we create a synthetic dataset and run the attack from one end to the other. Our results show that predicting these characteristics is indeed possible with alarming accuracy and re-identification attacks can be executed successfully.

The paper is structured as follows. In Section II we discuss how facial recognition works, the privacy risks of processing face embeddings and how re-identification attacks work. Next,

[1] Balatonfüred Student Research Group
[1,2] Department of Automation and Applied Informatics, Budapest University of Technology and Economics, Hungary.
(e-mail: fabian@aut.bme.hu; gabor.gulyas@aut.bme.hu)

DOI: 10.36244/ICJ.2020.2.7

in Section III, we introduce our attacker model. In Section IV we describe how we used different technologies in our research, and following in Sections V-VI we elaborate our results. Finally, Section VII summarizes our work.

II. RELATED WORK

A. Facial Recognition

The main motivation behind facial recognition is to make it possible to identify people, e.g. a person from a digital photo or video frame based on the face's unique characteristics. Despite the fact that it has only become widespread in recent years, the technology has been around for decades, although it wasn't as extensively used as today, because it had many open problems that hindered its performance and accuracy, like the lack of enough computational power and training data, which resulted in poor scalability.

However, the first milestone towards automated FR came in 1988 when Sirovich and Kirby came up with the Eigenface approach [15], which applies linear algebra (including principal component analysis) to recognize faces. Basically, it works by creating an average face and multiple so called Eigenfaces based on all faces available in a dataset, and then representing each new face as a vector made up of the coefficients of the linear combination of the average face and the Eigenfaces. Then the similarity between two faces depends on the distance metric between each face's vector, with a small distance corresponding to higher similarity. In 1991, Turk and Pentland further improved the Eigenface approach to also detect faces in images [16]. Since then, it was in the 2010s when FR technology significantly improved due to the usage of machine learning and deep neural networks. This was made possible by the large amount of training data and computing power available.

In our analysis we wanted to work with state-of-the-art facial recognition techniques that are publicly available in Python libraries and that could be run efficiently on a typical smart camera. One of the leading solutions is found in the dlib library [17], which uses the ResNet-34 structure deep neural network from [18], trained on the Labeled Faces in the Wild dataset (LFW) [19]. Another prominent method is implemented in the OpenCV library. This deep convolutional network uses the FaceNet structure [20] that directly maps face images into the Euclidean space using a triplet-based loss function based on large margin nearest neighbor classification (LMNN) [21]. This library achieves a 99.63% accuracy score on the LFW dataset [19].

Both of these techniques produce a 128 long vector of float values. When comparing the two methods, we found that the technique offered by dlib provides a better trade-off regarding less false positives, with a slightly higher rate of false negatives. Therefore we decided to work with it throughout our experiments.

B. Risks Related to Embeddings

Face embeddings should be considered biometric data by definition provided by the General Data Protection Regulation (Art 4. §14 in [22]): an embedding consists of data points that were extracted from the photo of a person that allow or enable the identification of the data subject. Due to their nature, biometric attributes capture features of the human body that one cannot be changed. Therefore, significant societal and privacy risks arise, which urges the need to analyze the impacts of this technology [2]. As we discussed previously, modern FR works by extracting templates from photos that need to be stored in a database or compared previously stored ones. If we consider the number of people represented in the images X, and the number of people who are part of a database Y, then FR can be used for authentication (X:1 Y:1), identification (X:1 Y:n) or tracking (X:1 Y: no need for a database). Depending on these various use cases, the risks can be more or less severe, e.g., a big central database means higher risks against malicious actors than a smaller database.

Further reasons for concern are that FR is not a perfect technology, risk appear that had been seen previously in automated decision making systems [23]. For example, FR can be discriminatory due to biases built into the technology, or one may find it difficult to explain in details how DML-based facial recognition works or why it had proposed a specific embedding in a certain situation.

Authors in [24] mention two potential threats regarding an attacker's abilities. One of the hazards is to masquerade the template owner, which means using the biometric template for reconstructing a 2D or 3D model of the template owner's face and using that model to trick a FR system. The other is the possibility of the attacker to do cross matching between multiple databases storing biometric templates, because biometrics are mostly immutable and the same or very similar templates could be stored in multiple databases for different applications. These risks motivate the use of biometric template protection (BTP) schemes that transform biometric templates to make their usage and storage safe, while also keeping their utility.

III. RISK AND ATTACKER MODEL

In our work, we consider re-identification attacks against a database of face embeddings. Since face embeddings are based on the face's unique characteristics and enable reconstructing faces, they may contain hints for demographic information as well. This can contribute to identification attacks.

Re-identification attacks are when an attacker combines multiple data sources to uncover the identities in the anonymous dataset. A common example is a health care provider who publishes data for research purposes after removing any PII (personally identifiable information) such as names, addresses, social security numbers, etc. However, as [25] showed, it can still be possible to re-identify people in that database by linking it with an additional database (e.g. publicly available voter database). Demographic data can be especially vulnerable against re-identification attacks, as [25] showed that the zip code, sex and date of birth provides a unique identifier for 87% of the US population based on census data.

These examples showed that tabular datasets are vulnerable for re-identification. It has been shown that large datasets, where the number of attributes is rather proportional to the

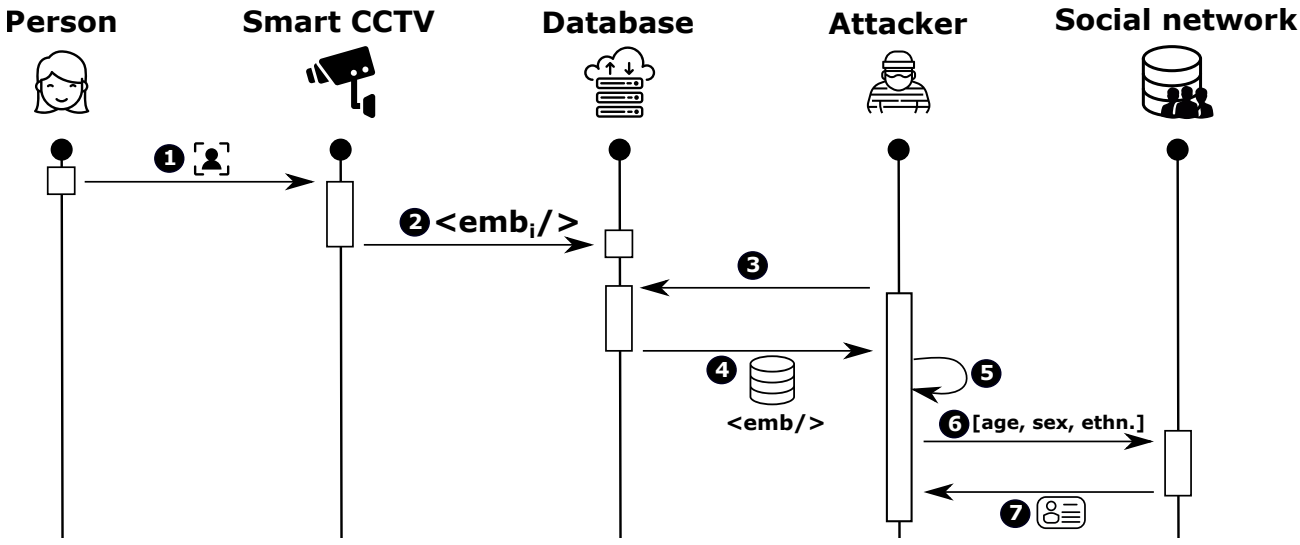

Fig. 1. The considered attack when a malicious third party reconstructs demographic data from embeddings and re-identifies data subjects by linking with another public database.

number of rows, can also be re-identified. Various examples include movie ratings [26], social networks [27], and credit card usage patterns [28]. As explained later, here we consider rebuilding attributes from embeddings that we consider later for re-identification.

In our case, let us consider the following FR system setup that may be deployed at a company, and the corresponding attacker model (see Figure 1). Smart cameras observe the company's various areas and extract the face embedding of employees appearing in the video footage (Step 1). These embeddings are then transferred and stored in a central database for later use either for tracking, automation, identification or other purposes (Step 2). The attacker then accesses these embeddings (Steps 3-4, e.g. en employee by stealing or an external person via hacking) and infers the data subjects' demographic information (age, sex and race) from them using a computer algorithm created for this task (Step 5). With this new information the attacker may now be able to do a successful re-identification attack by comparing the original data with another public data source, for example by looking up people on a social networking site (Steps 6-7).

The success of such an attack largely depends on Step 4 and Step 5 from Figure 1: how many embeddings the attacker can get, and how accurately they can predict demographic information from those embeddings. Thus, it is necessary to assess the potential attacker strength first. In our work, we assume a strong attacker who has access to all the embeddings stored in the database, and our main goal is to discover the level of prediction accuracy achievable regarding demographic data.

IV. METHODOLOGY

In order to estimate the potential success of attackers, on a real life dataset we considered the equivalence class distribution of demographic details. An equivalence class is a subset of elements that are equivalent to each other based on the demographic characteristics that we are trying to predict. In a database, the more people that are either unique or fall in small equivalence classes (e.g. at most 5 members), the higher

A. Technical Details

We carried out our analysis in the Python programming language, using open source libraries created for working on data science and machine learning (ML) applications (NumPy [29], pandas [30], Scikit-learn [31]). The face recognition library we used was face_recognition [32], which is a wrapper built around dlib [17] and uses dlib's state-of-the-art FR technology based on deep learning to detect faces in images and/or video frames and extract the face embeddings from them. While embeddings are hard for a human to interpret, a computer can compare two embeddings and calculate the mathematical distance between them, such as Euclidean or Manhattan distance, with the Euclidean distance being the most popular "best practice" choice for face recognition applications. These metrics can be used to determine whether the two embeddings belong to the same person or not. The lower the distance between two embeddings, the more likely it is that they belong to the same person. Usually, there is a distance threshold below which we consider embeddings to belong to the same person.

We used Random Forest Classifiers from the Scikit-learn library to build three ML models for predicting the age, sex and race from the embeddings. We chose a Random Forest Classifier as it is an easy to use ML model that doesn't require hyper parameter tuning and can be used easily even by non ML experts. It is an ensemble-tree based learning algorithm used to predict the class of test objects. Instead of training a single decision tree on the entire training data, the random forest works by training multiple decision trees on randomly sampled subsets of the training set (while also having the attributes randomly distributed), and then aggregating the votes of the

decision trees to conclude the final predicted class by majority voting.

For the data to train and test on, we used UTKFace [33], a public database containing over 23,000 photos from both sexes aged between 1 to over 100, from white, black, asian, indian and other races, where one image per person is included. Due to the fact that the various age, sex and race classes were not balanced, we sampled this data source to gain a more balanced dataset for training and testing (see the following subsection).

B. Our Methodology

Since the biggest majority of the people in UTKFace database are under the age of 80 and are either white, black, asian or indian, we only considered people fitting these constraints. There was a very low number of examples in dropped classes which would have led to poor training and prediction results. However, not all of the remaining classes were balanced. For example there were 2043 photos of white males aged between 20 and 40 years, while only 677 Asian males in the same age range.

So to achieve a relatively balanced training and testing data set, we had to apply data down sampling until we were left with 12192 photos, 1524 photos for each of the 8 race-sex pairs. Yet, the age distribution still was not completely balanced, as there were 2893 people (23.73%) aged between 1 and 20 years, 5515 (45.23%) aged between 21 and 40 years, 2452 (20.11%) aged between 41 and 60 years, while only 1332 people (10.93%) were aged between 61 and 80 years. While we accept this as it is rather life-like, this could hinder model performance. Furthermore, achieving a completely balanced dataset would have resulted in too few examples to train and test with.

The following step is to run the face_recognition library's face_encodings function on all the 12192 images, and storing the face embedding found for each. Since the image file names contain the necessary information about a person's demographics (as all the image file names follow the [age]_[gender]_[race]_[date&time].jpg pattern), the file names were used to create the training labels for each image. Equipped with this labeled data set, it is now possible to use Scikit-learn's RandomForestClassifier class to train a Random Forest Classifier for predicting the age, sex and race from face embeddings. In all models, we found that using a Random Forest of 100 trees can achieve the job (i.e. setting the *n_estimators* parameter to 100). Also, using Scikit-learn train_test_split function to split the data set into 80% training and 20% testing data made it possible to validate our models.

The simplest Random Forest Classifier to train was the one predicting the sex of people based on their face embeddings as this required only binary classification, while predicting the age and race required multi-class classification. Regarding age prediction, expecting the prediction of precise age values resulted in poor performance. First this may sound surprising, but it is impossible even for humans to predict a person's age with such precision. Thus some intervals needed to be defined for age prediction. Choosing narrow age ranges (1-10 years) also resulted in poor prediction accuracy. On the other hand, choosing a too wide age range (25 years and over) would have resulted in very poor utility regarding inference. As a viable trade-off, we divided people into 4 age groups: 1-20, 21-40, 41-60 and 61-80 years.

The results of our experiment are detailed in the following section.

V. MEASUREMENTS

As seen in Table I, which represents the sex prediction model's confusion matrix on the test data, the model achieved an accuracy score of 91.8%, and an F1 score of 91.8%. Looking at the confusion matrix it can be concluded that even such a simple model can correctly recognize with closely the same accuracy both males and females. Figure 2 shows the receiver operating characteristic (ROC) curve which achieved an area under curve (AUC) value of 97.6%.

Table II shows the confusion matrix of the age prediction model's performance on the test data. It can be seen that the age prediction model achieved an overall accuracy score of 77% and a weighted F1 score of 76.3%. As expected, this model's scores are moderately lower, because predicting a class that can be anywhere from 1 to 80 is a more complex problem than predicting sex, which is a simple binary classification. Also, the confusion matrix itself explains the lower scores as compared to the sex prediction: as discussed in the previous chapter, the data set was not completely balanced through all classes, so the ratio of people aged between 21-40 years was disproportionately high compared to other age groups. Summing up the values across the Truth rows, 23.65% of the people in the test data were aged between 1-20, 44.9% were between 21-40, 20.49% were between 41-60 and only 10.96% were between 61-80 year old. As a result, the model is better at predicting younger people's age, and it fails more often at predicting older ages. Moreover, possibly due to the fact that almost half the people in the dataset were between 21-40 years of age, the model often makes the mistake of predicting this age group even for 1-21 and 41-60 year age ranges, too.

Finally, Table III shows the confusion matrix regarding the race prediction model's performance on the test data.

The model achieved an accuracy score of 83.4%, and a weighted F1 score of 88.9%). Based on this, we can conclude that all the models achieve a considerable accuracy in the predictions. An interesting pattern to note is that the model makes more errors with people in the white race: the most common mistake the model makes is predicting indian, asian and black people to be white.

Summing up the results we can see that sex prediction works the best with 91.8% accuracy, better than the race prediction model's 83.4% accuracy which outperforms the age prediction model's 77% accuracy. While the age prediction model is not as good as the other two models, it still reaches an accuracy that can be dangerous from a privacy standpoint. However, the main takeaway is that the three demographics attributes can be used to re-identify people from face embeddings.

TABLE I
CONFUSION MATRIX OF THE SEX PREDICTION MODEL (ACCURACY=91.8%, RECALL=92.8%, PRECISION=90.9%)

Truth		Predicted Sex: *Male*	Predicted Sex: *Female*
	Male	45.54%	4.65%
	Female	3.59%	46.22%

TABLE II
CONFUSION MATRIX OF THE AGE PREDICTION MODEL (ACCURACY=77%, RECALL=77%, PRECISION=77.8%)

Truth		Predicted Age: *1-20*	Predicted Age: *21-40*	Predicted Age: *41-60*	Predicted Age: *61-80*
	1-20	17.63%	5.89%	0.04%	0.09%
	21-40	0.30%	42.17%	2.26%	0.17%
	41-60	0.04%	6.96%	11.44%	2.05%
	61-80	0.04%	1.02%	4.18%	5.72%

VI. EMBEDDING RE-IDENTIFICATION BY PREDICTING DEMOGRAPHICS

With the three Random Forest Classifier models trained, we were equipped to simulate a re-identification attack using face embeddings against a synthetic database. We carried out attack simulations against databases sizes of 10, 50 and 100 people, which are plausible database sizes for small or medium sized companies.

To construct the synthetic databases with realistic demographics data, we relied on census data from the University of California's Adult Data Set for Machine Learning [34]. This dataset contains over 30,000 records of different types of people including their demographic data (age, sex and race) and the ratio of people believed to be represented by every record. We used the latter weights to sample this dataset to build the smaller databases of 10, 50 and 100. For every person in each database, we then associated photos from the UTKFace dataset [33] that matched their age, race and sex, and used [32] to extract the corresponding facial embeddings from these photos, while taking care not to ever re-use photos that were part of the training data set. In order to suppress any potential bias coming from the randomness, we repeated each experiment with a new synthesized dataset 50 times.

Next, we used our models to predict the sex, age (in 20 year ranges) and race from each embedding, and tried to match the prediction results to people in the original database. By comparing matched records to their corresponding ones in the original database (the ground truth), we could find out how many people's demographic information were correctly

TABLE III
CONFUSION MATRIX OF THE RACE PREDICTION MODEL (ACCURACY=83.4%, RECALL=83.4%, PRECISION=95.2%)

Truth		Predicted Race: *White*	Predicted Race: *Black*	Predicted Race: *Asian*	Predicted Race: *Indian*
	White	24.46%	0.47%	0.13%	0.60%
	Black	3.03%	21.55%	0.04%	0.55%
	Asian	2.60%	0.17%	22.28%	0.17%
	Indian	4.61%	0.43%	0.38%	18.52%

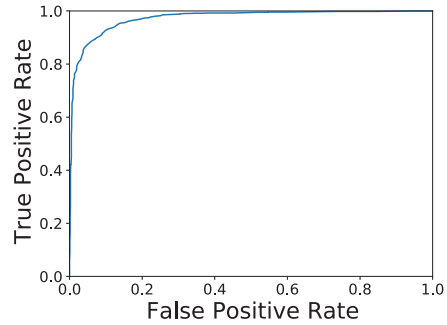

Fig. 2. The ROC curve for the sex prediction model (AUC=97.6%)

how many people's demographic information were correctly predicted. Also, as explained in Section IV, the smaller the size of a person's equivalence class is, the higher their risk of re-identification. So to consider the risks involved with these attacks, we measured the ratio of people falling in equivalence classes of different sizes (1, 2-5, 6-10, 11-20 and 20+). As stated above, we repeated this process 50 times for each smaller database to get an averaged out result.

Figure 3 shows our findings regarding equivalence class sizes. The most successful attacks can be carried out against the smallest database of 10 people, where 16% of all records fall in a unique equivalence class and are thus re-identified, and an additional 33.4% of records fall in an equivalence class of size 2-5, which still means considerable privacy risks. The risks are present even in the case of the databases of size 50 and 100, where the ratio of people falling in a unique equivalence class is 2.36% and 0.98% respectively, and the ratio of people falling in an equivalence class of size 2-5 is 12.72% and 7.18% respectively.

There is a considerable risk of re-identification for many people in all three database sizes simulated. If someone was unique, then we considered that as a successful re-identification. For the rest, the success of re-identification is proportional to the equivalence class size. We used the following metric to determine the overall risk of re-identification in each database size. If we consider the size of an equivalence class to be k, and the percentage of people that fall in that equivalence class based on the prediction is P, then the re-identification risk of that equivalence class is P/k. To get the expected proportion of people re-identified, one has to sum these values for all equivalence classes. In our experiments, these values were 28.90% for the database of 10, 10.38% for the database of 50, and 6.04% for the database of 100 people.

In conclusion, these results show that carrying out re-identification attacks by using face embeddings is indeed possible. Although as there are more people in the database, success of the attack degrades, chances of re-identification are never negligible.

VII. CONCLUSION

In this paper we discussed potential privacy and security risks associated with the widespread usage of facial recognition technologies, in particular the risks associated with pro-

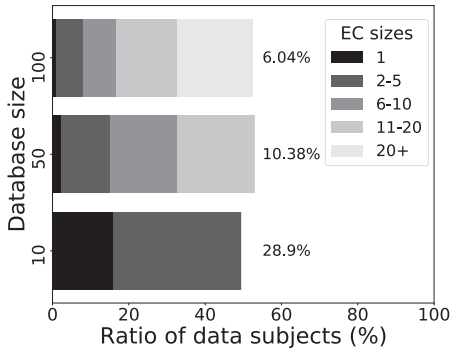

Fig. 3. The ratio of equivalence classes (EC) in the predicted databases (D) for various database sizes. Values in parentheses show the expected proportion of re-identified users.

cessing the concerned biometric identifiers. More specifically, we focused on attack that aim to re-identify facial embeddings based on using face embeddings to find out three key pieces of demographics data about the data subjects.

Our goal was to examine the level of accuracy achievable in predicting the sex, age and race from a face embedding. We used a publicly available facial database labeled with these demographic attributes to build a labeled training and testing dataset, and we trained a Random Forest Classifier to predict the sex, age and race from the embeddings.

Based on our findings, it is indeed possible to correctly predict someone's sex, age (within a 20 year range) and race from a face embedding with high accuracies: our models achieved a 90.9% accuracy score on sex prediction, a 83.4% accuracy score on race prediction and a 77% accuracy score on age prediction. As a result, we can consider our theory proven and state that the storing and processing of unprotected face embeddings pose considerable privacy risks as far as re-identification attacks and sensitive data leakage are concerned.

As the final conclusion, we state that further research is necessary to come up with privacy preserving ways to protect embeddings. One idea is to modify the face embeddings in such a way as to keep their utility (e.g. embeddings of the same person should remain close to each other in the vector space after the modification) while protecting them against reverse engineering attacks to make inference more difficult.

ACKNOWLEDGMENT

The research has been supported by the European Union, co-financed by the European Social Fund (EFOP-3.6.2-16-2017-00013, Thematic Fundamental Research Collaborations Grounding Innovation in Informatics and Infocommunications).

Project no. FIEK_16-1-2016-0007 has been implemented with the support provided from the National Research, Development and Innovation Fund of Hungary, financed under the Centre for Higher Education and Industrial Cooperation - Research infrastructure development (FIEK_16) funding scheme.

Icons made by Pixel perfect, fjstudio, Freepik, Pause08, surang, Smashicons from www.flaticon.com.

REFERENCES

[1] L. Introna and H. Nissenbaum, "Facial recognition technology: a survey of policy and implementation issues," 2010.

[2] C. Castelluccia and D. Le Métayer Inria, "Impact analysis of facial recognition," Feb. 2020, working paper or preprint.

[3] "Facial recognition: for a debate living up to the challenges," 2019.

[4] J. Goldenfein, "Facial recognition is only the beginning," 2020.

[5] E. . E. D. P. Board, "Guidelines 3/2019 on processing of personal data through video devices," 2019.

[6] T. Macaulay, "Automated facial recognition breaches gdpr, says eu digital chief," 2020.

[7] S. Stolton, "Leak: Commission considers facial recognition ban in ai 'white paper'," 2020.

[8] E. Commission, "White paper on artificial intelligence: a european approach to excellence and trust," Tech. Rep., 02 2020.

[9] T. Macaulay, "Washington state passes microsoft-approved facial recognition laws," 2020.

[10] D. Lee, "San francisco is first us city to ban facial recognition," 2019.

[11] FRA, "Facial recognition technology: fundamental rights considerations in the context of law enforcement," 2019.

[12] M. Kaya and H. Bilge, "Deep metric learning: A survey," *Symmetry*, vol. 11, p. 1066, 08 2019. [Online]. Available: DOI: 10.3390/sym11091066

[13] G. Mai, K. Cao, P. C. Yuen, and A. K. Jain, "On the reconstruction of face images from deep face templates," p. 1188–1202, May 2019. [Online]. Available: DOI: 10.1109/TPAMI.2018.2827389

[14] M. Gomez-Barrero and J. Galbally, "Reversing the irreversible: A survey on inverse biometrics," *Computers & Security*, vol. 90, p. 101700, 2020. [Online]. Available: DOI: 10.1016/j.cose.2019.101700

[15] L. Sirovich and M. Kirby, "Low-dimensional procedure for the characterization of human faces," *Josa* a, vol. 4, no. 3, pp. 519–524, 1987. [Online]. Available: DOI: 10.1364/josaa.4.000519

[16] M. Turk and A. Pentland, "Eigenfaces for recognition," Journal of cognitive neuroscience, vol. 3, no. 1, pp. 71–86, 1991.

[17] D. E. King, "Dlib-ml: A machine learning toolkit," *Journal of Machine Learning Research*, vol. 10, pp. 1755–1758, 2009.

[18] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778. [Online]. Available: DOI: 10.1109/cvpr.2016.90

[19] G. B. Huang, M. Ramesh, T. Berg, and E. Learned-Miller, "Labeled faces in the wild: A database for studying face recognition in unconstrained environments," University of Massachusetts, Amherst, Tech. Rep. 07-49, October 2007.

[20] F. Schroff, D. Kalenichenko, and J. Philbin, "Facenet: A unified embedding for face recognition and clustering," *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun 2015. [Online]. Available: DOI: 10.1109/CVPR.2015.7298682

[21] K. Q. Weinberger, J. Blitzer, and L. K. Saul, "Distance metric learning for large margin nearest neighbor classification," in In NIPS. MIT Press, 2006.

[22] E. Parliament and of the Council, "Regulation (eu) 2016/679 of the european parliament and of the council of 27 april 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing directive 95/46/ec (general data protection regulation)," 2016.

[23] E. . E. P. R. Service, "Understanding algorithmic decision-making: Opportunities and challenges," 2019.

[24] X. Dong, K. Wong, Z. Jin, and J.-L. Dugelay, "A cancellable face template scheme based on nonlinear multi-dimension spectral hashing," Cancun, MEXICO, 05 2019. [Online]. Available: DOI: 10.1109/iwbf.2019.8739179

[25] L. Sweeney, "Simple demographics often identify people uniquely," 2000, Working paper.

[26] A. Narayanan and V. Shmatikov, "Robust de-anonymization of large sparse datasets," in *Proc. of the 29th IEEE Symposium on Security and Privacy*. IEEE Computer Society, May 2008, pp. 111–125. [Online]. Available: DOI: 10.1109/SP.2008.33

[27] A. Narayanan and V. Shmatikov, "De-anonymizing social networks," in *2009 30th IEEE Symposium on Security and Privacy*, 2009, pp. 173–187. [Online]. Available: DOI: 10.1109/sp.2009.22

[28] Y.-A. de Montjoye, L. Radaelli, V. K. Singh, and A. ". Pentland, "Unique in the shopping mall: On the reidentifiability of credit card metadata," *Science*, vol. 347, no. 6221, pp. 536–539, 2015. [Online]. Available: DOI: 10.1126/science.1256297

[29] T. Oliphant, "NumPy: A guide to NumPy," USA: Trelgol Publishing, 2006.

[30] W. McKinney, "Data structures for statistical computing in python," in *Proceedings of the 9th Python in Science Conference*, vol. 445. Austin, TX, 2010, pp. 51–56. [Online]. Available: DOI: 10.25080/majora-92bf1922-00a

[31] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg et al., "Scikit-learn: Machine learning in python," *Journal of machine learning research*, vol. 12, no. Oct, pp. 2825–2830, 2011.

[32] A. Geitgey, "face recognition: The world's simplest facial recognition api for python and the command line," 2020.

[33] Y. Zhang Zhifei, Song and H. Qi, "Age progression/regression by conditional adversarial autoencoder," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2017. [Online]. Available: DOI: 10.1109/cvpr.2017.463

[34] D. Dua and C. Graff, "UCI machine learning repository," 2017.

István Fábián is a technical assistant at the Budapest University of Technology and Economics (BME) since 2019. He is a member of the Balatonfüred Student Research Group. His research interests include privacy and security in machine learning, and he is also working on projects related to IoT and Industry 4.0 in the BME Technology Center.

Gábor György Gulyás has been involved with Privacy Enhancing Technologies since 2005. In 2015 he obtained the degree of PhD Budapest University of Technology and Economics (BME). The focus of his thesis was on how privacy and anonymity could be preserved in social networks against largescale re-identification attacks. Between 2015 and 2018 he was a PostDoc and Research Engineer in the Privatics team at INRIA (France). There, he was working on research projects related to web privacy and at the intersection of machine learning and privacy. Since 2019 he is a research fellow at BME with a special focus on (but not limited to) the privacy issues related to the IoT and machine learning technologies.