Predictive modeling toward identification of sex from lip prints - machine learning in cheiloscopy

Agustín F. Sabelli, Parag Chatterjee and Ma Florencia Pollo-Cattaneo

Grupo de Estudio en Metodologías de Ingeniería en Software (GEMIS), Universidad Tecnológica Nacional Facultad Regional Buenos Aires, Buenos Aires, Argentina

Abstract

Cheiloscopy is a forensic investigation technique that deals with identification of humans based on lips traces. Lip traces hold multifarious features and could be analyzed in different ways to identify the links with personal identifying features. Machine Learning holds strong application in this domain, especially for pattern recognition and further interpretation. This work is focused on a brief survey of existing machine learning approaches in Cheiloscopy. Also, a comparative study of predictive models has been presented based on an original dataset of lip prints where supervised models have been used to predict the biological sex of the persons using their lip traces. Machine learning on one hand automatizes the identification process and poses a significant potential in analyzing huge number of lip traces with considerable accuracy.

Keywords

Cheiloscopy, Machine Learning, Lip Prints, Biometrics, Image Processing, Pattern Recognition

1. Introduction

The outer surface of the human lips has many ridges and depressions that form a distinct pattern called "lip impression". The study of lines, fissures, wrinkles and stretch marks on the lip is called "Cheiloscopy". Although Cheiloscopy is a relatively new field among the large number of identification tools available to forensic experts, extremely useful information such as the identity of a person can be obtained from it. This is because they remain relatively stable and show gender differences [1]. Originally, Cheiloscopy has been a manual procedure where tools such as magnifying glasses, especial lamps and microscopes were used to analyze lip prints, making it prone to human errors. Fortunately, computational algorithms could be used to mitigate those errors. Furthermore, among numerous algorithms capable of solving these situations, the optimal algorithm needs to be selected, based on the efficiency, lines of code, memory utilized or a combination of all of them. In addition to this, for many applications there is not enough knowledge to transform the inputs into the desired outputs pertaining to forensics. The power of data science stands highly significant in this perspective, where the insights obtained from data could complement and enrich the traditional knowledge of forensics, which is essentially the essence of Machine Learning (ML).

ICAIW 2021: Workshops at the Fourth International Conference on Applied Informatics 2021, October 28–30, 2021, Buenos Aires, Argentina

agustinsabelli@hotmail.com (A. F. Sabelli); paragc@ieee.org (P. Chatterjee); fpollo@frba.utn.edu.ar (M. F. Pollo-Cattaneo)

ML is a subfield of Artificial Intelligence (AI); AI is defined as the intelligence exhibited by an artificial entity to solve complex problems. This entity is generally assumed to be a computer or machine. In other words, AI is the ability of an entity to learn from data and make decisions like a human being without being explicitly programmed. Unlike humans, machines do not require breaks, are able to analyze huge amounts of data simultaneously and present a low error rate [2]. There are many efficient automatic methods in Cheiloscopy using algorithms like dynamic time warping, top-hat transform, vote counting and the Hough transform [3]. Nonetheless, not a substantial work has been performed in the recent years to obtain, apart from the identity, meaningful information of a person such as the sex or age.

This work focuses on the predictive aspect of machine learning algorithms in Cheiloscopy and compares different supervised learning algorithms that can be used to identify the biological sex of a person based on their lip traces. First, the basic pipeline of a biometric authentication system is briefly described. For each step, the state-of-the art is reviewed, analyzed and discussed. Then, the implementation of each of the stages for this particular use case is explained. Moreover, the performance is evaluated for each classification algorithm in the results section. Finally, the conclusions are drawn based on the application and system performance, identifying the future lines of research as well.

2. Machine Learning in Cheiloscopy

In the field of biometrics, ML stands out for its ability to increase precision in the identification process. Biometric characteristics taken first instance are not al-ways the same as those taken a second time. Consequently, the use of machine learning techniques such as artificial neural networks, fuzzy logic, evolutionary computing, etc., has increased in demand [4].

An automated biometric system aims to correctly predict the identity of a per-son based on a biometric sample or check if it matches an existing one stored in a data base. This is done in five steps (see Figure 1).

Figure 1: Basic pipeline of a biometric authentication system

Every step can be seen as a separate ML task. For each task, it is necessary to find the optimal parameters and settings in order to improve the accuracy of the whole authentication process.

2.1. Input data

The input data refers to the raw data obtained directly from the sensor or data source. For this case, that would be the lip impressions. These can be collected from people by pressing their lips, with lipstick previously applied, against a cellophane paper or tape. Alternatively, they can be obtained from the surface of body parts, fabrics, or objects (see Figure 2) by using a brush and revealing powder [5].

Figure 2: Lip impression obtained using conventional revealing powder

Within the scope of this work, the only current open access databases found related to Cheiloscopy samples are SUT-Lips-DB [6] and the Biometrics Research at University of Silesia[1].

2.2. Segmentation

Segmentation or identification is the process in which the region of interest (ROI) is extracted from the input data, which can be done by techniques like Object Detection (DO). This computer vision technique creates a bounding box around each object found (See Figure 3). In Cheiloscopy, the ROI is the mucosal part called the Klein zone. This zone contains the characteristic patterns of the lips [7]. Furthermore, another important role that DO plays is to filter and eliminate any image that does not match a lip impression.

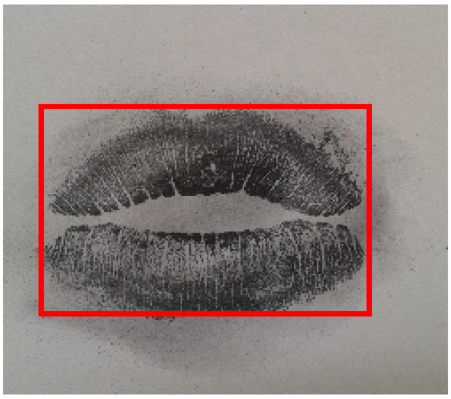


Figure 3: Digitized lip impression and its bounding box after applying DO

Alternatively, Image Segmentation (IS) is another option for segmentation. This technique creates a pixel mask for each object in the image, giving much more image granularity and enabling detailed analysis [8, 9] (See Figure 4).

[1]Biometrics Research at University of Silesia http://www.biometrics.us.edu.pl

Figure 4: Digitized lip impression after applying IS

2.3. Pre-processing

Generally, the data obtained from different data sources are not standardized and the pre-processing aims at solving that. In general, this process is carried out in several steps that reduce the complexity and increase the accuracy of the classifying algorithms [18]. Because it is not practical to perform a different pre-processing method for every single image, each of them is manipulated and trans-formed in such a way that any algorithm can process it.

Table 1
State-of-the-art of pre-processing techniques used in Cheiloscopy

Authors	Grayscale conversion	Brightness/contrast adjustment	Binarization	Normalization	Undesirable elements filtering	Background removal	Morphological operations	Figure detection	Feature enhancement
Smacki et al., [10]	•		•			•		•	•
Wrobel et al., [11]	•		•		•	•	•		•
Jain et al., [12]	•	•			•			•	
Wrobel and Froelich [13]	•							•	
Mousavi and Zarrabi [14]	•			•	•		•	•	
Travieso et al., [15]	•		•	•	•	•	•		•
Niu et al., [16]	•			•					
Lopez-Sanchez et al., [17]		•							

The pre-processing stage is of great importance in Cheiloscopy since lip impressions usually have many undesirable elements: fragments of skin, hair, and other undesired elements, which generate noise. For this reason, images of lip impressions should be cleaned and have any undesirable elements removed from them [11].

From the literature review of articles published during the last 5 years obtained from relevant and recognized scientific sources (Scopus and IEEE Xplore), a lack of specific or standardized way of pre-processing digitalized lip impressions for Cheiloscopy was inferred. In this respect, Table 1 shows the most common processes.

2.4. Feature Extraction

In the process of feature extraction, an initial set of data is reduced to more manageable groups for further processing, in this case, leading to dimensionality reduction. Because data often contains a large number of variables, it requires a lot of computing resources to process them [19]. In the case of lip impressions, the features of the Klein zone and the labial region are extracted.

2.4.1. Feature extraction using algorithms

Wrobel et al., [11], analyze the grooves' bifurcations of the upper and lower lips for their simplicity and unique pattern. To find such bifurcations, all black pixels in the digitized lip print are analyzed. The lip impression is then defined by a set of bifurcation systems determined by the Euclidean distance between their centers and their orientation angle (calculated based on the three angles between the bifurcations). Smacki et al., [10], obtain the lip pattern using the Top-Hat transform and filtering with special structures. Jain et al., [12], use Fast-Match, a Template Matching algorithm, to obtain a matrix of features. Wrobel and Froelich [13] and Mousavi and Zarrabi [14], extract the segments found in the lips using the Hough transform. These are later used to find the straight lines that make up the lip patterns. Travieso et al., [15], measured the point-to-point height and width of the labial contour divided into four regions. Lopez-Sanchez et al., [17], implement various dimensionality reduction techniques (Random projection, Principal Component Analysis, Independent Component Analysis, Simple Autoencoder) that are combined with different machine learning classifiers. Niu et al., [16], extract Gabor and Local Binary Pattern (LBP) characteristics from lip prints. Norhikmah et al., [20], make use of Two-dimensional Principal Component Analysis (2DPCA) as a feature extraction method. Wrobel et al., [21], mean distances, curvatures, contour shapes and lip area. These measurements form a vector of characteristics for each individual.

2.4.2. Manual Feature Extraction

Although there are different systems for classifying lip grooves, Kazuo Suzuki and Yasuo Tsuchihashi designed one that classifies them into six different types (see Figure 5) [22]. This proposal is superior as it provides greater detail and is easy to understand [23]. It was also concluded that the morphological pattern of the grooves is unique and exclusive [22].

In the literature, several papers illustrated used a traditional process of collecting lip prints using cellophane paper, eventually scanning and analyzing them using Adobe Photoshop

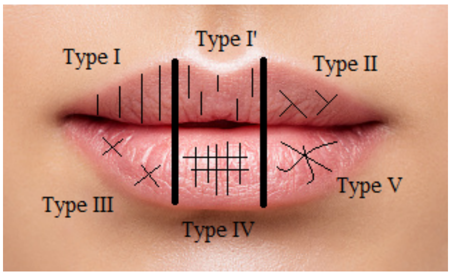


Figure 5: Groove classification proposed by Kazuo Suzuki and Yasuo Tsuchihashi

6.0 [24, 25, 26]. These are subsequently studied and classified according to the Suzuki and Tsuchihashi classification.

2.5. Classification

Classification is the process by which a set of data goes through a categorization process to assign different classes (often referred to as a target, label, or categories). This can be done on structured or unstructured data. Since there are no specific guidelines to figure out which algorithms to apply to a specific problem, it is recommended to conduct controlled experiments

Table 2
State-of-the-art of ML algorithms used in Cheiloscopy

Authors	ML is not used	Decision Tree or similar	Naive Bayes	K*	Radial Basis Function Classifiers	Rule-Based Classifier	Fuzzy clustering	Hidden Markov Models	Support Vector Machine	K-Nearest	Multilayer Perceptron	Self-Organizing Map
Smacki et al., [10]	•											
Wrobel et al., [11]		•	•	•	•	•						
Jain et al., [12]	•											
Wrobel and Froelich [13]							•					
Mousavi and Zarrabi [14]								•				
Travieso et al., [15]								•	•			
Niu et al., [16]									•			
Lopez-Sanchez et al., [17]									•	•	•	
Norhikmah and Haris[20]												•

and find out which algorithm and configuration performs best [27]. This is evident in the following table (Table 2), where the algorithms used to process the labial images are listed.

3. Materials and Methods — Predictive models in Cheiloscopy

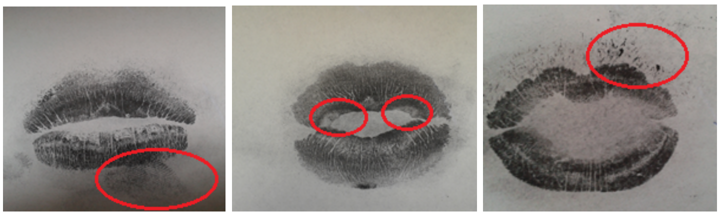
This work was focused to review and compare different supervised learning algorithms in terms of their performance and effectiveness to identify the biological sex of a person based on their lip prints. The methodology is presented in five steps, including the entire workflow from dataset preparation and preprocessing till the modeling and prediction.

3.1. Dataset preparation

Primarily, two common datasets were considered for the study — the Biometrics Research database from the University of Silesia, and the SUT-Lips database. The first database does not include the sex of the subjects in the metadata and was thus discarded, and the latter was selected as the final data source. The SUT-Lips-DB contains 50 folders, each having several lip traces as JPG files only for one person. However, after rejecting the lip prints with blurry appearance, 43 were selected based on their definition and clarity. Furthermore, for each folder only the best image was selected, reducing the selection to only one print per person. Subsequently, the final data set contains lip traces of 26 females and 17 males.

3.2. Data segmentation

Due to high level of noise and undesired elements in the image, most of the lip traces' ROI proved to be very hard to extract by automatic means and hence data segmentation was done on a selective and manual basis. Among the unwanted elements in the images, there were facial hair, finger prints and oral mucosa prints (see Figure 6).



(a) Finger print (b) Oral mucosa (c) Facial hair

Figure 6: Lips images

Adobe Photoshop was used to extract the ROI out of the lip prints; with the help of the magnetic loop, a tool used to make automatic selections, it was possible to separate the lower lip from the upper lip (see Figure 7). The advantage of the magnetic loop is that as points are selected, it automatically detects and picks the best path based on the contours. This tool has the peculiarity of adapting to the image that is below the area with which we are working. For this reason, it is very useful for selecting contours.

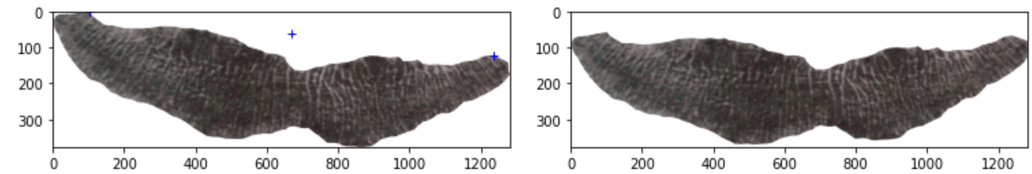
Figure 7: ROI of the lip print

After the segmentation, the traces of the upper and lower lip are saved in Portable Network Graphics (PNG) format in separate files. The naming convention for these files was `<index><sex>-<upper/lower>`. This is important, because they are treated differently at the beginning of the next stage depending on whether it is an upper or lower lip.

3.3. Pre-processing

First, based on the naming convention the image is vertically flipped if it is an upper lip. In case of a lower lip trace, it is not. This is because the algorithm that handles the feature selection later on was coded taking into consideration a lower lip trace. Nevertheless, the code is robust enough to deal with this.

The transparent pixels (of alpha channel value zero) are replaced by pixels of white color (equivalent to [255, 255, 255, 0] in [R, G, B, A] notation). After that, the image is converted to grayscale and horizontally aligned. To accomplish the latter, the corners of the lip are searched. Once the corners are found, the anchor point that will be used to perform the rotation which will horizontally align the lip is calculated. This point is the middle spot between the corners (see Figure 8).

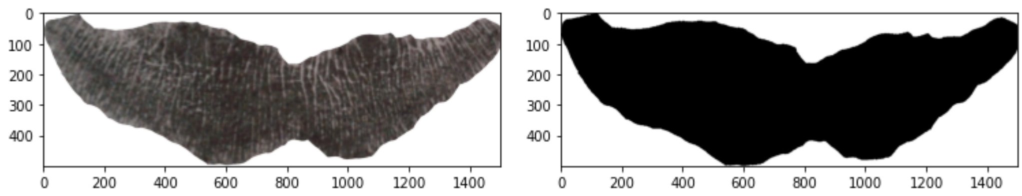


(a) Corners and anchor points

(b) Horizontally aligned lip

Figure 8: Lips corners

The next step is to get rid of the blank space left after the rotation that does not provide any information. To do this, the minimum bounding rectangle that can enclose the print is calculated. Taking this rectangle into account, the image is cropped. Afterwards, the image is resized so that all prints are the same size. This process is known as normalization and in this case it is 1500x500 pixels. Finally, as we are only interested in the shape, the image is binarized (see Figure 9).



(a) White space removed and normalization (b) Binarization

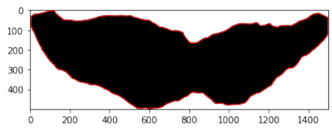
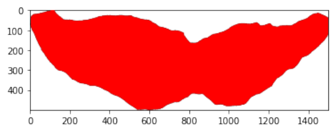
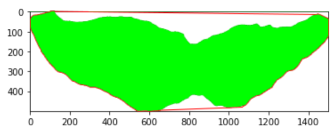
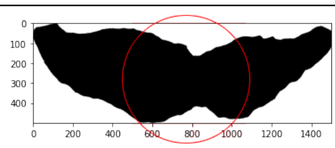
Figure 9: Lips Normalization

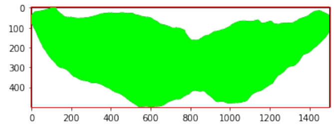
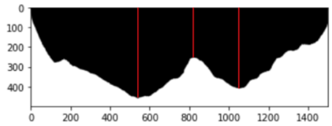
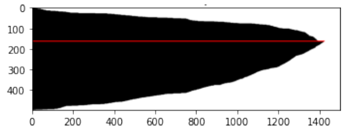
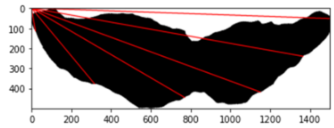
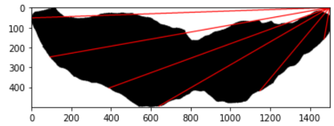
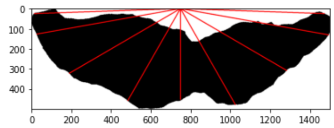
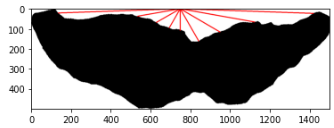
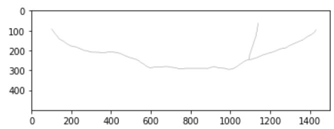
3.4. Feature selection

For all images the same biometric features have been determined. These were mainly inspired by Wrobel et al., work [21]. Each set of lip print-based features are denoted as $f_n = [f_1, \ldots, f_n]$ (see Table 3). For each side of the lip, 40 features have been extracted making 80 in total for the whole lip trace. These have been exported to a `.csv` file that is later used to feed the classification algorithms.

In addition, the features were standardized by removing the mean and scaling to unit variance. This was done to avoid bad behavior from the ML estimators used in the classification stage.

Table 3
Measurement of the lip print-based features

#	Type of measuring	Feature Vector	Visualization
(a)	Perimeter	$f_1 = [p]$	
(b)	Area	$f_2 = [a]$	
(c)	Solidity	$f_3 = [s]$	
(d)	Equivalent diameter	$f_4 = [ed]$	

(e)	Extent	$f_4 = [ex]$	
(f)	Main peaks and valley of the vertical projection	$f_5 = [p_1, v_1, p_2]$	
(g)	Maximum length of the horizontal projection	$f_6 = [l]$	
(h)	Distance from the left corner to the outer edges $[80°, 50°, 30°, 20°, 10°, 2°]$	$f_7 = [ld_1, ld_2, ld_3, ld_4, ld_5, ld_6]$	
(i)	Distance from the right corner to the outer edges $[80°, 50°, 30°, 20°, 10°, 2°]$	$f_8 = [rd_1, rd_2, rd_3, rd_4, rd_5, rd_6]$	
(j)	Distance from the upper center to the outer edges $[2°, 10°, 30°, 60°, 90°, 120°, 150°, 170°, 178°]$	$f_9 = [cd_1, cd_2, cd_3\ , cd_4, cd_5, cd_6, cd_7, cd_8, cd_9]$	
(k)	Distance from the upper center to the inner edges $[2°, 10°, 30°, 60°, 90°\ 120°, 150°, 170°, 178°]$	$f_{10} = [id_1, id_2, id_3, id_4, id_5, id_6, id_7, id_8, id_9]$	
(l)	Length of the skeletonized lip print	$f_{11} = [sk]$	

3.5. Classification

Since the number of features collected from the previous stage was substantially high, only the most relevant ones were preserved. The Extra Tree Classifier was used to perform this dimensionality reduction. Also known as Extremely Randomized Trees, this is a type of ensemble learning technique composed of a large number of decision trees where the final decision is obtained taking into account the prediction of every tree. For this type of classifier, all the features and splits (questions about the data) are selected at random. In this forest structure, the relevance of a feature is given by the Gini importance. To achieve the transformation of data from a high-dimensional space into a low-dimensional space, each feature is ordered in descending order according to this value. Finally, only the top k features are selected.

As a result, only a certain number of features (Table 4) out of the originally 80 features were kept. The final number of features depends on the best possible result for each classification algorithm. Furthermore, with the intention of getting better and more stable results the dataset was split into different sequences of train and test portions (also known as cross-validation) [28, 29]. Hence, it was split into 5 different training and validation datasets. Since the k-fold approach was used [30], and k is equal to 5, the train to test ratio was 34:9 (80% for training an 20% for test) for each fold.

As stated before, the number of samples in the cleaned dataset was pretty small (43 in total). In addition to this, since the data has been labeled and the goal is to predict or classify future observations, supervised learning was chosen. In particular, the following algorithms were selected:

3.5.1. Logistic Regression (LR)

LR is a statistical learning method for classification. The term “logistic” refers to the “log odds” probability that is modeled. The term “odds” is defined as the ratio of the probability that an event occurs to the probability that it doesn’t. It seeks to predict the effect of a series of variables on a binary response variable and classify observations by estimating the probability that an observation is in a particular category [31, 32].

3.5.2. Multilayer Perceptron (MLP)

MLP is a robust and nonlinear neural network model that operates as an approximation function. It is the expansion of a simple neural network, with multiple hidden layers that allow solving extremely complex problems. MLP uses back-propagation as its learning algorithm, a generalization of the Least Mean Squared rule. One of the major problems of this model is over-fitting, mainly due to large number of hidden layers being used [33].

3.5.3. Support Vector Machine (SVM)

SVM works under the principle of margin calculation. It can be simply defined as a prediction tool that looks for a particular division line called a hyperplane that easily separates datasets or classes, thus avoiding overtraining of the data. This hyperplane is generated in an iterative manner, which is used to minimize the error. In practice, SVM uses a technique called the “kernel trick” where a low-dimensional input space is transformed into a higher dimensional space. This is because a higher dimensional transformation can allow us to separate data in order to make classification predictions [34].

3.5.4. Naïve Bayes (NB)

NB is typically used for classification and clustering purposes. It is called naïve (naive), since it assumes that all variables contribute to the classification and are mutually co-related. This technique is based on Bayes theorem and is used when the dimensionality of the inputs is high. Using Bayes theorem, we can find the probability of A happening, given that B has occurred. NB algorithms are fast and fairly easy to implement. However, their biggest disadvantage is the need of independent predictors. In most cases, these predictors are dependent and the performance suffers from it [35].

3.5.5. K-Nearest Neighbor (K-NN)

K-NN is an algorithm where given "n" training vectors, identifies the "k" nearest neighbors of "c" (a feature vector that we want to estimate its' class), regardless of labels. It then predicts the class of "c" based on the neighbors who are majority. The main advantages of this algorithm are its simplicity and straight forward implementation. On the other hand, the main draw-back is the complexity in searching the nearest neighbors for each sample [31, 36].

4. Results

Performance statistics for the different classification algorithms are shown in Table 4. Only the top combinations of features and classification algorithms are displayed.

Table 4
Performance statistics

Classifier	# Features	Accuracy	f1 score	AUC
K-NN	28	0.82	0.86	0.80
LR	23	0.79	0.80	0.75
NB	24	0.77	0.81	0.77
SVM	25	0.70	0.77	0.78
MLP	25	0.65	0.66	0.59

The optimal number of features differs across classifiers, but overall it's less than a half of the original 80 obtained from the feature selection stage. The accuracy of a ML classification algorithm is one way to measure how often the algorithm classifies a data point correctly. The f1 score and the Area Under the Curve (AUC) values are also of utter importance. This is because the first one indicates the precision and recall of the model, whereas the second represents how much the model is capable of distinguishing between classes (male or female) [35]. In general, the Accuracy and AUC is above 75% for all the models, while the f1 score is greater than 80%. The best accuracy for the sex prediction was verified in the K-NN model (0.82). The top f1 score (0.86) and AUC (0.80) can also be found in this particular model. Contrastingly, the lowest accuracies are detected from MLP (0.65). The same is the case for the f1 score (0.66) and AUC (0.59).

5. Conclusions

This paper focused on the use of different supervised learning algorithms to identify the biological sex of a person based on their lip traces. We also presented a summary of the state-of-the-art in Cheiloscopy in the context of ML. An original dataset was created out of the features automatically extracted from one of the only two open-source databases available for lip prints. This dataset was later used to use in the classification algorithms and their performance was compared and analyzed in the previous section. Comparing the different types of models used, K-NN provided the best performance, with all the models providing satisfactory accuracy in determining the biological sex of a person based on their lip traces, thus validating the focal objective of this work. The principal challenge was in the aspect of image segmentation, which was done manually, considering the high amount of noise and number of undesired elements. Scarcity of open access databases containing images of lip prints posed another challenge, especially with respect to the training of algorithms. Also, fine tuning the hyper-parameters of the estimators for each classification model was a significant challenge; however, it was resolved using a hyperparameter tuning function [37]. In the future, further work could be performed in the domain of unsupervised algorithms and deep learning as well, provided the availability of sufficiently big datasets. Similarly, it might provide further information like age that could still be extracted out of lip prints.

References

[1] A. Kumar, S. N. Prasad, V. Kamal, S. Priya, M. Kumar, A. Kumar, Importance of cheiloscopy, Int J Oral Care Res 4 (2016) 48–52.

[2] L. Rouhiainen, Artificial Intelligence: 101 things you must know today about our future, Lasse Rouhiainen, 2018.

[3] S. Sandhya, R. Fernandes, Lip print: an emerging biometrics technology-a review, in: 2017 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC), IEEE, 2017, pp. 1–5.

[4] P. Akulwar, N. A. Vijapur, Secured multi modal biometric system: A review, in: 2019 Third International conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud)(I-SMAC), IEEE, 2019, pp. 396–403.

[5] S. Kannan, K. Muthu, S. Muthusamy, P. Sidhu, Cheiloscopy-a vital tool in crime investigation, Int J Forensic Sci Pathol 3 (2015) 89–93.

[6] D. Mrozek, Sut-lips-db - a database of lips traces, 2017. URL: https://dx.doi.org/10.21227/H2R04P. doi:10.21227/H2R04P.

[7] L. Chatra, T. Peter, A. Ahsan, et al., Cheiloscopy, International Journal of Forensic Odontology 1 (2016) 48.

[8] F. Velosa, H. Florez, Edge solution with machine learning and open data to interpret signs for people with visual disability., in: CEUR Workshops Proceedings, 2020, pp. 15–26.

[9] J. Yepes Zuluaga, F. Yepes-Calderon, Tensor domain averaging in diffusion imaging of small animals to generate reliable tractography, ParadigmPlus 2 (2021) 1–19.

[10] L. Smacki, J. Luczak, Z. Wrobel, Lip print pattern extraction using top-hat transform,

in: Proceedings of the 9th International Conference on Computer Recognition Systems CORES 2015, Springer, 2016, pp. 337–346.
[11] K. Wrobel, P. Porwik, R. Doroz, Effective lip prints preprocessing and matching methods, in: Proceedings of the 9th International Conference on Computer Recognition Systems CORES 2015, Springer, 2016, pp. 347–357.
[12] S. Jain, V. Poojitha, M. Bhatia, A cheiloscopic approach for unique identification among indian subpopulation, in: Intelligent Communication, Control and Devices, Springer, 2018, pp. 1729–1738.
[13] K. Wrobel, W. Froelich, Recognition of lip prints using fuzzy c-means clustering, Journal of Medical Informatics & Technologies 24 (2015).
[14] A. S. Mousavi, H. Zarrabi, A hidden markov model (hmm) scheme for lip based identification utilizing vertical grooves angles, International Journal of Computer Applications 975 (2018) 8887.
[15] C. M. Travieso, A. G. Ravelo-García, J. B. Alonso, J. M. Canino-Rodríguez, M. K. Dutta, Improving the performance of the lip identification through the use of shape correction, Applied Intelligence 49 (2019) 1823–1840.
[16] B. Niu, J. Sun, Y. Ding, Lip print recognition using gabor and lbp features, DEStech Transactions on Computer Science and Engineering (2016).
[17] D. López-Sánchez, A. G. Arrieta, J. M. Corchado, Machine learning methods for automatic cheiloscopy on facial images, in: 8th International Conference on Imaging for Crime Detection and Prevention (ICDP 2017), IET, 2017, pp. 55–60.
[18] C. Zhu, D. Gao, Influence of data preprocessing, Journal of Computing Science and Engineering 10 (2016) 51–57.
[19] D. Sanyal, N. Bosch, L. Paquette, Feature selection metrics: Similarities, differences, and characteristics of the selected models., International Educational Data Mining Society (2020).
[20] M. K. Norhikmah, S. K. H. Angriawan, Implementation of 2dpca and som algorithms to determine sex according to lip shapes, in: 2019 4th International Conference on Information Technology, Information Systems and Electrical Engineering (ICITISEE), IEEE, 2019, pp. 101–106.
[21] K. Wrobel, R. Doroz, P. Porwik, J. Naruniec, M. Kowalski, Using a probabilistic neural network for lip-based biometric verification, Engineering Applications of Artificial Intelligence 64 (2017) 112–127.
[22] K. Suzuki, Y. Tsuchihashi, Personal identification by means of lip prints, J Forensic Med 17 (1970) 52–57.
[23] M. Bajpai, N. Pardhe, B. Chandolia, M. Arora, Cheiloscopy-an overview of its limitations and future per-spectives, J For Med Leg Aff 1 (2016) 106.
[24] S. Nadeem, G. Manzoor, S. Pervez, Lip forensics; cheiloscopic pattern among population of punjab, pakistan., The Professional Medical Journal 26 (2019) 1177–1182.
[25] P. Kesarwani, A. Choudhary, Establishing role of cheiloscopy in person identification and sex determination, Journal of Orofacial Research (2017) 4–7.
[26] B. S. Sharma, V. Gupta, H. Vij, E. Sharma, N. Tyagi, S. Singh, et al., Cheiloscopy: A tool for antemortem identification, Indian Journal of Dental Sciences 9 (2017) 176.
[27] J. Brownlee, Machine learning mastery with weka, Ebook. Edition 1 (2019) 4.

[28] D. Berrar, Cross-validation., 2019.
[29] J. Hernandez, K. Daza, H. Florez, Alpha-beta vs scout algorithms for the othello game, in: CEUR Workshops Proceedings, volume 2846, 2019, pp. 65–79.
[30] S. L. Developers, Cross-validation: Evaluating estimator performance, 2018.
[31] S. Ray, A quick review of machine learning algorithms, in: 2019 International conference on machine learning, big data, cloud and parallel computing (COMITCon), IEEE, 2019, pp. 35–39.
[32] H. Florez, S. Singh, Online dashboard and data analysis approach for assessing covid-19 case and death data, F1000Research 9 (2020).
[33] H. Taud, J. Mas, Multilayer perceptron (mlp), in: Geomatic Approaches for Modeling Land Change Scenarios, Springer, 2018, pp. 451–455.
[34] M. Somvanshi, P. Chavan, S. Tambade, S. Shinde, A review of machine learning techniques using decision tree and support vector machine, in: 2016 international conference on computing communication control and automation (ICCUBEA), IEEE, 2016, pp. 1–7.
[35] B. Mahesh, Machine learning algorithms-a review, International Journal of Science and Research (IJSR).[Internet] 9 (2020) 381–386.
[36] J. Hernandez, H. Florez, An experimental comparison of algorithms for nodes clustering in a neural network of caenorhabditis elegans, in: International Conference on Computational Science and Its Applications, Springer, 2021, pp. 327–339.
[37] Scikit-learn developers, Tuning the hyper-parameters of an estimator, 2007.