



HAL
open science

Computational diagnostic methods on 2D photographs: A review of the literature

Quentin Hennocq, Roman Hossein Khonsari, Vincent Benoît, Marlène Rio,
Nicolas Garcelon

► To cite this version:

Quentin Hennocq, Roman Hossein Khonsari, Vincent Benoît, Marlène Rio, Nicolas Garcelon. Computational diagnostic methods on 2D photographs: A review of the literature. *Journal of Stomatology, Oral and Maxillofacial Surgery*, 2021, 122 (4), pp.e71-e75. 10.1016/j.jormas.2021.04.003 . hal-04360135

HAL Id: hal-04360135

<https://inria.hal.science/hal-04360135>

Submitted on 22 Jul 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

1 **Title: Computational diagnostic methods on 2D photographs: a review of the**
2 **literature.**

3 Running title: Computational diagnostic methods on 2D photographs

4

5 Quentin **Hennocq**, M.D.^{1,2}; Roman Hossein **Khonsari**, M.D. PhD.^{1,2}; Vincent **Benoît**,
6 PhD.¹; Marlène **Rio**, M.D.^{1,3}; Nicolas **Garcelon**, PhD.¹

7

8 1. Institut Imagine, Paris Descartes University-Sorbonne Paris Cité, Paris, France.

9 2. Department of Maxillo-Facial Surgery and Plastic Surgery, Hôpital Universitaire

10 Necker – Enfants Malades, Assistance Publique – Hôpitaux de Paris ; Centre de

11 Référence des Malformations Rares de la Face et de la Cavité Buccale

12 MAFACE, Filière Maladies Rares TeteCou ; Faculté de Médecine, Université de

13 Paris, Paris, France.

14 3. Department of Genetics, IHU Necker-Enfants Malades, University Paris

15 Descartes, Paris, France.

16

17 **Corresponding author:**

18 Quentin Hennocq, M.D.

19 Institut Imagine ; 24, boulevard du Montparnasse, 75015 Paris

20 Phone : +33142161305

21 Fax : +33142161301

22 Email: quentin.hennocq@aphp.fr

23

24 No financial assistance was received in support of the study.

25 **ABSTRACT**

26

27 Here we provide a literature review of all the methods reported to date for analyzing 2D
28 pictures for diagnostic purposes. Pubmed was used to screen the MEDLINE database
29 using MeSH (Medical Subject Heading) terms and keywords. The different recognition
30 steps and the main results were reported. All human studies involving 2D facial
31 photographs used to diagnose one or several conditions in healthy populations or in
32 patients were included. We included 1515 articles and 27 publications were finally
33 retained. 67% of the articles aimed at diagnosing one particular syndrome versus
34 healthy controls and 33% aimed at performing multi-class syndrome recognition. Data
35 volume varied from 15 to 17,106 patient pictures. Manual or automatic landmarks were
36 one of the most commonly used tools in order to extract morphological information
37 from images, in 22/27 (81%) publications. Geometrical features were extracted from
38 landmarks based on Procrustes superimposition in 4/27 (15%). Textural features were
39 extracted in 19/27 (70%) publications. Features were then classified using machine
40 learning methods in 89% of publications, while deep learning methods were used in
41 11%. Facial recognition tools were generally successful in identifying rare conditions in
42 dysmorphic patients, with comparable or higher recognition accuracy than clinical
43 experts.

44 **Keywords:** literature review; machine learning; deep learning; photograph;
45 dysmorphology; diagnosis

46 **Declarations of interest:** none.

47

48 **INTRODUCTION**

49

50 Shape analysis on 2D pictures can involve computations based on pre-defined rules; in
51 machine learning approaches, these rules are replaced by the ability of the model to
52 learn from examples (1). Applications of machine learning are increasing in healthcare
53 (1), in a variety of fields such as radiology (2), dermatology (3) and surgery (4). Inputs
54 or features are converted into outputs or labels. Deep learning is defined as a subtype of
55 machine learning in which artificial neural networks are used. These neural networks
56 can encode relationships between features and labels that are not accessible to the
57 human brain, because of their complexity or because they do not match the way our
58 brain processes data (1).

59 Examples of relevant inputs in healthcare are X-rays, texts, pathology slices and clinical
60 photographs. Among these inputs, clinical photographs are particularly important in the
61 diagnosis of craniofacial malformations. In fact, 30-40% of the 7000 rare diseases have
62 craniofacial anomalies (5). Many among these 2-3000 syndromes have minor facial
63 features or are very uncommon: their screening requires the eye of an expert in
64 dysmorphology. Due to this specific clinical issue, there have been a recent increase in
65 the number of scientific publications dedicated to the diagnosis of rare conditions using
66 computational methods on facial photographs (6).

67 Here we provide a literature review of all the methods used for computational diagnosis
68 on facial photographs. This overview will be of use for physicians and scientists
69 interested in syndrome recognition, and more generally, in facial recognition.

70

71

72 **METHODS**

73

74 We performed a review of the literature according to the framework described by
75 Arksey and O'Malley (7). We aimed at reporting all articles describing the
76 computerized identification of a disease and/or a syndrome based on 2D facial
77 photographs. PubMed was used to screen the MEDLINE database using MeSH
78 (Medical Subject Heading) terms along with keywords from titles and abstracts (see
79 request Table 1).

80 All human studies involving 2D facial photographs used to diagnose one or several
81 conditions in healthy or patient populations were included. Only studies with diagnosis
82 confirmation by clinical experts and/or molecular investigations were retained. Articles
83 on prenatal diagnosis, extra-facial diseases and dermatological diseases of the face (e.g.
84 facial acnea or naevi) were excluded. Studies assessing treatment results, such as
85 craniofacial surgery results, were excluded, as well as studies dedicated to the detection
86 of pain or other emotions. We also excluded studies based on the diagnosis of patients
87 using a pre-existing algorithm, such as *Face2Gene* (8).

88 One reviewer (QH) screened all the titles and/or abstracts. When exclusion criteria were
89 not clear after reading title and abstract, the full text was assessed. The relevant
90 publications were then charted with the following items: year of publication, first
91 author, country of affiliation, study design or objective, disease of interest, origin of
92 data and number of patients. We reported the pre-processing of the pictures, *i.e.* every
93 treatment the authors applied to the photograph before extracting features or using
94 classifiers. Concerning feature extraction, we reported the use of: (1) landmarks, with
95 manual and/or automatic placement, (2) geometric features, *i.e.* distances, angles or

96 other morphometric data and (3) texture features, based on the spatial arrangement of
97 intensities on the image. We listed the classifiers and the validation methods of machine
98 learning designs. We considered that authors had used machine learning methods when
99 predictions were made on a test set with a classifier or a model trained on a training set.

100

101 **RESULTS**

102

103 With the pre-defined request, we listed 1515 results on PubMed. After applying all the
104 exclusion criteria on reading title, abstract and full text when necessary, we included 27
105 articles. These publications were fully read and charted (Table 2). The region of
106 affiliation of the first author was North America – USA only in 37%, Europe –
107 Germany, UK and Croatia in 33%, and Asia – China, India and Turkey in 30%.

108

109 ***1. Study designs and conditions***

110 We defined two types of studies depending of the number of diseases to be diagnosed.

111 The first and most common study design – in 18/27 (67%) publications – intended to
112 diagnose one condition in the general population versus healthy controls using facial
113 photographs. Facial pictures of the following conditions were tested versus healthy
114 controls: Down syndrome (10–13), Cornelia de Lange syndrome (16), Rubinstein-Taybi
115 syndrome (16), 22q11.2 deletion syndrome (17), Laron syndrome (18), Williams-
116 Beuren syndrome (19), acromegaly (20–24), Cushing disease (25), Turner syndrome
117 (26), Chronic Fatigue Syndrome (27) and schizophrenia (28).

118 The second most prevalent study design – in 9/27 (33%) publications – consisted in
119 using multi-syndrome classification, *i.e.* in differentiating syndromes from each other

120 and from the general population (8,29–38). For example, Ferry et al. (29) aimed at
121 diagnosing 8 syndromes – Angelman, Apert, Cornelia de Lange, Down, fragile X,
122 progeria, Treacher-Collins and Williams-Beuren syndromes – in the general population.
123 Gurovich et al. (8) aimed at diagnosing 216 different syndromes in a population of
124 affected patients.

125 The conditions reported in the included studies were:

- 126 – genetic syndromes: Down (29), Angelman (8), Prader-Willi (34), Cornelia de Lange
127 (8), X fragile (34), Apert (29), Progeria (29), Treacher-Collins (29), Williams-
128 Beuren (32), Sotos (30), Cri-du-chat (34), Smith-Lemli-Opitz (34), 22q11.2 deletion
129 (34), Turner (8), Noonan (8), Mucopolysaccharidosis (34), Rubinstein-Taybi (8),
130 Hurler (36), Wolf-Hirschhorn (36) and 183 others in (8),
- 131 – endocrine conditions with a facial phenotype (e.g. acromegaly and Cushing
132 syndrome) (23)
- 133 – psychiatric conditions: schizophrenia (28) and chronic fatigue syndrome (27).

134

135 **2. Data description**

136 The photographic data had diverse origins: (1) pictures taken during clinics
137 (9,10,12,15,17–25,27,30,32–36,39–41), (2) pictures from articles or books
138 (14,16,17,29,38), (3) pictures from pre-existing cohorts (8,37,42,43) and (4) pictures
139 publicly available on the internet (11,29). The data volume varied from 15 pictures of
140 affected patients (9) to 17,106 for Gurovich et al. (8). The average number of affected
141 (non-control) patients in the studies was 844 (+/- 3272); more precisely, 114 (+/- 138)
142 affected individuals were included in average in order to detect one syndrome versus the
143 general population and 2085 (+/- 5309) affected individuals in order perform multi-

144 syndrome classification. There was one photograph per patient in all publications.
145 Authors used age-, sex-, (15,19,22,25,40) and ethnicity-matched controls (12) in 6/27
146 (22%) studies.

147

148 **3. Image pre-processing**

149 Pre-processing corresponded to the preparation of pictures before extracting data and
150 applying classifiers. It is a useful step in machine learning: images are grouped (face,
151 profile), resized and transformed (rotation, cropping, lighting) to optimize analysis.

152 An automated facial detection system was used in several studies to sort frontal views
153 from lateral views and then crop the pictures. The most common algorithm in this field
154 was the *Haar cascade* tool developed by Viola and Jones (1). One author used manual
155 cropping, scaling and rotation to maximize the classifier performances (9). This
156 approach was not adapted to large datasets. Manual or automated cropping was useful
157 for little standardized databases, for instance when several individuals were present on
158 pictures (*e.g.* the child and his mother) (6).

159

160 **4. Image processing**

161 All authors applied two exclusive image processing approaches: (1) use of geometric
162 and/or textural features in order to describe phenotypes or (2) deep learning process in
163 order to detect a syndrome without describing phenotypes.

164

165 **Landmarks positioning**

166 Landmarks are defined by reproducible points of interest on an image. They are one of
167 the most common tools used to extract phenotypic data, as in 22/27 (81%) publications.

168 In 7/27 (26%) publications, the authors placed landmarks manually and automatic
169 landmarking was used in 16/27 (59%) publications.

170 In 4/27 (15%) publications (9,11,24,41), the authors did not use landmarks to analyze
171 faces and phenotypes and preferred assessing the overall shape and textures.

172

173 ***Frontalization and facial alignment.***

174 Frontalization is a process modifying a frontal image in order to improve the orientation
175 of the frontal view. Kong et al. (21) used alignment, based on Procrustes
176 superimposition (47,48), in order to compute frontal pictures from photos with capture
177 angles varying from -45° to $+45^\circ$ around the standard coronal plane. The final frontal
178 face was produced using the corresponding symmetric sides, which raised numerous
179 issues in conditions with facial asymmetry.

180

181 ***Geometric features extraction***

182 Geometric features were extracted from landmarks and used for disease detection. The
183 main advantage of using landmarks was to be more flexible with pose changes and
184 different angles of view. One straightforward approach was to measure Euclidean
185 distances and angles from landmarks (22,23,25,26,35,36).

186 Another approach to geometrical information extraction, used in 4/27 (15%)
187 publications, was Procrustes analysis (10,16,29,49,50). Procrustes distance was defined
188 as the deformation necessary to fit a dysmorphic face on a control face via an iterative
189 least-squares process. Each of the faces were centered and scaled using a centroid (51)
190 to minimize sum of squared differences between the landmarks. Douglas and

191 Mutsvangwa (52) concluded that Procrustes-based approaches were the best methods
192 for shape analysis in terms of prediction accuracy.

193

194 ***Texture features extraction***

195 A texture extraction system was used in 18/27 (67%) publications, in association with
196 geometric features extraction (8,11,12,14,22,23,25–27,32,37,41,48) or independently
197 (9,11,41). The most common methods were Gabor Wavelet Transformation (GWT)
198 used in 12/27 (44%) (54,55) and/or a Local Binary Patterns (LBP) in 7/27 (26%)
199 publications (56). These methods provide information about the surface located
200 between the landmarks. For example, zygomatic hypoplasia in Treacher Collins
201 syndrome would be better detected by analyzing the shadows on a photo than by
202 measuring angles between landmarks.

203

204 The information on geometric / texture features created extensive data matrices, with
205 the need to reduce dimensionality before incorporation into classifiers. For this purpose,
206 Principal Component Analysis (PCA) was used in 17/27 (70%) publications.

207 Gurovich et al. (8) did not use any geometric or textural features extraction step for
208 prediction making after facial alignment based on different landmarks. Their approach
209 is referred to as deep learning because it corresponds to a *black box* approach
210 diagnosing dysmorphic features without going through a phenotype description process.
211 When using deep learning, authors cannot easily explain what contributed to the
212 diagnosis of a particular syndrome.

213

214 **5. Classification**

215 Machine learning was defined by using a training set to make predictions on a
216 validation or test set. Based on this, 3/27 (11%) publications were not using machine
217 learning methods.

218 In these studies, diverse classifiers were used. The simplest classifiers were Generalized
219 Linear Models (GLM) such as logistic regression and linear discriminant analysis, as
220 used in 4/27 (15%) publications (16,21). Their output was a linear combination of input
221 facial data.

222 SVM – used in 13/27 (48%) publications – was a classification method aiming at
223 separating two data sets with distance maximization based on an optimal separating
224 hyperplane (57). Bounds between datasets and the optimal separating hyperplane were
225 referred to as ‘support vectors’ (68).

226 The k-nearest neighbors (kNN) rule – used in 11/27 (41%) publications – classified
227 each unlabeled example by the majority label of its k-nearest neighbors in the training
228 set (59). It stored instances of previously earned images. Classification was obtained
229 from a majority vote of the nearest neighbors (21). The kNN classifier was particularly
230 adapted for multiclass detection.

231 Random forests (RT) – used in 2/27 (7%) publications – were another technique (21)
232 consisting in growing many classification trees. The forest chose the classification tree
233 having the best prediction. In addition, a subset of features was randomly selected from
234 the optional features in order to grow the tree at each node. Random forests were
235 particularly adapted for large number of weak classifiers.

236 Deep learning – Convolutional Neural Networks (CNN) or Artificial Neural Networks
237 (AAN) – was used in 3/27 (11%) publications. Kong et al. (21) also used CNN with
238 convolutional layers, pooling layers and fully-connected layers. Gurovich et al. (8) used

239 Deep Convolutional Neural Networks (DCNN) (60) for each of the cropped facial areas,
240 which were then combined by averaging to form a model for a multiclass problem.
241 Kong et al. (21) compared the performances of these machine learning techniques
242 (linear functions, SVM, kNN, RF and CNN): better classification abilities were found in
243 terms of accuracy for SVM with detected features and for CNN without detected
244 features.

245

246 **6. Clinical results**

247 Saraydemir et al. (9) and Zhao et al. (10) found an overall accuracy of respectively 97%
248 and 97.9% with 15 and 24 patients for diagnosing Down syndrome in a healthy
249 population. Both these performances were higher than clinical experts. Kruszka and his
250 team found an accuracy of 95% for detecting 22q11.2 deletion syndrome in 156 patients
251 vs. the general population (17), 94% for Down syndrome (12) in 129 patients, 90% for
252 Williams-Beuren syndrome in 286 patients (19), 94% for Cornelia de Lange in 246
253 patients (15) and 89% for Noonan syndrome in 161 patients (40). Kong et al. (21) used
254 the F1-score in order to evaluate the performances of their classifier in detecting
255 acromegaly in 527 patients. They obtained a F1-score of 92% with landmarks and SVM
256 and 91% without landmarks and with deep learning methods. Accuracies of 86% and
257 82% were obtained to detect acromegaly among a healthy population in 24 (20) and 57
258 patients (22).

259 Ferry et al. (29) used the *Clustering Improvement Factor* (CIF) after implementing kNN
260 on 8 syndromes and 1363 patients. Treacher-Collins syndrome was then clustered 23.5-
261 fold better than by chance alone, Down syndrome 11.6-fold better, and Williams-
262 Beuren 9.4-fold-better. The overall accuracy for detecting one syndrome vs. the 7 others

263 was 76% and the overall accuracy for detecting syndromes vs. a healthy population was
264 95%. Another method for displaying results in a multi-syndrome classification design
265 was to provide a Top-5 or Top-10 accuracy, *i.e.* to evaluate the probability of a
266 condition to be ranked as the 5 or 10 most probable syndromes by the classifier.
267 Gurovich et al. (8), in 17,106 patients with 216 syndromes, found a Top-10 accuracy of
268 91%, a Top-5 accuracy of 85% and a Top-1 accuracy of 61%. Kuru et al. (38) found
269 53% accuracy on a multi-class syndrome detection of 15 conditions with 92 patients.
270 Boehringer et al. (32) had an overall accuracy of 21% for multi-syndrome detection on
271 14 conditions in 202 patients, which is very low compared to the performances of the
272 clinical experts.

273

274 **DISCUSSION**

275

276 Machine learning approaches used to diagnose clinical conditions on 2D photographs
277 generally followed similar frameworks (Figure 1): either features were extracted from
278 photographs to diagnose a disease or the syndrome was detected directly without feature
279 extraction by deep learning. In case of features extraction, the steps were: image pre-
280 processing, image processing with or without landmark placement, extraction of
281 geometrical and/or textural features, reduction of the dimensions of the input, and
282 classification and validation using training and test sets. Procrustes analysis was a
283 reliable technique for the extraction of morphometric parameters (52). Texture
284 extraction techniques seemed of interest in adding information to the geometric features
285 (27).

286 Deep learning methods are a promising tool with much space for performance
287 improvement (8). The number of publications using deep learning was however limited
288 because this approach requires large data volumes – 17,106 patients in Gurovich et al.
289 (8) – in order to train the model. The challenge in the future years will be to build large
290 photographic databases and train neural networks in order to improve predictions.
291 Comparing publications and diagnosis performances of machine learning tools was
292 difficult due to the variety of assessment parameters: Boehringer (32) used accuracy,
293 Zhao (10) precision and recall, Dudding-Byth (37) top-10, top-5 and top-1 accuracies,
294 and Ferry (29) the Clustering Improvement Factor (CIF). Evidence-based comparisons
295 of machine learning in facial recognition would greatly benefit from more homogenous
296 result assessment approaches.

297

298 **CONCLUSION**

299 Facial recognition approaches are efficient in identifying rare conditions using 2D
300 photographs. Machine-learning algorithms improve recognition accuracy. Deep learning
301 techniques such as CNNs are promising, but larger datasets are needed. Based on the
302 current results of available algorithms, we can predict that collaborative studies with
303 massive shared datasets could shortly lead to the development of algorithms with almost
304 perfect diagnostic abilities.

305 **REFERENCES**

306

307 1. Rajkomar A, Dean J, Kohane I. Machine Learning in Medicine. *N Engl J Med*.
308 2019 Apr 4;380(14):1347–58.

309 2. Choy G, Khalilzadeh O, Michalski M, Do S, Samir AE, Pianykh OS, et al.
310 Current Applications and Future Impact of Machine Learning in Radiology. *Radiology*.
311 2018 Aug;288(2):318–28.

312 3. Novoa RA, Gevaert O, Ko JM. Marking the Path Toward Artificial Intelligence–
313 Based Image Classification in Dermatology. *JAMA Dermatol*. 2019 Oct
314 1;155(10):1105–6.

315 4. Loftus TJ, Tighe PJ, Filiberto AC, Efron PA, Brakenridge SC, Mohr AM, et al.
316 Artificial Intelligence and Surgical Decision-Making. *JAMA Surg* [Internet]. 2019 Dec
317 11 [cited 2019 Dec 14]; Available from:
318 <https://jamanetwork.com/journals/jamasurgery/fullarticle/2756311>

319 5. Hart TC, Hart PS. Genetic studies of craniofacial anomalies: clinical
320 implications and applications. *Orthod Craniofac Res*. 2009 Aug;12(3):212–20.

321 6. Thevenot J, López MB, Hadid A. A Survey on Computer Vision for Assistive
322 Medical Diagnosis From Faces. *IEEE J Biomed Health Inform*. 2018 Sep;22(5):1497–
323 511.

324 7. Arksey H, O'Malley L. Scoping studies: towards a methodological framework.
325 *Int J Soc Res Methodol*. 2005 Feb 1;8(1):19–32.

326 8. Gurovich Y, Hanani Y, Bar O, Nadav G, Fleischer N, Gelbman D, et al.
327 Identifying facial phenotypes of genetic disorders using deep learning. *Nat Med*. 2019
328 Jan;25(1):60–4.

- 329 9. Saraydemir S, Taşpınar N, Eroğul O, Kayserili H, Dinçkan N. Down syndrome
330 diagnosis based on Gabor Wavelet Transform. *J Med Syst.* 2012 Oct;36(5):3205–13.
- 331 10. Zhao Q, Rosenbaum K, Okada K, Zand DJ, Sze R, Summar M, et al. Automated
332 down syndrome detection using facial photographs. In: 2013 35th Annual International
333 Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). 2013.
334 p. 3670–3.
- 335 11. Burçin K, Vasif NV. Down syndrome recognition using local binary patterns
336 and statistical evaluation of the system. *Expert Syst Appl.* 2011 Jul 1;38(7):8690–5.
- 337 12. Kruszka P, Porrás AR, Sobering AK, Ikolo FA, Qua SL, Shotelersuk V, et al.
338 Down syndrome in diverse populations. *Am J Med Genet A.* 2017;173(1):42–53.
- 339 14. Basel- Vanagaite L, Wolf L, Orin M, Larizza L, Gervasini C, Krantz ID, et al.
340 Recognition of the Cornelia de Lange syndrome phenotype with facial dysmorphology
341 novel analysis. *Clin Genet.* 2016;89(5):557–63.
- 342 15. Dowsett L, Porrás AR, Kruszka P, Davis B, Hu T, Honey E, et al. Cornelia de
343 Lange syndrome in diverse populations. *Am J Med Genet A.* 2019;179(2):150–8.
- 344 16. Dalal AB, Phadke SR. Morphometric analysis of face in dysmorphology.
345 *Comput Methods Programs Biomed.* 2007 Feb 1;85(2):165–72.
- 346 17. Kruszka P, Addissie YA, McGinn DE, Porrás AR, Biggs E, Share M, et al.
347 22q11.2 Deletion Syndrome in Diverse Populations. *Am J Med Genet A.* 2017
348 Apr;173(4):879–88.
- 349 18. Schaefer GB, Rosenbloom AL, Guevara-Aguirre J, Campbell EA, Ullrich F,
350 Patil K, et al. Facial morphometry of Ecuadorian patients with growth hormone receptor
351 deficiency/Laron syndrome. *J Med Genet.* 1994 Aug;31(8):635–9.
- 352 19. Kruszka P, Porrás AR, de Souza DH, Moresco A, Huckstadt V, Gill AD, et al.

353 Williams-Beuren Syndrome in Diverse Populations. *Am J Med Genet A*. 2018
354 May;176(5):1128–36.

355 20. Miller RE, Learned- Miller EG, Trainer P, Paisley A, Blanz V. Early diagnosis
356 of acromegaly: computers vs clinicians. *Clin Endocrinol (Oxf)*. 2011;75(2):226–31.

357 21. Kong X, Gong S, Su L, Howard N, Kong Y. Automatic Detection of
358 Acromegaly From Facial Photographs Using Machine Learning Methods.
359 *EBioMedicine*. 2017 Dec 15;27:94–102.

360 22. Schneider HJ, Kosilek RP, Günther M, Roemmler J, Stalla GK, Sievers C, et al.
361 A Novel Approach to the Detection of Acromegaly: Accuracy of Diagnosis by
362 Automatic Face Classification. *J Clin Endocrinol Metab*. 2011 Jul 1;96(7):2074–80.

363 23. Kosilek RP, Frohner R, Würtz RP, Berr CM, Schopohl J, Reincke M, et al.
364 Diagnostic use of facial image analysis software in endocrine and genetic disorders:
365 review, current results and future perspectives. *Eur J Endocrinol*. 2015 Oct
366 1;173(4):M39–44.

367 24. Learned-Miller E, Lu Q, Paisley A, Trainer P, Blanz V, Dedden K, et al.
368 Detecting acromegaly: screening for disease with a morphable model. *Med Image*
369 *Comput Comput-Assist Interv MICCAI Int Conf Med Image Comput Comput-Assist*
370 *Interv*. 2006;9(Pt 2):495–503.

371 25. Kosilek RP, Schopohl J, Grunke M, Reincke M, Dimopoulou C, Stalla GK, et al.
372 Automatic Face Classification of Cushing’s Syndrome in Women – A Novel Screening
373 Approach. *Exp Clin Endocrinol Diabetes*. 2013 Oct;121(9):561–4.

374 26. Chen S, Pan Z, Zhu H, Wang Q, Yang J-J, Lei Y, et al. Development of a
375 computer-aided tool for the pattern recognition of facial features in diagnosing Turner
376 syndrome: comparison of diagnostic accuracy with clinical workers. *Sci Rep [Internet]*.

377 2018 [cited 2019 Dec 9];8. Available from:
378 <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6006259/>

379 27. Chen Y, Liu W, Zhang L, Yan M, Zeng Y. Hybrid facial image feature
380 extraction and recognition for non-invasive chronic fatigue syndrome diagnosis.
381 *Comput Biol Med.* 2015 Sep 1;64:30–9.

382 28. Buretic-Tomljanovic A, Petaros A, Suc E, Nincoletti S, Rubeša G, Tomljanović
383 D, et al. Craniofacial morphologic and anthropometric features of Croatian
384 schizophrenia patients and non-psychiatric controls - A pilot study. *Anthropol Anz Ber*
385 *Über Biol-Anthropol Lit.* 2012 Nov 1;69:379–97.

386 29. Ferry Q, Steinberg J, Webber C, FitzPatrick DR, Ponting CP, Zisserman A, et al.
387 Diagnostically relevant facial gestalt information from ordinary photos. *eLife* [Internet].
388 2014 [cited 2019 Dec 9];3. Available from:
389 <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4067075/>

390 30. Boehringer S, Vollmar T, Tasse C, Wurtz RP, Gillessen-Kaesbach G,
391 Horsthemke B, et al. Syndrome identification based on 2D analysis software. *Eur J*
392 *Hum Genet.* 2006 Oct;14(10):1082–9.

393 31. Shukla P, Gupta T, Saini A, Singh P, Balasubramanian R. A Deep Learning
394 Frame-Work for Recognizing Developmental Disorders. In: 2017 IEEE Winter
395 Conference on Applications of Computer Vision (WACV). 2017. p. 705–14.

396 32. Boehringer S, Guenther M, Sinigerova S, Wurtz RP, Horsthemke B, Wieczorek
397 D. Automated syndrome detection in a set of clinical facial photographs. *Am J Med*
398 *Genet A.* 2011;155(9):2161–9.

399 33. Loos HS, Wieczorek D, Würtz RP, Malsburg C von der, Horsthemke B.
400 Computer-based recognition of dysmorphic faces. *Eur J Hum Genet.* 2003

401 Aug;11(8):555–60.

402 34. Vollmar T, Maus B, Wurtz RP, Gillessen-Kaesbach G, Horsthemke B,
403 Wieczorek D, et al. Impact of geometry and viewing angle on classification accuracy of
404 2D based analysis of dysmorphic faces. *Eur J Med Genet.* 2008 Jan 1;51(1):44–53.

405 35. Cerrolaza JJ, Porras AR, Mansoor A, Zhao Q, Summar M, Linguraru MG.
406 Identification of dysmorphic syndromes using landmark-specific local texture
407 descriptors. In: 2016 IEEE 13th International Symposium on Biomedical Imaging
408 (ISBI). 2016. p. 1080–3.

409 36. Özdemir ME, Telatar Z, Eroğul O, Tunca Y. Classifying dysmorphic syndromes
410 by using artificial neural network based hierarchical decision tree. *Australas Phys Eng
411 Sci Med.* 2018 Jun 1;41(2):451–61.

412 37. Dudding-Byth T, Baxter A, Holliday EG, Hackett A, O'Donnell S, White SM, et
413 al. Computer face-matching technology using two-dimensional photographs accurately
414 matches the facial gestalt of unrelated individuals with the same syndromic form of
415 intellectual disability. *BMC Biotechnol.* 2017 19;17(1):90.

416 38. Kuru K, Niranjan M, Tunca Y, Osvank E, Azim T. Biomedical visual data
417 analysis to build an intelligent diagnostic decision support system in medical genetics.
418 *Artif Intell Med.* 2014 Oct 1;62(2):105–18.

419 39. Knaus A, Pantel JT, Pendziwiat M, Hajjir N, Zhao M, Hsieh T-C, et al.
420 Characterization of glycosylphosphatidylinositol biosynthesis defects by clinical
421 features, flow cytometry, and automated image analysis. *Genome Med.* 2018
422 09;10(1):3.

423 40. Kruszka P, Porras AR, Addissie YA, Moresco A, Medrano S, Mok GT, et al.
424 Noonan Syndrome in Diverse Populations. *Am J Med Genet A.* 2017 Sep;173(9):2323–

- 425 34.
- 426 41. Shu T, Zhang B. Non-invasive Health Status Detection System Using Gabor
427 Filters Based on Facial Block Texture Features. *J Med Syst*. 2015 Feb 27;39(4):41.
- 428 42. Valentine M, Bihm DCJ, Wolf L, Hoyme HE, May PA, Buckley D, et al.
429 Computer-Aided Recognition of Facial Attributes for Fetal Alcohol Spectrum
430 Disorders. *Pediatrics* [Internet]. 2017 Dec 1 [cited 2019 Dec 9];140(6). Available from:
431 <https://pediatrics.aappublications.org/content/140/6/e20162028>
- 432 43. Kuru K, Niranjana M, Tunca Y, Osvank E, Azim T. Biomedical visual data
433 analysis to build an intelligent diagnostic decision support system in medical genetics.
434 *Artif Intell Med*. 2014 Oct 1;62(2):105–18.
- 435 44. Viola P, Jones M. Rapid object detection using a boosted cascade of simple
436 features. In: *Proceedings of the 2001 IEEE Computer Society Conference on Computer
437 Vision and Pattern Recognition CVPR 2001* [Internet]. Kauai, HI, USA: IEEE Comput.
438 Soc; 2001 [cited 2019 Dec 9]. p. I-511-I-518. Available from:
439 <http://ieeexplore.ieee.org/document/990517/>
- 440 47. Cootes TF, Edwards GJ, Taylor CJ. Active appearance models. In: Burkhardt H,
441 Neumann B, editors. *Computer Vision — ECCV'98*. Berlin, Heidelberg: Springer;
442 1998. p. 484–98. (Lecture Notes in Computer Science).
- 443 48. Ashraf AB, Lucey S, Cohn JF, Chen T, Ambadar Z, Prkachin KM, et al. The
444 Painful Face – Pain Expression Recognition Using Active Appearance Models. *Image
445 Vis Comput*. 2009 Oct;27(12):1788–96.
- 446 49. Rohlf F, Slice D. Extensions of the Procrustes Method for the Optimal
447 Superimposition of Landmarks. *Syst Zool*. 1990 Mar 1;39.
- 448 50. Rohlf FJ. Bias and error in estimates of mean shape in geometric

449 morphometrics. *J Hum Evol.* 2003 Jun 1;44(6):665–83.

450 51. Peters JR, Campbell RM, Balasubramanian S. Characterization of the age-
451 dependent shape of the pediatric thoracic spine and vertebrae using generalized
452 procrustes analysis. *J Biomech.* 2017 Oct 3;63:32–40.

453 52. Douglas TS, Mutsvangwa TEM. A review of facial image analysis for
454 delineation of the facial phenotype associated with fetal alcohol syndrome. *Am J Med*
455 *Genet A.* 2010;152A(2):528–36.

456 53. Liehr T, Acquarola N, Pyle K, St- Pierre S, Rinholm M, Bar O, et al. Next
457 generation phenotyping in Emanuel and Pallister-Killian syndrome using computer-
458 aided facial dysmorphology analysis of 2D photos. *Clin Genet.* 2018;93(2):378–81.

459 54. Shen L, Bai L. A review on Gabor wavelets for face recognition. *Pattern Anal*
460 *Appl.* 2006 Oct 1;9(2):273–92.

461 55. Wiskott L, von der Malsburg C. Recognizing Faces by Dynamic Link Matching.
462 *NeuroImage.* 1996 Dec 1;4(3):S14–8.

463 56. Ojala T, Pietikainen M, Maenpaa T. Multiresolution gray-scale and rotation
464 invariant texture classification with local binary patterns. *IEEE Trans Pattern Anal*
465 *Mach Intell.* 2002 Jul;24(7):971–87.

466 57. Gumus E, Kilic N, Sertbas A, Ucan ON. Evaluation of face recognition
467 techniques using PCA, wavelets and SVM. *Expert Syst Appl.* 2010 Sep 1;37(9):6404–8.

468 58. Vapnik VN. *Statistical learning theory.* New York: Wiley; 1998. 736 p.
469 (Adaptive and learning systems for signal processing, communications, and control).

470 59. Weinberger KQ, Saul LK. Distance Metric Learning for Large Margin Nearest
471 Neighbor Classification. :38.

472 60. Li H, Lin Z, Shen X, Brandt J, Hua G. A convolutional neural network cascade

473 for face detection. In: 2015 IEEE Conference on Computer Vision and Pattern
474 Recognition (CVPR) [Internet]. Boston, MA, USA: IEEE; 2015 [cited 2019 Dec 9]. p.
475 5325–34. Available from: <http://ieeexplore.ieee.org/document/7299170/>
476

477 **FIGURES AND TABLE CAPTIONS**

478

479 **Table 1.** Pubmed request.

*(diagnosis[Title] OR detection[Title] OR dysmorphic[Title/Abstract] OR
dysmorphology[Title/Abstract] OR disease[Title/Abstract] OR syndrome[Title/Abstract])*

AND *(face[Title/Abstract] OR faces[Title/Abstract] OR facial[Title/Abstract])*

AND *("face-classification"[Title/Abstract] OR "facial analysis technology"[Title/Abstract] OR
"artificial intelligence"[MeSH Terms] OR "biometric identification"[Mesh] OR "Medical
informatics"[MeSH Terms] OR "Image Processing, Computer-Assisted"[MeSH Terms])*

480

481

482 **Table 2.** General characteristics for each of the studies, and method of classification.

483 Publications not using machine learning (n = 3) were excluded. SVM = Support Vector

484 Machine. kNN = k-Nearest Neighbors. RF = Random forests. CNN = Convolutional

485 Neural Network. AAN = Artificial Neural Networks.

486

Multisyndrome	Data volume	Number of syndromes	Linear functions	SVM	kNN	RF	CNN / ANN
Boehringer, 2006	147	10		✓	✓		
Vollmar, 2007	200	14		✓	✓		
Boehringer, 2011	202	14		✓	✓		
Ferry, 2014	1363	8			✓		
Kuru, 2014	92	15			✓		
Shu, 2015	13	100		✓	✓		
Özdemir, 2018	124	5			✓		✓
Gurovich, 2019	17,106	216					✓
Monosyndrome	Condition / syndrome						
Schaefer, 1994	49	Laron		✓			

Learned-Miller, 2006	24	Acromegaly	✓			
Miller, 2011	24	Acromegaly	✓			
Saraydemir, 2011	15	Down	✓	✓		
Schneider, 2011	57	Acromegaly	✓			
Buretic-Tomljanovic, 2012	58	Schizophrenia	✓			
Kosilek, 2013	20	Cushing	✓			
Zhao, 2014	100	Down	✓	✓	✓	
Chen, 2015	294	Chronic fatigue syndrome		✓		
Kong, 2017	527	Acromegaly	✓	✓	✓	✓
Kruszka, 2017	156	22q11.2	✓			
Kruszka, 2017	129	Down	✓			
Kruszka, 2017	161	Cornelia de Lange	✓			
Chen, 2018	32	Turner	✓			
Kruszka, 2018	286	Williams-Beuren	✓			
Dowsett, 2019	246	Cornelia de Lange	✓			

487

488