bioRxiv preprint doi: https://doi.org/10.1101/2022.12.27.521999; this version posted December 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

A 3D Clinical Face Phenotype Space of Genetic Syndromes using a Triplet-Based Singular Geometric Autoencoder

Soha S. Mahdi, Harold Matthews, Michiel Vanneste, Nele Nauwelaers, Shunwang Gong, Giorgos Bouritsas, Gareth S Baynam, Peter Hammond, Richard Spritz, Ophir D Klein, Benedikt Hallgrimsson, Hilde Peeters, Peter Claes

***Abstract*— Clinical diagnosis of syndromes benefits strongly from objective facial phenotyping. This study investigates facial dysmorphism of genetic syndromes by building and investigating a low-dimensional metric space referred to as the clinical face phenotypic space (CFPS). As a facial matching tool for clinical genetics, such CFPS can enhance clinical diagnosis. It helps to interpret facial dysmorphisms of a subject by placing them within the space of known dysmorphisms. In this paper, a triplet loss-based autoencoder developed by geometric deep learning (GDL) is trained using multi-task learning, which combines supervised and unsupervised learning approaches. Experiments are designed to illustrate the following properties of CFPSs that can aid clinicians in narrowing down their search space: A CFPS can 1) classify and cluster syndromes accurately, 2) generalize to novel syndromes, and 3) preserve the relatedness of genetic diseases, meaning that clusters of phenotypically similar disorders reflect functional relationships between genes. This model is composed of three main components: 1) an encoder based on GDL that optimize distances between individuals in the CFPS therefore adding to the classifier's power. 2) a decoder that improves both classification and clustering performance by reconstructing a face from an embedding in a CFPS, 3) a singular value decomposition layer to maintain orthogonality and optimal variance distribution across dimensions. This allows for the selection of an optimal number of CFPS dimensions as well as improving the classification, reconstruction, and generalization capabilities of the CFPS.**

This work was supported by the National Institutes of Health (1-R01-DE027023), The Research Fund KU Leuven (BOF-C1, C14/20/081) and The Research Program of the Research Foundation - Flanders (Belgium) (FWO, G078518N).

S. S. Mahdi is with the Vrije Universiteit Brussel, ETRO. Building Ke. 3.15, Bd du Triomphe 26, 1050 Ixelles, Belgium (e-mail: soha.sadat.mahdi@vub.be)

H. Matthews, N.Nauwelaers, and P. Claes are with the KU Leuven, ESAT/PSI - UZ Leuven, MIRC. UZ Herestraat 49 - box 7003, 3000 Leuven, Belgium (e-mail: harry.matthews@kuleuven.be, nele.nauwelaers@kuleuven.be, peter.claes@kuleuven.be).

M. Vanneste, P. Hammond and H. Peeters are with KU Leuven, Department of Human Genetics. Laboratory for Genetic Epidemiology, ON I Herestraat 49 - box 606, 3000 Leuven, Belgium (e-mail: michiel.vanneste@kuleuven.be, watervole@icloud.com, hilde.peeters@uzleuven.be). H. Matthews and P. Claes are also affiliated with this department.

S. Gong and G. Bouritsas are with the Imperial College London, Department of Computing. 569, Huxley Building, South Kensington Campus, London, United Kingdom (e-mail: shunwang.gong16@imperial.ac.uk, g.bouritsas@imperial.ac.uk).

G. Baynam is with the Curtin University, School of Earth and Planetary Sciences, Faculty of Science and Engineering, Curtin University, Perth, Western Australia, Australia, and also with the Western Australian Register of Developmental Anomalies, King Edward Memorial Hospital, Perth, Western Australia, Australia (email: Gareth.Baynam@health.wa.gov.au)

R. Spritz is with the University of Colorado, School of Medicine, Human Medical Genetics and Genomics Program, Aurora, Colorado, United States (email: richard.spritz@cuanschutz.edu)

O. D. Klein is with the University of California, San Francisco, Departments of Orofacial Sciences and Pediatrics, and Institute for Human Genetics, San Francisco, United States (email: Ophir.Klein@ucsf.edu)

B. Hallgrimsson is with the Cumming School of Medicine, Department of Cell Biology and Anatomy, and Alberta Children's Hospital Research Institute, University of Calgary, Calgary, Alberta, Canada (email: bhallgri@ucalgary.ca)

***Index Terms*— Clinical Genetics, Computer-aided Diagnosis, Deep Phenotyping, 3D Shape Analysis, Geometric Deep Learning, Precision Public Health.**

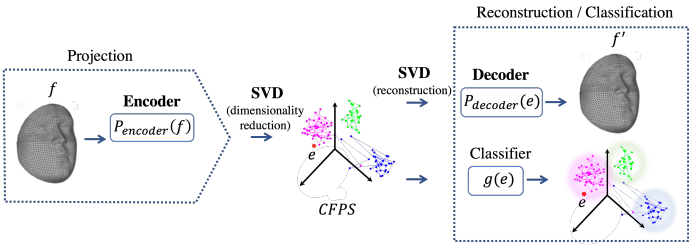


Fig. 1: The complete model consists of three main components: a triplet-based encoder, a singular value decomposition (SVD) layer, and a decoder. Projection function pSGE for geometric model (alternatively p_{LDA} for baseline) projects a facial mesh f into a facial embedding e in the CFPS. A facial mesh f' is reconstructed from the embedding e with decoding function p_{SGD}. Note that the reconstruction is not possible within the baseline. Classification from the embedding space into syndrome groups is performed by a classification function g, which in this work constitutes a simple K-nearest-neighbor classifier.

I. Introduction

Many genetic syndromes are associated with mild to severe facial dysmorphism. Observations from studies on the human phenome suggest that similar phenotypes are related to functionally related genes [1]. As first introduced by Ferry et al. [2], facial dysmorphisms can be modeled in a lower dimensional latent space or metric space, known as a Clinical Face Phenotype Space (CFPS), using metric learning techniques. A CFPS has three main properties. First, distances in a CFPS are a measure of phenotypic similarity, and patients' faces are clustered based on diagnostically relevant phenotypic features. Second, a well-trained CFPS generalizes to dysmorphic syndromes that were not used in the training, and therefore it can be used to explore or cluster novel syndromes or smaller groups of similar patients [3]. Third, a CFPS recapitulates known relationships of genetic diseases, meaning that clusters of phenotypically similar disorders reflect functional relationships among the genes involved [3]. With these properties, individuals without a confirmed diagnosis can be rapidly compared to each other and to known phenotypic groupings in the space, positioning them in the spectrum of known relationships between the phenotype and clinical and molecular diagnosis. Thus, the space can be used to propose hypothetical clinical and molecular diagnoses (e.g., by classification from embedding coordinates). This can facilitate variant interpretation in a genome wide NGS analysis or direct towards targeted sequencing as part of a clinical diagnostic workup [4].

bioRxiv preprint doi: https://doi.org/10.1101/2022.12.27.521999; this version posted December 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Multiple challenges are involved in constructing a CFPS based on metric learning or classification. On the one hand, there is considerable overlap between different syndromes with the same or similar dysmorphic features being present in related [5] and ostensibly unrelated disorders. For example hypertelorism is a feature of both Apert and Wolf-Hirschhorn syndrome [6], [7]. On the other hand, there can be substantial phenotypic variation within a particular syndrome [1], [8]–[10]. Therefore, encoding facial shape into a metric space with a minimum overlap between syndromes and small variations within syndromes is a complicated task. To tackle this challenge, the first component of our proposed model is a supervised (deep) metric learner based on triplet loss function that optimize such a CFPS directly to discriminate different groups by learning the between-group dissimilarities and within-group similarities. However, by solely focusing on discriminating clues, information about general facial topology is overlooked. Therefore, the second challenge is encoding general facial similarity irrespective of the syndrome group. To address this issue, we aim to simultaneously preserve the facial topology in the space: following unsupervised dimensionality reduction techniques such as autoencoders, we combine the metric learning encoder with a facial decoder as the second component of our model. In essence, facial structure and similarity are learned by the combination of supervised (metric learning) and unsupervised (dimensionality reduction) learning paradigms. The intention is that, in combination, this should yield a space that prioritizes diagnostically relevant facial features (by virtue of the metric learning) and also can meaningfully represent inter-patient variation in general (by virtue of the decoder). This latter property is essential to meaningfully encode patients from unseen syndromes. Another advantage of the decoder is the ability to reconstruct a face from a sampled embedding from the CFPS. However, since no structure is imposed on the space during training such encoder-decoder, sampling from the resulting latent space is not reliably done. To facilitate sampling, we incorporate the third component, a singular value decomposition (SVD) layer, with which the dimensions of the CFPS are transformed to an orthonormal basis. This SVD layer was first introduced in [11] to impose orthogonality and optimal distribution of variance across dimensions of the latent space learned by an autoencoder, with their extra power in modeling non-linearity and performing additional tasks.

Related works fall into one or both of the following categories: 1) syndrome classification and 2) building and/or experimenting with clinical face phenotype spaces (CFPSs) . The majority of approaches for syndrome classification use 2D photographs (recent reviews are available in [12]–[14]) because of their availability. The most popular 2D classification tool among clinicians today is Face2Gene (DeepGestalt), introduced by the company Facial Dysmorphology Novel Analysis (FDNA). This technology comprises multiple deep convolutional neural networks (DCNNs) designed for classification, each applied to and specialized for different facial regions. The outputs of all specialized classifiers are combined to give a ranked list of candidate disorders [15]. The tool is supported by a large and diverse training data set, which is continuously expanded as it is contributed to by users and currently supports the identification of approximately 300 disorders [16]. However, as opposed to 3D images, 2D images only encode information about 3D shapes indirectly. The importance of 3D photographs has been emphasized in the field [17]–[20] and they are becoming more popular with the increased accessibility of 3D imaging devices [21]. The recent acquisition of large-scale 3D image datasets of participants with genetic syndromes [22] have allowed some attempts at large-scale learning from 3D photographs. Hallgrimsson et al. [23] classified 63 genetic syndromes using linear classifiers and sparse anatomical landmarks as features. Bannister et al. [24] classified 47 genetic syndromes from facial surface scans using a normalizing flow architecture. Mahdi et al. [25] introduced a multi-scale part-based metric learner for classification of 14 syndromes form surface scans represented by $\sim$8000 landmarks. The concept of CFPS was first introduced in [2] where principal component (PC) scores, representing variation in landmark coordinates and local pixel intensity variation around the landmarks, were derived from 2D images of 8 disorders and subjected to the linear largest margin nearest neighbor metric learner. They investigated the relationship between phenotypic distance in the space and genetic distance (based on protein-to-protein interaction networks) and found the magnitude of the relationship to be non-zero, demonstrating to some extent the space recapitulates the functional relationships among the involved genes. The largest CFPS to date is 'GestaltMatcher' [16], which utilizes the same underlying architecture as Face2Gene but interprets the extracted feature representations directly as a location in a CFPS. Until recently the limited availability of 3D images has hampered work in 3D. To the best of our knowledge the only other 3D CFPS [24], used an invertible normalizing flow model to produce the first non-gaussian 3D facial CFPS. They report mean sensitivity of 43% across 48 syndromes as well as generating modal, randomly sampled, and counterfactual 3D faces using demographic information. Alternatively, in this work, we aim to learn a CFPS from dense 3D meshes of patients from 51 syndrome groups and a group of controls. In contrast to Bannister et al., we use spiral convolutional operators to allow learning directly from 3D data [26], [27], instead of first transforming the data into principal component scores. We first develop a triplet loss-based encoder to learn a non-linear low dimensional metric in a supervised manner [28], [29], and we further extend this model by attaching a decoder block and an SVD layer to the network to enable unsupervised facial reconstruction and orthonormal basis. We validate the properties of the CFPS by first, investigating the clustering performance of the space using classification and clustering metrics. We show that the components added to the encoder both contribute to the classification performance. Second, we show the capacity of the decoder to generate realistic-looking synthetic faces by generating faces at the center of syndrome clusters whereby it can be visually assessed if they display the known facial gestalt of certain disorders. Extending the validation efforts of previous CFPSs, we implement and execute an extensive test of the generalization to novel syndromes in which a CFPS is trained based on a subset of syndromes, and then the clustering characteristics of the projection of the left-out syndromes are assessed and compared with existing 3D linear metric learning techniques. We prove a better generalization of our complete model compared to the linear baseline and also compared to the models without the SVD layer. We also assess whether classification and clustering performance conforms to prior clinical knowledge about the presence of facial dysmorphism and we bring evidence for the recapitulation of the phenotypic relationship among related genetic diseases.

II. Materials and Methods

A. Data

3D facial images of the dataset used in this project were sourced from:

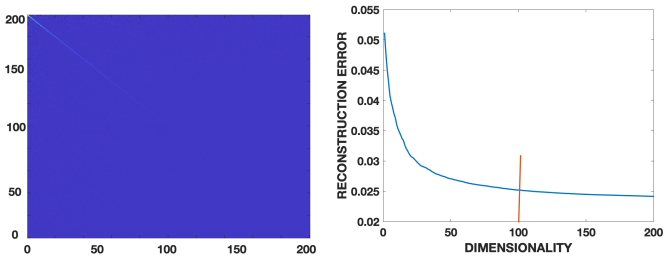
1) The FaceBase repository[1], "Developing 3D Craniofacial Morphometry Data and Tools to Transform Dysmorphology, FB00000861", collected at patient support groups in the USA, Canada, and the UK [22], [23].
2) The Western Australian Health Department. This collection is from the database of the Health Department of Western

[1] www.facebase.org

bioRxiv preprint doi: https://doi.org/10.1101/2022.12.27.521999; this version posted December 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

TABLE I: Data demographics: Syndrome group name, sample size (N), mean and standard deviation of age (M±SD), the female/male ratio (F/M), and the category .

Name	Size	Age Range	Sex Ratio	Category	Name	Size	Age Range	Sex Ratio	Category
Williams	221	17.57 ± 13.9	0.46	A	BBS	87	26.33 ± 14.78	0.48	C
22q11.2 Del	180	10.74 ± 6.03	0.49	A	Neurofibromatosis	85	20.18 ± 18.01	0.54	C
Wolf Hirschhorn	155	11.03 ± 9.42	0.57	A	Loeys Dietz	84	25.38 ± 17.15	0.57	C
Smith Magenis	129	14.32 ± 9.09	0.55	A	Joubert	75	10.57 ± 8.58	0.48	C
Down	117	21.64 ± 11.14	0.49	A	Ectodermal Dysplasia	71	15.09 ± 15.32	0.28	C
Prader Willi	96	19.34 ± 13.24	0.51	A	Rett	70	13.32 ± 10.54	0.89	C
Fragile X	77	17.65 ± 12.56	0.3	A	Cardiofaciocutaneous	59	12.21 ± 8.55	0.53	C
Achondroplasia	70	22.62 ± 18.34	0.59	A	Klinefelter	57	22.91 ± 14.58	0	C
Rubinstein Taybi	63	13.54 ± 11.73	0.52	A	Mucopolysaccharidosis	57	21.51 ± 13.51	0.47	C
Costello	58	12.39 ± 9.24	0.66	A	Alstrom	52	21.28 ± 9.4	0.54	C
Cohen	33	18.27 ± 10.46	0.52	A	Fibrodysplasia Ossificans Progressiva	50	21.81 ± 12.37	0.56	C
Pitt Hopkins	29	8.53 ± 5.7	0.62	A	Fabry	48	32.37 ± 16.53	0.44	C
Pallister Killian	23	9.59 ± 7.19	0.26	A	Sotos	45	17.92 ± 12.22	0.49	C
Crouzon	22	10.22 ± 6.27	0.55	A	Russell Silver	44	10.18 ± 10.32	0.34	C
Smith Lemli Opitz	19	11.75 ± 7.05	0.32	A	Cockayne	41	12.15 ± 7.37	0.44	C
Apert	13	14.55 ± 10.73	0.62	A	Pseudoachondroplasia	35	28.06 ± 20.53	0.51	C
Coffin Lowry	12	13.76 ± 9.16	0.08	A	Osteogenesis Imperfecta	31	16.72 ± 14.61	0.68	C
Cornelia de Lange	183	12.1 ± 9.14	0.54	B	1p36 Del	29	8.82 ± 7.83	0.62	C
Noonan	155	14.06 ± 12.51	0.45	B	Trisomy 18	27	8.79 ± 8.71	0.85	C
Angelman	106	9.97 ± 7.57	0.47	B	Beckwith Wiedemann	26	9.68 ± 6.66	0.42	C
Stickler	45	22.31 ± 17.45	0.62	B	EED CLP	20	23.24 ± 17.71	0.65	C
Treacher Collins	39	18.48 ± 13.5	0.49	B	Vander Woude	16	10.16 ± 4.75	0.56	C
Kabuki	37	12.09 ± 6.62	0.65	B	Goltz	14	9.4 ± 5.03	0.86	C
Coffin Siris	16	12.08 ± 9.65	0.63	B	Rhizo Chondro Punct	13	7.57 ± 5.53	0.69	C
Marfan	153	26.34 ± 16.85	0.58	C	Zellweger Syndrome	11	7.33 ± 9.49	0.09	C
Turner	102	24.25 ± 19.03	0.98	C	Controls	100	30.94 ± 11.64	0.72	CONTROL



(a) (b)

Fig. 2: (a) Covariance matrix of the scores for the 150-dimensional space learned by the triplet-based singular autoencoder. (b) Reconstruction error as a function of dimensionality of the space learned by singular autoencoder. The reconstruction error reported here is based on the normalized faces of the training data.

Australia. Images were collected between 2009 and 2018, and were recruited primarily through the Genetic Services of Western Australia, but also at complementary sites including Australian hospitals and patient support groups. [30]

3) Peter Hammond's legacy 3D dysmorphology dataset hosted at the KU Leuven, Belgium. Patients were recruited at patient support groups across the United States, UK and Italy between 2002 and 2013. At initial recruitment, diagnosis was as reported by families and/or suggested by clinical geneticists attending the meetings; some patients were in contact over several years and molecular diagnoses were reported by parents or by collaborating clinical geneticists. [31]

From these three collections combined, groups with a minimum of 10 individuals were selected. Approximately, 59%, 40%, and <1% of the data used in this work are collected by the first, second, and the third listed source respectively. In total, the dataset comprised 3,285 3D facial images of 51 different syndromes and one group of 138 control individuals that are unrelated to the patients with known genetic syndromes. Demographic characteristics of the dataset are given in Table I. According to clinical knowledge and based on clinical assessment of the available images, the syndromic groups were assigned to one of the following categories by two clinical experts (co-authors HP and MV):

(a) Genetic conditions that can be diagnosed based on typical facial characteristics and that are genetically homogeneous, i.e. they are caused by one single gene or recurrent chromosomal anomaly.

(b) Genetic conditions that can be diagnosed based on typical facial characteristics and that are to some extend genetically heterogeneous, i.e. more than one gene for this clinical condition is known.

(c) Genetic conditions that are usually not diagnosed based on facial features, i.e. for these conditions facial features are not typical.

For syndromes in categories A and B, the facial features direct the clinician towards the molecular diagnosis. So, we expect subjects within these groups to have a distinct facial phenotype in general. However, in contrast to A, the genetic heterogeneity in disorders of category B introduces uncertainty to the relationship between the typical face and the underlying gene. In practical terms, based on known genotype-phenotype correlations these disorders may be more phenotypically diverse than A. Category C includes syndromes that in clinical practice are not diagnosed based on the facial features but based on other clinical symptoms. However, for these syndromes, clinicians do not claim that there is no recognizable gestalt. Therefore, in contrast to A and B, the presence of a distinctive facial gestalt is unclear, but not necessarily absent. This study was approved by the ethical review board of KU Leuven and University Hospitals Gasthuisberg, Leuven (S56392, S60568).

B. Preprocessing

For pre-processing, after cleaning the raw image by removing hair and ears, a 3D face template was non-rigidly registered to each face using Meshmonk [21]. Each 3D face shape is therefore described as a manifold triangle mesh $F = (V, \mathcal{E}, \Phi)$, where $V = \{v_i\}_i = 1^{8,321}$ is a $8,321 \times 3$ dimensional matrix, containing 8,321 3D vertices $v_i = (x_i, y_i, z_i)$ defining the mesh geometry, $\mathcal{E}$ and Φ are set of

bioRxiv preprint doi: https://doi.org/10.1101/2022.12.27.521999; this version posted December 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

edges and faces which define the mesh topology. $\mathcal{E}$ and Φ are fixed Since all our meshes have the same topology as the template.

C. Pipeline Design

As a first and the baseline approach towards learning a CFPS, we used Linear Discriminant Analysis (LDA), which is a supervised linear metric learner and classifier. Alternatively, we proposed a multi-component metric learner based on geometric deep models that consist of the following components: 1) a triplet-loss encoder, 2) a decoder, and 3) an SVD layer. Fig. 1 shows the main components constructing the model. A triplet-based encoder is a supervised deep-metric learner that relates individuals in terms of group membership and consists of three identical encoder networks. Such encoders are trained with triplets of the data comprising an anchor (f_a), positive (f_p), and negative sample (f_n). In each triplet, the anchor and positive samples are from the same class, while the anchor and negative samples are from different classes. The output of the network for a given triplet is a lower dimensional embedding of each element of the triplet $(e_a, e_p, e_n) = (p_{GE}(f_a), p_{GE}(f_p), p_{GE}(f_n))$. A decoder (p_{GD}) then reconstructs the image data from the embeddings $(f'_a, f'_p, f\prime_n) = (p_{GD}(e_a), p_{GD}(e_p), p_{GD}(e_n))$. To keep the power of this triplet-based autoencoder (encoder-decoder) in modeling non-linearity and at the same time to obtain orthogonal dimensions and optimal distribution of variance across different CFPS dimensions, an SVD layer was added. These three components were combined, and spiral-based geometric architectures were used to form our triplet-based singular geometric autoencoder (TB-SGAE).

1) CFPS based on Linear Discriminant Analysis (LDA): LDA maximizes the distance between the mean of all faces in each class (between-class scatter S_b) and minimizes the spreading within the class itself (within-group scatter S_w). Since our original meshes were densely sampled and had much higher dimensions thanare represented by many more variables than the number of training examples, we first applied principal component analysis (PCA) to the original meshes to reduce the dimensions as to avoid overfitting during LDA. The first 100 dimensions (preserving 99.16% of datavariation) were used as input to LDA. The lower-dimensional space projection for LDA was constructed by Fisher's criterion $\underset{a}{\operatorname{argmax}} \frac{|a^T S_b a|}{|a^T S_w a|}$, where a is considered as the lower-dimensional space projection matrix, also called Fisher's criterion. A projected embedding of a facial shape f in the resulting CFPS was then calculated as:

$$p_{LDA} : F \longrightarrow E, \quad p_{LDA}(f) = e_{PCA}.a \tag{1}$$

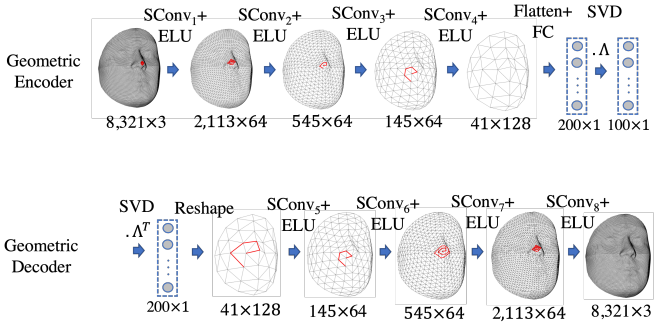


Fig. 3: The architecture of a singular geometric autoencoder (SGAE) with a singular value decomposition (SVD) layer. λ contains right singular vectors of the SVD. Once trained, the geometric encoder constitutes the projection function p_{SGE} , and the geometric decoder constitutes the decoding function p_{SGD}.

We took the maximum number of dimensions for LDA which is equal to the number of classes -1 (=51).

2) Triplet-Based Geometric Encoder (TB-GE): A geometric encoder (GE) refers to a 3D face to a CFPS encoder that uses spiral convolutional operators. In a triplet-based geometric encoder (TB-GE) the feature representations of patients within the same syndrome group are situated closer to each other than patients from a different syndrome group. Once trained, a GE learns a function p_{GE} that maps an input mesh $f \in F$ to a low dimensional embedding $e \in E$ in CFPS:

$$p_{GE} : F \longrightarrow E, \quad e = p_{GE}(f) \tag{2}$$

The triplet-loss loss function for training the TB-GE is:

$$t = max\left(\| e_a - e_p \|_2^2 - \| e_a - e_n \|_2^2 + \alpha, 0\right) \tag{3}$$

Where α is the margin between paired positive and negative samples of the triplet and following [32], it was set to 0.2. Changing this parameter did not significantly change the outcomes (data not shown).

3) Decoder: The geometric decoder (GD) function p_{GD} reconstructs a facial mesh f' from an encoder e:

$$p_{GD} : E \longrightarrow F, \quad f\prime = p_{GD}(e) \tag{4}$$

The geometric autoencoder (GAE) function can therefore be formulated as:

$$p_{GAE} : F \longrightarrow F, \quad f' = p_{GD}(p_{GE}(f)) \tag{5}$$

For 3D facial shapes as image data, the reconstruction loss function was the mean absolute error or the absolute difference between v_i, vertices of the input shape f, and v'_i, the corresponding vertices of the reconstructed output f', averaged over all data samples:

$$r_f = \frac{1}{N_{dataset}} \sum_{j=1}^{N_{dataset}} mean(D_j), \quad (D_j = mean|v'_i - v_i|_{i=1}^{8,321}) \tag{6}$$

For a triplet of anchor, positive, and negative within a batch, the reconstruction loss was therefore calculated as:

$$r = mean(r_a, r_p, r_n) \tag{7}$$

Where r_a, r_p, r_n are the reconstructions of the anchor, positive, and negative samples respectively. The final loss function used for training the GAE was:

$$l = r + \lambda \times t \tag{8}$$

Where λ is the weight given to the triplet loss t from 3 to regularize the scale of the two loss values. Increasing λ puts more emphasis on the triplet loss and less on the reconstruction loss and vice versa. With $\lambda = 1$, the triplet loss t was about 10 times larger than the reconstruction loss. Therefore, we rebalanced the total loss by setting $\lambda = 0.1$, so that both losses had an equal contribution. To train the network, data was provided in triplets selected by random triplet mining from all possible triplets within a training batch.

4) SVD layer: decorrelation of the CFPS dimensions: To ensure orthogonal dimensions, they were calculated from a set of low dimensional embeddings E by singular value decomposition of $E : US\Lambda^T = E$, where S is a diagonal matrix of singular values in descending order of magnitude, U contains left singular vectors and Λ contains right singular vectors. Individual facial embeddings with orthogonal dimensions were then calculated as part of the projection function p_{SGE} defined as:

$$p_{SGE} : F \longrightarrow E, \quad e = p_{GE}(f).\Lambda \tag{9}$$

The decoder and the autoencoder functions were then re-defined as:

$$p_{SGD} : E \longrightarrow F, \quad f' = .\ p_{GD}(\Lambda^T.e) \tag{10}$$

bioRxiv preprint doi: https://doi.org/10.1101/2022.12.27.521999; this version posted December 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

$$p_{SGAE} : F \longrightarrow F\prime, f' = \ p_{SGAE}(f) = p_{SGD}(p_{SGE}(f)) \quad (11)$$

This added SVD layer facilitated the selection of the number of CFPS dimensions which is the main bottleneck of the model and affects the capacity of the learned space. Therefore, it was important to define the smallest number of dimensions that was large enough for reconstructing syndromic and normal facial variation. To do so, we trained an autoencoder with the SVD layer using a reconstruction loss only and called it a singular autoencoder (SAE). An SAE is composed of an autoencoder with the last layer of the encoder and the first layer of the decoder replaced by a low-rank singular value decomposition [33]. The same architecture as our TB-SGAE encoders and a large dimensional latent channel (=200) was used. Training data included all syndromic and control groups. As observed in [11] the covariance matrix in Fig. 2a, the variance was distributed over the different dimensions in such a way that for any given number of dimensions, the explained amount of variance was maximized. Therefore, to select the proper dimension, we calculated and plotted the reconstruction loss as a function of the latent dimension in Fig. 2b. Based on this, 100 dimensions were close to saturation for the reconstruction of the original faces.

5) Spiral-based SAE Architecture : We used geometric deep learning to learn directly from the 3D facial meshes and efficiently leverage the underlying geometry by using spiral convolution operators [34]. The architecture of our GE is illustrated in Fig. 3. The spiral convolutional (Sconv) layer in this figure consists of first, convolving spirals on vertices of the mesh in the current layer, and second, down- or up-sampling the current mesh to obtain input for the next layer. Each Sconv layer is followed by an exponential linear unit (ELU). A spiral convolution is a filter consisting of learned weights w, which is applied to a sequence of neighborhood vertices. That means,

$$\forall v \in V \ , \ h'(v) = \sum w_i^T h(S_i(v)) \quad (12)$$

where $h(v)$ is the input representation of vertex v, $h'(v)$ the output representation, and $S_i(v)$ the i^th neighbor of v in the spiral [27]. The sequence was defined as a spiral around a central vertex, starting in an arbitrary direction and then proceeding in a counterclockwise direction until a fixed length was reached.

In a geometric encoder based on spiral convolutions, aside from the convolution operator, a pooling operator for meshes must be incorporated. Established mesh decimation techniques used in many geometric deep learning methods reduce the number of vertices such that a good approximation of the original shape remains, but they result in irregularly sampled meshes at different steps of resolution. In contrast we used a 3D mesh down and up-sampling scheme that retains the property of equidistant mesh sampling as defined in [35]. Starting from five initial points, the refinement is done with loop subdivision by splitting each triangular face of the mesh into four smaller triangles by connecting the midpoints of the edges. The last up-sampled mesh has 8,321 vertices and an average resolution of 2mm, meaning that the average edge length is 2mm. For our geometric encoder, the five highest levels of resolution (shown in Fig. 3) are kept, and their output is passed through the fully connected layers of our encoder. In-house experiments showed that other sampling schemes are equally effective and can be used instead. The number of spirals in each layer was chosen empirically based on the previous and related works [**?**], [35], as well as in other in-house projects where similar facial data structures are used. The length of the spiral filters was set to 19 for the first two layers with the highest resolution, and a length of 6 was chosen for the following layers with lower resolution. These choices were made such that for higher resolution meshes two-ring neighbors (=19 vertices) and for lower resolution meshes one-ring neighbors (=6 vertices) are covered by a spiral filter. Larger spiral lengths were initially tested for the first layers in a geometric autoencoder and no significant improvement in reconstruction performance was observed. A shorter spiral length, covering one-ring neighbors (9 vertices), was also tested for the first layers, and the difference in performance was not significant. Therefore, to decrease the computation cost, one can choose a spiral length of nine over 16. Since we have a fixed topology enforced on all faces, the spirals were determined only once, on the template mesh.

6) Classification: To take advantage of the CFPS as a classification tool, and to compare the CFPS obtained from SGAE with the one obtained from LDA as dimensionality reduction (baseline), a K-nearest neighbor (KNN) classifier with K=10 was applied to the projection of individuals into the CFPSs based on TB-SGAE and baseline (Equations 12 and 1 respectively). A KNN algorithm computes the distances between a test image and all the examples in the projected training set and then selects the K closest to the test image and votes for the most frequent label. We set K=10, since it is the minimum group size in our dataset.

7) Training: All models were trained on an NVIDIA GeForce RTX 2080 Ti, 64 GB RAM, with PyTroch 1.1.0. The Adam optimizer was used for 600 epochs with a batch size of 30 (limited by the maximum GPU memory), an initial learning rate of 1e-4 chosen based on experiments ran for a range of (1e-1,1e-8), and a decay rate of 0.99 was applied after each epoch.

D. Experiments

A core assumption underpinning the development of a CFPS is that they define a clinically meaningful and useful model of the variation within and among classes. Therefore, we designed a series of four experiments, investigating different and complementary aspects of the CFPSs obtained. The complete network is TB-SGAE, and the utility of each component was tested by comparing performance to performance of networks with specific components removed. These are TB-SGE without the decoder and TB-GAE without the SVD layer.

First, **experiment 1**, evaluated and compared clustering and classification capacity of our CFPSs and that of the linear baseline. A 5-fold cross-validation was performed. In each of five folds, 20% of data for each group separately was selected randomly and devoted to the test set, and the remaining 80% was used for training the CFPS followed by the classifier when needed. The classification performance was assesses as well as clustering assessed by the within-group variance (WGV), and between-group distances (BGD). The classification measures are:

- Sensitivity $= \frac{true\ positives}{true\ positives\ +\ false\ negatives}$, measuring how well the classifier can identify true positives,
- Specificity $= \frac{true\ negatives}{true\ negatives\ +\ false\ positives}$, measuring how well the classifier can identify true negatives,
- Balanced accuracy $= \frac{Sensitivity\ +\ Specificity}{2}$, the mean of sensitivity and specificity which is especially useful when the classes are imbalanced,
- Adjusted Rand Index, compares two categorizations of the data: one based on the true labels and one based on the labels predicted by the classifier.

WGV and BGD are two distance-based measures of the variance within the space. WGV is defined as the median embedding distance of all individuals in each group to the average embedding of the group and BGD is defined as the median of distances between the average embedding of a syndrome and all other syndromes in the data. Therefore, in a well-clustered space, lower WGV and higher BGD is expected. Generally, in comparison of the two methods, the

bioRxiv preprint doi: https://doi.org/10.1101/2022.12.27.521999; this version posted December 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

difference between these metrics in terms of statistical significance is measured using a paired two-tailed Wilcoxon signed rank test. To be able to compare BGD of different models, the distance scales of all embeddings are normalized for each model. We assessed whether categories A, B, and C relate to the classification performance of our CFPS, to assess if the space recapitulates clinical knowledge of the syndrome groups.

Furthermore, we expected that, in general, phenotypic characteristics being syndrome uniqueness (the median of distances between the average shape of a syndrome to all other average syndrome shapes), cohesion (the median shape distance, measured on landmarks, of all individuals in each group to the average shape of the group), and severity (the average shape distance between the subjects with a syndrome and the mean shape for controls [23]) should predict accuracy to a substantial degree, and this may also be impacted by sample size. Given the correlations among the phenotypic predictors, it is difficult to investigate their effects independently. To do so, predictors were combined into a single latent variable using a PLS regression of accuracy (in each space) onto the phenotypic predictors and sample size, with one latent component.

Second, **experiment 2**, investigated the ability to reconstruct faces from the CFPS using the training and out-of-fold (OOF) error of reconstruction. The training error is the same as the reconstruction loss and is computed based on the samples in the trainset. This error explains the ability of a model to efficiently capture shape variation in a compact representation for a given number of dimensions. The OOF error describes the model's capability to capture shape variance in unseen, or non-training data. Therefore, it is computed as the mean absolute error for all samples from the test set. To investigate the effect of SVD layer, the training and OOF error for a model without this layer (TB-GAE) were measured.

Third, **experiment 3**, investigated the ability to generalize into the clustering of syndrome groups not used for building the CFPS. For this purpose, six syndrome groups with various recognizable phenotypic features were left out during training. Once trained, individuals in these groups were projected to the space. Then, the BGD, WGV and clustering improvement factor (CIF) CIF was computed and compared between various spaces. The CIF, first introduced in [2], determines the improvement in clustering over randomly distributed faces, and therefore measures structure in the CFPS controlling for the composition of the database. Considering a syndrome with Np positive and N_n negative instances in the space, the CIF is defined as the expected rank (r) of nearest positive match under random ranking over observed average rank (r) of nearest positive match:

$$CIF = \frac{E(r)}{O(r)} \tag{13}$$

To measure the effect of each component, the results for the complete model (TB-SGAE) were compared with same architectures without the decoder (TBS-GE) and without the SVD layer (TB-GAE). Furthermore, to investigate the added value of metric learning, we made a comparison to an unsupervised and linear baseline being the spaces spanned by principal components only.

The syndrome groups that were left out for this experiment are:

- Apert syndrome is a condition caused by a single gene characterized by typical craniofacial dysmorphism present in all patients.These individuals are all expected to be significantly different from controls.
- Cockayne syndrome is a condition where the facial features are mainly determined by a progressive loss of facial fatty and muscular tissue. This means that more typical faces are expected in adolescents and adult patients.
- Stickler syndrome is genetically heterogenous and characterized by midfacial hypoplasia which can be mild and therefore sometimes hard to recognize or to distinguish from normal facial variation. Therefore the range of severity in Stickler is broader than for example in Apert syndrome which can always be recognized in every single patient. Stickler is expected to cluster, but probably not as tight as Apert however the patients with Stickler syndrome in our dataset belong to the more severe end of the spectrum of dysmorphism.
- Cohen syndrome is a condition caused by a single gene characterized by typical dysmorphic features that may be variable in severity.
- Fragile X syndrome is an X linked condition caused by a single gene characterized by typical dysmorphic features that may be variable in severity. The male patients are usually recognized from a combination of typical behavior, intellectual disability, and sometimes dysmorphism.
- Coffin-Siris syndrome is genetically heterogenous and characterized by a combination of facial and other physical features. The facial dysmorphism is expected to be different from controls in analyses, although the group may not form a tight cluster because of phenotypic and genetic heterogeneity. This group is expected to be the least recognizable in this selection.

In **experiment 4**, known relationships between specific syndrome groups were verified in the CFPS. We investigated the positioning of four syndrome groups that are known to show considerable phenotypic similarity and overlap. Noonan, Costello, Cardiofaciocutaneous syndrome, and Neurofibromatosis Type I (NF1) are members of an etiologically related group of disorders, collectively known as the RASopathies [36]. These syndromes are all caused by the overactivation of the RASMAPK pathway. To determine if these four groups were closer to each other than expected by chance, a p-value was calculated for the average distance between cluster centers within an empirical null distribution estimated by recalculating the statistic 1000 times, each time randomly selecting 4 groups from the dataset. To gain visual feedback on the CFPS structure, a 2-dimensional visualization of the 100-dimensional CFPS was generated using the Uniform Manifold Approximation and Projection (UMAP) algorithm [37].

III. Results

For the first experiment, Table II reports the average and the standard error of the classification and clustering measures over the five cross-validation folds. The results in the first and the second row are based on the CFPS obtained by TB-SGAE and the baseline being PCA+LDA, respectively. The distributions of metrics, averaged within each syndrome group over the five folds are shown using boxplots in Fig. 4a. The p-value of the statistical test comparing TB-SGAE and the baseline is reported on top for each metric. According to the results, the performance of the TB-SGAE was significantly higher than the baseline (p-value <0.05) for balanced accuracy, specificity, ARI, and syndromes within the space showed increased BGD, and reduced WGV. For sensitivity, the results were not significantly different. To investigate the effect of the additional components of the model (decoder and SVD layer), Fig. 4a also compares a TB-SGAE with a TB-SGE and a TB-GAE to investigate the contribution of the decoder and the SVD layer respectively. The decoder significantly improved the classification measures: specificity, balanced accuracy, and ARI, while the sensitivity was not significantly different. At the same time, BGDs were significantly higher and WGVs were significantly lower, consistent with improved clustering. Removing the SVD layer significantly decreased all performance indicators but sensitivity. Fig. 4b shows the distributions of metrics, averaged within

bioRxiv preprint doi: https://doi.org/10.1101/2022.12.27.521999; this version posted December 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

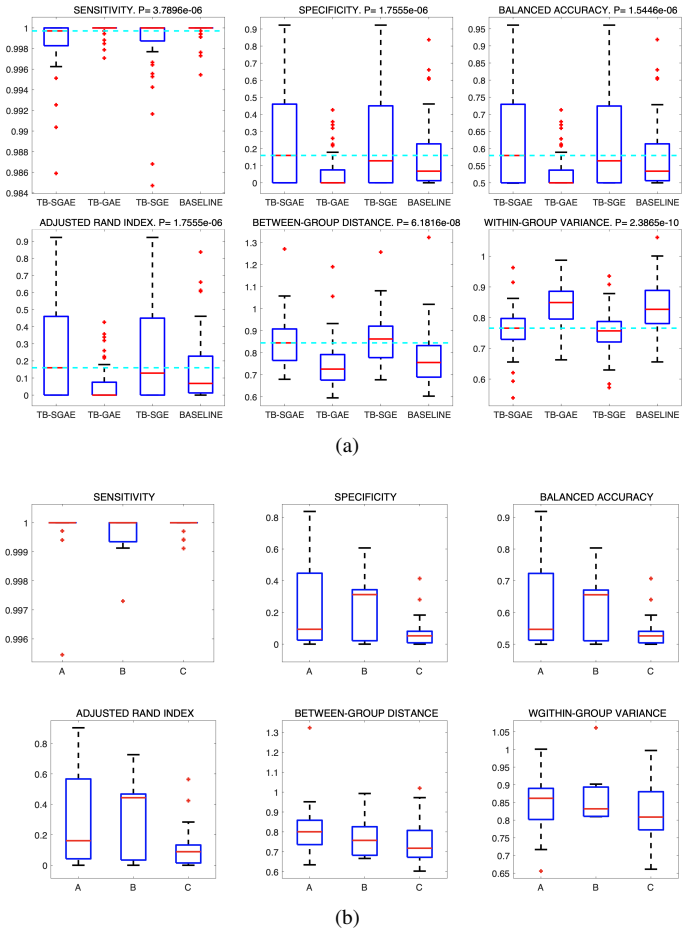


Fig. 4: (a) The average group-level metrics reported for complete model (TB-SGAE), model without the SVD layer (TB-GAE), model without the decoder (TB-SGE) and baseline (PCA+LDA). Classification and clustering measures are based on KNN classifier with K=10. (b) The average group-level metrics for all groups in category A, B, and C based on TB-SGAE.

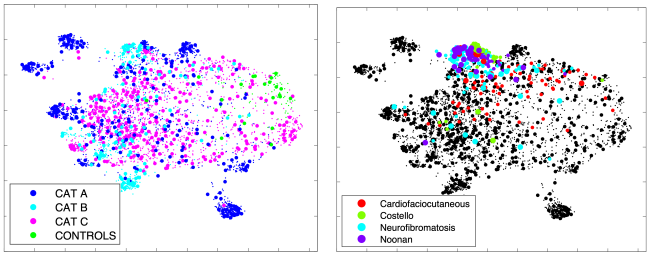


Fig. 5: (a) 2D UMAP visualization of the trainset (smaller dots, and test set (larger dots) into the space, colored by categories. (b) Colored 2D UMAP visualization of the four RASopathies together with the rest of the trainset (smaller dots) and test set (larger dots) colored in black.

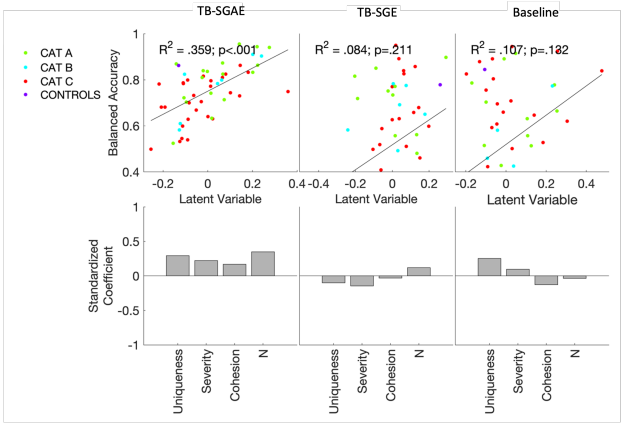


Fig. 6: PLS regression of accuracy onto phenotypic predictors and sample size. The bottom row plots the standardized coefficients of the linear combination defining the latent variable. The top row plots the regression of accuracy onto the latent variable. Columns correspond to the different spaces.

each syndrome group over the five folds, stratified by the clinical categorization (A, B, or C). Syndromes in categories A and B had a higher median sensitivity, specificity, balanced accuracy, and ARI than those in category C, with B showing slightly lower values than A. Furthermore, in line with the previous observation, and consistent with clinical expectation categories A and B have higher BGDs and lower WGVs, indicating in general they are in more isolated regions of the space than the syndromes in group C and have lower internal variation. Fig. 5a shows the 2D visualization of the CFPS obtained from TB-SGAE using the UMAP algorithm. The projection of individuals in the train set (smaller dots) and test set (larger dots) are colored by their categorization. The PLS regression of accuracy onto phenotypic predictors and sample size is shown in Fig. 6. The standardized coefficients of the linear combination and the regression of accuracy onto the derived latent variable are shown in Fig. 6. Accuracy in the CFPS of the TB-SGAE is significantly predicted by the phenotypic measures and sample size. The training and OOF error of reconstruction for TB-SGAE were 0.1597 and 0.1705 respectively. Reconstruction error per vertex is shown in Fig. 7a. The error bar is scaled in millimeters. The average error per vertex was less than 1 mm. Nevertheless, the heatmaps indicate that regions around the mouth, nose, and eyes had relatively higher errors. The lips and mouth regions are sensitive to expression variation, introducing extra complexity for the model to learn. The training and OOF of the model without the SVD layer (TB-GAE) were 0.1650 and 0.1710 respectively, which is slightly more than that of the complete model. To visually assess the precision and smoothness of the reconstructions from the CFPS of the complete model, the average test-set projections can be reconstructed. Fig. 7b shows recunstructions for four groups of Achondroplasia, Wolf Hirschhorn, Apert and williams from Category A syndromes.

The next experiment investigated the ability of the space to generalize to unseen syndromes. Six groups of syndromes were left out from the train set and the CFPS was trained based on the 47 remaining groups. The unseen groups and the test set were merged and projected to the space. Based on these projections the BGD, WGV, and the CIF were computed. Fig. 8 shows the average results for TB-SGAE and the baseline (PCA+LDA) over five folds of data. The median BGD of novel syndromes was significantly higher and the median WGV was significantly lower in TB-SGAE than in the baseline, and CIF was higher for five out of the six syndromes for TB-SGAE. Fig. 8 additionally compares TB-SGAE with TB-SGE (without decoder), TB-GAE (without the SVD layer), the space obtained from PCA+LDA baseline and the first 100 principal components. BGD and WGV and CIF were not significantly different with or without the decoder ($P<0.005$). However, there was a considerable improvement in all

bioRxiv preprint doi: https://doi.org/10.1101/2022.12.27.521999; this version posted December 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

TABLE II: The comparison of discriminating metrics computed for both TB-SGAE (+LDA for classification measures) and the linear baseline comprising of PCA + LDA. Reported metrics are: sensitivity, specificity, balanced accuracy (acc), adjusted rand index (ARI), embedding between-group distance (BGD), and within-group variance (WGV).

Model	Sensitivity	Specificity	Balanced acc	ARI	BGD	WGV
TB-SGAE	0.9987 ± 0.0003	0.2454 ± 0.0361	**0.6221 ± 0.018**	0.2454 ± 0.0361	0.857 ± 0.0148	0.7598 ± 0.0094
BASELINE	0.9998 ± 0.0001	0.1464 ± 0.0265	**0.5731 ± 0.0133**	0.1464 ± 0.0265	0.7761 ± 0.0175	0.8305 ± 0.0105

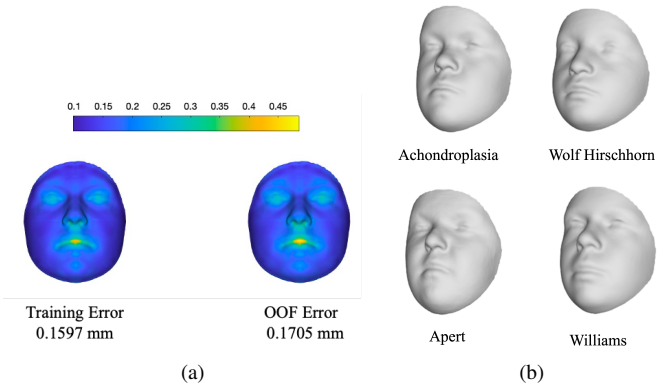


Fig. 7: (a) The training error of reconstruction (left) and out-of-fold error of reconstruction (right) (b) The reconstruction of the average embedding of individuals with Achondroplasia, Wolf Hirschhorn, Apert and williams using the geometric decoder.

three metrics when the SVD layer was added to the TB-GAE. The space of principal components had generalization power higher than a random performance. In addition, applying the LDA transformation to the PCA scores increased the CIF of all the syndrome groups but Cohen syndrome. Note that the maximum dimensionality of the LDA-based CFPS was bounded by the maximum number of classes in the training set minus one (i.e. 46 in this study). Therefore, we further investigated and compared the CIF results using a TB-SGAE space with 100 dimensions (used before and determined based on an SAE, see above) and 46 dimensions (equivalent to the LDA-based baseline), and observed no statistical difference (p=0.84). This is not entirely unexpected thanks to the SVD layer, which results in the most variance being coded the lower components, making it easier to reduce dimensionality, with the minimum loss of data information. For the last experiment, The UMAP plot in Fig. 5b shows the RASopathies grouped in the upper corner which confirms the proximity of these groups in the CFPS. The statistical test also indicated that within the CFPSs based on TB-SGAE, TB-SGE, and the baseline (PCA+LDA), the average distance between RASopathies cluster centers in the normalized CFPSs were lower than 99.7%, 99.6%, and 99.5% (respectively) of random selections of four groups.

IV. Discussion

In this work, we build a CFPS that models the range of facial dysmorphism present in 51 syndromes alongside general facial variations from a group of controls. To this end, we proposed a triplet-based singular geometric autoencoder for multi-task learning, to simultaneously learn facial shape variation and reconstruction, in an unsupervised way, and group discriminations with the supervision of syndrome labels.

The existing CNNs for syndrome classification or building CFPSs are mostly based on large-scale 2D photographs of patients with genetic syndromes. By now, large-scale databases of 3D photographs of clinical populations have been collected. Considering the expected growth in the popularity and accessibility of portable 3D imaging hardware, building systems that are applicable to this imaging modality is essential so as to fully exploit the 3D shape information contained in such images. With the recently developed field of GDL, CNNs are now directly applicable to 3D images. This eliminates the need for any domain transformation. Therefore, in this work, we aimed for building a CFPS based on 3D facial images using spiral convolutional operators with which we facilitate both syndrome classification and facial reconstruction. Once learned, we evaluated the main properties of the CFPS, being clustering of syndromes, generalization to novel syndromes, and the recapitulation of related genetic diseases. We also assessed the reconstruction precision from the CFPS and investigated the phenotypic shape predictors of the classification. We compared the performance of our space to a linear baseline which consists of PCA for dimensionality reduction and LDA, a linear metric learner. Similar work on 2D data [2], estimates the factor by which clustering is improved compared to random chance (CIF). Compared to such random performance, LDA is a much more difficult baseline to match or improve on. In fact, for statistical shape analysis, LDA and its regularized variants were and still are strong and popular methods that are also used and outperformed many other classifiers in the 3D syndrome classification published in [23].

Our proposed model consists of three main components. The first is a triplet-based encoder which was used in the recent syndrome classification work in [25] to optimize the distances among individuals belonging to different syndrome groups. In the triplet-loss function, the focus is on learning the CFPS such that the distances are a measure of similarity and group membership and therefore it contributes to the classification and clustering power of the space. The second component is a decoder that not only allows the reconstruction of a face from an embedding in a CFPS but also improves the classification and clustering performance of the system. The decoder does not diminish (nor improve) the generalization capacity of the CFPS. The third component, being an SVD layer, makes it simpler to select the dimensionality of the space without retraining and also improves the classification, reconstruction, and generalization aspects of the CFPS.

We showed that the CFPS built based on the complete model (TB-SGAE) outperforms the classification and the clustering performances of the linear model which consists of PCA for dimensionality reduction and LDA for metric learning. We then compared the results to a TB-SGE and observed higher BGD and lower WGV in the encoding, which was not surprising since the network's attention during learning is expanded away from these aspects only. Interestingly, this property not only did not diminish the classification and clustering performance, but it improved them, illustrating the use of coding general facial variation and similarity beyond simple supervised learning of the CFPS. In other words, two facially similar individuals belonging to different groups are embedded closer together in the CFPS that preserves the topology, while this information is overlooked when the CFPS that only focuses on the phenotypic features that are related to group discrimination. Although a direct comparison with the state-of-the-art 3D syndrome diagnosis in [16] is not available,

bioRxiv preprint doi: https://doi.org/10.1101/2022.12.27.521999; this version posted December 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

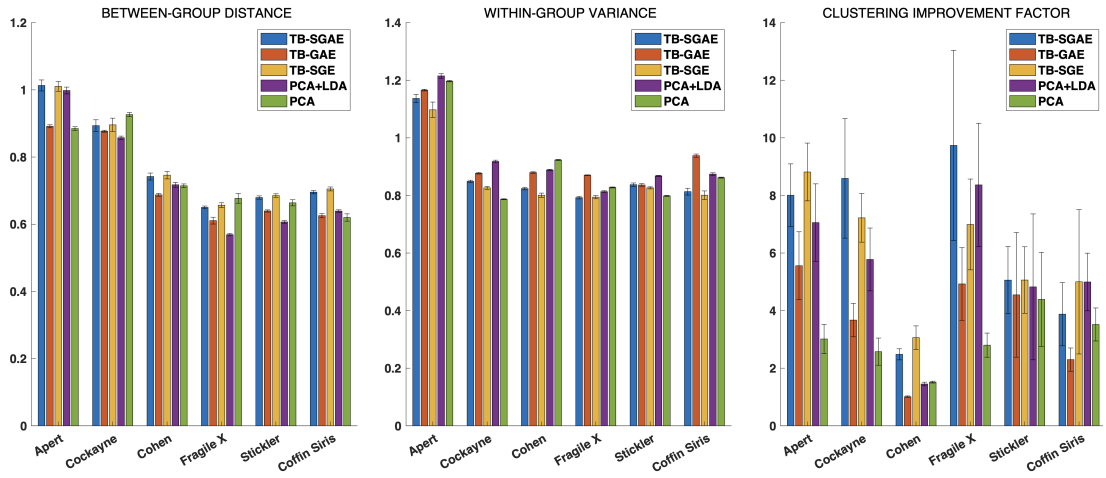


Fig. 8: The comparison of the between-group distance, within-group variance (WGV), and clustering improvement factor (CIF) for individuals in the six left-out groups of the experiment 3, projected to the CFPSs obtained by TB-SGAE, TB-GAE, TB-SGE, PCA+LDA (baseline) and PCA. Error bars indicate the standard error of the mean over five folds.

the classification results are competitive. It is important to note that our classification results are affected by sample size which is due to the data imbalance (some groups have as low as 10 individuals). We believe that this can improve with an expanded dataset.

Furthermore, we investigated whether the CFPS is inline with the expectations of clinicians. Experts in clinical genetics categorized syndromes into three classes. Two include syndromes that are phenotypically distinctive (A and B) and the third class of syndromes that are not necessarily phenotypically distinctive (C). All the classification and clustering metrics but the WGV differ in the expected direction significantly for the phenotypically distinctive categories A and B, compared to C. Groups in category C show smaller WGV, contrary to expectation, however this is highly variable depending on the syndrome. Hence, the CFPS largely conformed to expectation. Graphically, this is shown in the UMAP visualization (Fig. 5a) which shows that syndromes in categories A and B generate more isolated and clear clusters around the corners while category C groups show less clear cluster boundaries and are positioned close together around the center of the mapped embeddings. Category A and B are also more distant from controls than category C.

With the ability to construct a face from an embedding, we can visualize and hence explain and understand the embeddings better. The decoder component facilitates the reconstruction of encoded facial shapes with less than 1 mm error. In addition, thanks to the orthonormal dimensions, we have a coordinate system in the space that properly spans every vector and thus allows us to explore and interpolate the space more structurally. In other words, one can manipulate one dimension without changing the values on other dimensions, this is a property of a vector basis or coordinate system which is not available in a default autoencoder.

Furthermore, we evaluated the clustering generalization onto novel syndromes that have not been included in the training set. The novel syndromes projected into the CFPS are significantly more unique and cohesive in our complete network than that of the linear baseline. We also computed and compared the CIF for the novel syndromes. This comparison shows that the clustering improves from a random chance for both baseline and TB-SGAE. Compared to the linear baseline, the improvement is stronger for five out of six novel syndromes within our CFPS. Moreover, results of the comparison between the CFPS built by the same models with and without the SVD layers indicated that the added structure imposed on the latent space by adding the SVD layer makes it significantly more generalizable to novel syndromes. The comparison between TB-SGAE and the TB-SGE shows that the generalization performance is not affected by adding the decoder. However, it still is a solid addition that further enables the ability to visualize and therefore interpret the metric space by generating faces. The comparison of the generalization power between the supervised linear metric learning approach (PCA+LDA) and that of the unsupervised and linear PCA shows that in five out of six left out groups supervised learning improves the performance. The group that has superior performance with PCA (Cohen syndrome) is known to be clinically difficult to recognize from the face, and this observation suggests that the supervision of metric learning has less influence on groups that have little to no facial clues for diagnosis. It also worth mentioning that the CIF obtained by PCA suggests that despite the unsupervised nature of this method, it still is powerful enough to improve the clustering factor considerably compared to random performance.

Finally, we tested for expected low distance among the four RASopathies, an etiologically related group of disorders caused by mutations in genes encoding the RAS/MAPK pathway. Results of the statistical test indicated that the four groups are significantly closer together in the CFPS of TB-SGAE, TB-SGE, and the linear baseline compared to random selections of other four syndrome groups in each of these spaces. This test provided us with one piece of evidence towards the recapitulation of the relatedness in the CFPS, although it does not demonstrate any improvement of the deep metric learners over the linear baseline, in which our measure of this recapitulation was already close to ceiling. When genetic data are available, more comprehensive tests, correlating measures of genetic similarity to phenotypic distance should be performed. For example genetic similarity can be based on protein-to-protein interaction as per Ferry et al. [2] or distance based on patterns of DNA methylation [38].

The proposed clinical face phenotype space can facilitate interpretation of facial dysmorphism of a subject by positioning them within the space of known facial dysmorphisms. The generalization and classification improvements of the space demonstrate the space prioritizes facial variations that are disturbed by genetic anomalies, facilitating the assessment of dysmorphism for new subjects even when they do not belong to one of the groups used to learn the metric space. In this way, the CFPS can contribute to the identification of novel genetic

bioRxiv preprint doi: https://doi.org/10.1101/2022.12.27.521999; this version posted December 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

disorders by matching individuals with facial dysmorphism caused by the same rare and novel disorder. These clusteres are prime targets for genetic investigations to unravel the etiology of novel genetic disorders, as illustrated by [3]. This work can also be technically expanded by implementing other existing geometric deep learning methods, aside from the spiral convolutions used in this work, such as PointNet++ [39], which enable learning from unstructured 3D data. This makes it possible to learn directly from the original 3D scans and shortcuts the pre-processing steps of the pipeline used in this work. However, a steeper learning curve is expected since extensive data normalization is less trivial to implement. Additionally, inspired by [24] we can combine normalizing flow architecture with our geometric deep learning-based approach to possibly improve the performance and flexibility of the CFPS. For our deep learning framework, further improvement is expected with additional data which will be collected and processed in the future. Finally, designing more in-depth experiments on genotype-phenotype correlation can reveal valuable information for clinicians and experts in the field.

V. Conclusion

In this work, we proposed a CFPS learner based on 3D facial images and GDL techniques for large-scale syndrome analysis. The proposed model consists of the base component being a geometric encoder, which is further expanded by our additional components being a geometric decoder, with which high-precision facial shapes are reconstructed from an embedding in the CFPS, and a singular value decomposition layer to encode a structured facial mesh into an orthonormal 100-dimensional CFPS. We used a multi-task learning approach to train the model in an end-to-end manner. The loss function combines the supervised triplet-loss function with the unsupervised reconstruction-loss. In summary, we showed that supervised and unsupervised learning strategies both improve the clustering factor compared to a random performance. Moreover, supervised learning leads to superior performance compared to unsupervised learning only. Lastly, the proposed GDL-based model learns a CFPS that outperforms the linear metric learning baseline (consisting of PCA and LDA), in both syndrome classification or clustering and generalization to novel syndromes. We proved the contribution of each added component in the classification, reconstruction, and generalization capacity of the CFPS. More precisely, we showed that the attached decoder not only facilitated the ability to reconstruct patient faces and generate synthetic faces but also improved the classification performance of the model. In addition, the orthonormal base of the CFPS facilitated by the SVD layer has considerably impacted the classification, reconstruction, and more importantly the generalization capacity of the space. We also showed that the space strongly replicates clinical expectations such that the classification and clustering measures obtained from the CFPS relate to the categorization of syndromes. Further the proximity of the four RASopathies, characterized by mutations in functionally-related genes is reflected in the CFPS. The resulting CFPS can potentially narrow the search space for diagnosing new instances of the syndromes that are represented in the space, objectively assess facial similarity between undiagnosed patients who share a rare and novel disorder, and facilitate targeted sequencing of genomic regions to identify causal variants.

References

[1] M. Oti and H. G. Brunner, "The modular nature of genetic diseases," *Clinical Genetics*, vol. 71, pp. 1–11, Jan. 2007.

[2] Q. Ferry, J. Steinberg, C. Webber, D. R. FitzPatrick, C. P. Ponting, A. Zisserman, and C. Nellåker, "Diagnostically relevant facial gestalt information from ordinary photos," *eLife*, vol. 3, p. e02020, June 2014.

[3] F. Marbach, C. F. Rustad, A. Riess, D. ukić, T.-C. Hsieh, I. Jobani, T. Prescott, A. Bevot, F. Erger, G. Houge, M. Redfors, J. Altmueller, T. Stokowy, C. Gilissen, C. Kubisch, E. Scarano, L. Mazzanti, T. Fiskerstrand, P. M. Krawitz, D. Lessel, and C. Netzer, "The Discovery of a LEMD2-Associated Nuclear Envelopathy with Early Progeroid Appearance Suggests Advanced Applications for AI-Driven Facial Phenotyping," *The American Journal of Human Genetics*, vol. 104, pp. 749–757, Apr. 2019.

[4] T.-C. Hsieh, M. A. Mensah, J. T. Pantel, D. Aguilar, O. Bar, A. Bayat, L. Becerra-Solano, H. B. Bentzen, S. Biskup, O. Borisov, O. Braaten, C. Ciaccio, M. Coutelier, K. Cremer, M. Danyel, S. Daschkey, H. D. Eden, K. Devriendt, S. Wilson, S. Douzgou, D. ukić, N. Ehmke, C. Fauth, B. Fischer-Zirnsak, N. Fleischer, H. Gabriel, L. Graul-Neumann, K. W. Gripp, Y. Gurovich, A. Gusina, N. Haddad, N. Hajjir, Y. Hanani, J. Hertzberg, K. Hoertnagel, J. Howell, I. Ivanovski, A. Kaindl, T. Kamphans, S. Kamphausen, C. Karimov, H. Kathom, A. Keryan, A. Knaus, S. Köhler, U. Kornak, A. Lavrov, M. Leitheiser, G. J. Lyon, E. Mangold, P. M. Reina, A. M. Carrascal, D. Mitter, L. M. Herrador, G. Nadav, M. Nöthen, A. Orrico, C.-E. Ott, K. Park, B. Peterlin, L. Pölsler, A. Raas-Rothschild, L. Randolph, N. Revencu, C. R. Fagerberg, P. N. Robinson, S. Rosnev, S. Rudnik, G. Rudolf, U. Schatz, A. Schossig, M. Schubach, O. Shanoon, E. Sheridan, P. Smirin-Yosef, M. Spielmann, E.-K. Suk, Y. Sznajer, C. T. Thiel, G. Thiel, A. Verloes, I. Vrecar, D. Wahl, I. Weber, K. Winter, M. Wiśniewska, B. Wollnik, M. W. Yeung, M. Zhao, N. Zhu, J. Zschocke, S. Mundlos, D. Horn, and P. M. Krawitz, "PEDIA: prioritization of exome data by image analysis," *Genetics in Medicine*, vol. 21, pp. 2807–2814, Dec. 2019.

[5] H. Matthews, M. Vanneste, K. Katsura, D. Aponte, M. Patton, P. Hammond, G. Baynam, R. Spritz, O. D. Klein, B. Hallgrimsson, H. Peeters, and P. Claes, "Refining nosology by modelling variation among facial phenotypes: the RASopathies," *Journal of Medical Genetics*, pp. jmedgenet–2021–108366, July 2022.

[6] S. Kreiborg and M. M. Cohen, "Ocular Manifestations of Apert and Crouzon Syndromes: Qualitative and Quantitative Findings," *Journal of Craniofacial Surgery*, vol. 21, pp. 1354–1357, Sept. 2010.

[7] A. Dickmann, R. Parrilla, A. Salerni, G. Savino, I. Vasta, M. Zollino, S. Petroni, and G. Zampino, "Ocular manifestations in Wolf-Hirschhorn syndrome," *Journal of American Association for Pediatric Ophthalmology and Strabismus*, vol. 13, pp. 264–267, June 2009.

[8] J. E. Allanson, A. Bohring, H.-G. Dörr, A. Dufke, G. Gillessen-Kaesbach, D. Horn, R. König, C. P. Kratz, K. Kutsche, S. Pauli, S. Raskin, A. Rauch, A. Turner, D. Wieczorek, and M. Zenker, "The face of Noonan syndrome: Does phenotype predict genotype," *American Journal of Medical Genetics Part A*, vol. 152A, pp. 1960–1966, July 2010.

[9] P. Hammond, M. Suttie, R. C. Hennekam, J. Allanson, E. M. Shore, and F. S. Kaplan, "The face signature of fibrodysplasia ossificans progressiva," *American Journal of Medical Genetics Part A*, vol. 158A, pp. 1368–1380, June 2012.

[10] A. D. Kline, J. F. Moss, A. Selicorni, A.-M. Bisgaard, M. A. Deardorff, P. M. Gillett, S. L. Ishman, L. M. Kerr, A. V. Levin, P. A. Mulder, F. J. Ramos, J. Wierzba, P. F. Ajmone, D. Axtell, N. Blagowidow, A. Cereda, A. Costantino, V. Cormier-Daire, D. FitzPatrick, M. Grados, L. Groves, W. Guthrie, S. Huisman, F. J. Kaiser, G. Koekkoek, M. Levis, M. Mariani, J. P. McCleery, L. A. Menke, A. Metrena, J. O'Connor, C. Oliver, J. Pie, S. Piening, C. J. Potter, A. L. Quaglio, E. Redeker, D. Richman, C. Rigamonti, A. Shi, Z. Tümer, I. D. C. Van Balkom, and R. C. Hennekam, "Diagnosis and management of Cornelia de Lange syndrome: first international consensus statement," *Nature Reviews Genetics*, vol. 19, pp. 649–666, Oct. 2018.

[11] N. Nauwelaers, H. Matthews, Y. Fan, B. Croquet, H. Hoskens, S. Mahdi, A. El Sergani, S. Gong, T. Xu, M. Bronstein, M. Marazita, S. Weinberg, and P. Claes, "Exploring palatal and dental shape variation with 3D shape analysis and geometric deep learning," *Orthodontics & Craniofacial Research*, p. ocr.12521, Aug. 2021.

[12] J. Thevenot, M. B. Lopez, and A. Hadid, "A Survey on Computer Vision for Assistive Medical Diagnosis From Faces," *IEEE Journal of Biomedical and Health Informatics*, vol. 22, pp. 1497–1511, Sept. 2018.

[13] Q. Hennocq, R. H. Khonsari, V. Benoît, M. Rio, and N. Garcelon, "Computational diagnostic methods on 2D photographs: A review of the literature," *Journal of Stomatology, Oral and Maxillofacial Surgery*, p. S2468785521000781, Apr. 2021.

[14] O. Agbolade, A. Nazri, R. Yaakob, A. A. Ghani, and Y. K. Cheah, "Down Syndrome Face Recognition: A Review," *Symmetry*, vol. 12, p. 1182, July 2020.

[15] Y. Gurovich, Y. Hanani, O. Bar, G. Nadav, N. Fleischer, D. Gelbman, L. Basel-Salmon, P. M. Krawitz, S. B. Kamphausen, M. Zenker, L. M.

bioRxiv preprint doi: https://doi.org/10.1101/2022.12.27.521999; this version posted December 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Bird, and K. W. Gripp, "Identifying facial phenotypes of genetic disorders using deep learning," *Nature Medicine*, vol. 25, pp. 60–64, Jan. 2019.

[16] T.-C. Hsieh, A. Bar-Haim, S. Moosa, N. Ehmke, K. W. Gripp, J. T. Pantel, M. Danyel, M. A. Mensah, D. Horn, S. Rosnev, N. Fleischer, G. Bonini, A. Hustinx, A. Schmid, A. Knaus, B. Javanmardi, H. Klinkhammer, H. Lesmann, S. Sivalingam, T. Kamphans, W. Meiswinkel, F. Ebstein, E. Krüger, S. Küry, S. Bézieau, A. Schmidt, S. Peters, H. Engels, E. Mangold, M. Kreiß, K. Cremer, C. Perne, R. C. Betz, T. Bender, K. Grundmann-Hauser, T. B. Haack, M. Wagner, T. Brunet, H. B. Bentzen, L. Averdunk, K. C. Coetzer, G. J. Lyon, M. Spielmann, C. Schaaf, S. Mundlos, M. M. Nöthen, and P. Krawitz, "GestaltMatcher: Overcoming the limits of rare disease matching using facial phenotypic descriptors," preprint, Genetic and Genomic Medicine, Jan. 2021.

[17] "The use of 3D face shape modelling in dysmorphology," *Archives of Disease in Childhood*, vol. 92, pp. 1120–1126, Dec. 2007.

[18] J. Cox-Brinkman, A. Vedder, C. Hollak, L. Richfield, A. Mehta, K. Orteu, F. Wijburg, and P. Hammond, "Three-dimensional face shape in Fabry disease," *European Journal of Human Genetics*, vol. 15, pp. 535–542, May 2007.

[19] J. Meulstee, L. Verhamme, W. Borstlap, F. Van der Heijden, G. De Jong, T. Xi, S. Bergé, H. Delye, and T. Maal, "A new method for three-dimensional evaluation of the cranial shape and the automatic identification of craniosynostosis using 3D stereophotogrammetry," *International Journal of Oral and Maxillofacial Surgery*, vol. 46, pp. 819–826, July 2017.

[20] A. C. E. Hurst, "Facial recognition software in clinical dysmorphology," *Current Opinion in Pediatrics*, vol. 30, no. 6, pp. 701–706, 2018.

[21] H. L. Rudy, N. Wake, J. Yee, E. S. Garfein, and O. M. Tepper, "Three-Dimensional Facial Scanning at the Fingertips of Patients and Surgeons: Accuracy and Precision Testing of iPhone X Three-Dimensional Scanner," *Plastic and Reconstructive Surgery*, vol. 146, pp. 1407–1417, Dec. 2020.

[22] O. Klein, W. Mio, R. Spritz, and B. Hallgrimsson, "Developing 3D Craniofacial Morphometry Data and Tools to Transform Dysmorphology," 2019. Version Number: 1 type: dataset.

[23] B. Hallgrímsson, J. D. Aponte, D. C. Katz, J. J. Bannister, S. L. Riccardi, N. Mahasuwan, B. L. McInnes, T. M. Ferrara, D. M. Lipman, A. B. Neves, J. A. J. Spitzmacher, J. R. Larson, G. A. Bellus, A. M. Pham, E. Aboujaoude, T. A. Benke, K. C. Chatfield, S. M. Davis, E. R. Elias, R. W. Enzenauer, B. M. French, L. L. Pickler, J. T. C. Shieh, A. Slavotinek, A. R. Harrop, A. M. Innes, S. E. McCandless, E. A. McCourt, N. J. L. Meeks, N. R. Tartaglia, A. C.-H. Tsai, J. P. H. Wyse, J. A. Bernstein, P. A. Sanchez-Lara, N. D. Forkert, F. P. Bernier, R. A. Spritz, and O. D. Klein, "Automated syndrome diagnosis by three-dimensional facial imaging," *Genetics in Medicine*, June 2020.

[24] J. Bannister, M. Wilms, D. Aponte, D. Katz, O. D. Klein, F. P. Bernier, R. Spritz, B. Hallgrimsson, and N. D. Forkert, "A Deep Invertible 3D Facial Shape Model For Interpretable Genetic Syndrome Diagnosis," *IEEE Journal of Biomedical and Health Informatics*, pp. 1–1, 2022.

[25] S. S. Mahdi, H. Matthews, N. Nauwelaers, M. Vanneste, S. Gong, G. Bouritsas, G. S. Baynam, P. Hammond, R. Spritz, O. D. Klein, B. Hallgrimsson, H. Peeters, M. Bronstein, and P. Claes, "Multi-Scale Part-Based Syndrome Classification of 3D Facial Images," *IEEE Access*, vol. 10, pp. 23450–23462, 2022.

[26] G. Bouritsas, S. Bokhnyak, S. Ploumpis, S. Zafeiriou, and M. Bronstein, "Neural 3D Morphable Models: Spiral Convolutional Networks for 3D Shape Representation Learning and Generation," in *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, (Seoul, Korea (South)), pp. 7212–7221, IEEE, Oct. 2019.

[27] S. Gong, L. Chen, M. Bronstein, and S. Zafeiriou, "SpiralNet++: A Fast and Highly Efficient Mesh Convolution Operator," in *2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW)*, (Seoul, Korea (South)), pp. 4141–4148, IEEE, Oct. 2019.

[28] E. Hoffer and N. Ailon, "Deep Metric Learning Using Triplet Network," in *Similarity-Based Pattern Recognition* (A. Feragen, M. Pelillo, and M. Loog, eds.), Lecture Notes in Computer Science, (Cham), pp. 84–92, Springer International Publishing, 2015.

[29] F. Schroff, D. Kalenichenko, and J. Philbin, "FaceNet: A unified embedding for face recognition and clustering," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, (Boston, MA, USA), pp. 815–823, IEEE, June 2015.

[30] S. Kung, M. Walters, P. Claes, P. LeSouef, J. Goldblatt, A. Martin, S. Balasubramaniam, and G. Baynam, "Monitoring of Therapy for Mucopolysaccharidosis Type I Using Dysmorphometric Facial Phenotypic Signatures," in *JIMD Reports, Volume 22* (J. Zschocke, M. Baumgartner, E. Morava, M. Patterson, S. Rahman, and V. Peters, eds.), vol. 22, pp. 99–106, Berlin, Heidelberg: Springer Berlin Heidelberg, 2015. Series Title: JIMD Reports.

[31] P. Hammond and M. Suttie, "Large-scale objective phenotyping of 3D facial morphology," *Human Mutation*, vol. 33, pp. 817–825, May 2012.

[32] F. Schroff, D. Kalenichenko, and J. Philbin, "FaceNet: A unified embedding for face recognition and clustering," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, (Boston, MA, USA), pp. 815–823, IEEE, June 2015.

[33] N. Halko, P. G. Martinsson, and J. A. Tropp, "Finding Structure with Randomness: Probabilistic Algorithms for Constructing Approximate Matrix Decompositions," *SIAM Review*, vol. 53, pp. 217–288, Jan. 2011.

[34] M. M. Bronstein, J. Bruna, Y. LeCun, A. Szlam, and P. Vandergheynst, "Geometric Deep Learning: Going beyond Euclidean data," *IEEE Signal Processing Magazine*, vol. 34, pp. 18–42, July 2017. Conference Name: IEEE Signal Processing Magazine.

[35] S. S. Mahdi, N. Nauwelaers, P. Joris, G. Bouritsas, S. Gong, S. Bokhnyak, S. Walsh, M. D. Shriver, M. Bronstein, and P. Claes, "3D Facial Matching by Spiral Convolutional Metric Learning and a Biometric Fusion-Net of Demographic Properties," in *2020 25th International Conference on Pattern Recognition (ICPR)*, (Milan, Italy), pp. 1757–1764, IEEE, Jan. 2021.

[36] W. E. Tidyman and K. A. Rauen, "The RASopathies: developmental syndromes of Ras/MAPK pathway dysregulation," *Current Opinion in Genetics & Development*, vol. 19, pp. 230–236, June 2009.

[37] L. McInnes, J. Healy, N. Saul, and L. Großberger, "UMAP: Uniform Manifold Approximation and Projection," *Journal of Open Source Software*, vol. 3, p. 861, Sept. 2018.

[38] E. Aref-Eshghi, J. Kerkhof, V. P. Pedro, M. Barat-Houari, N. Ruiz-Pallares, J.-C. Andrau, D. Lacombe, J. Van-Gils, P. Fergelot, C. Dubourg, V. Cormier-Daire, S. Rondeau, F. Lecoquierre, P. Saugier-Veber, G. Nicolas, G. Lesca, N. Chatron, D. Sanlaville, A. Vitobello, L. Faivre, C. Thauvin-Robinet, F. Laumonnier, M. Raynaud, M. Alders, M. Mannens, P. Henneman, R. C. Hennekam, G. Velasco, C. Francastel, D. Ulveling, A. Ciolfi, S. Pizzi, M. Tartaglia, S. Heide, D. Héron, C. Mignot, B. Keren, S. Whalen, A. Afenjar, T. Bienvenu, P. M. Campeau, J. Rousseau, M. A. Levy, L. Brick, M. Kozenko, T. B. Balci, V. M. Siu, A. Stuart, M. Kadour, J. Masters, K. Takano, T. Kleefstra, N. de Leeuw, M. Field, M. Shaw, J. Gecz, P. J. Ainsworth, H. Lin, D. I. Rodenhiser, M. J. Friez, M. Tedder, J. A. Lee, B. R. DuPont, R. E. Stevenson, S. A. Skinner, C. E. Schwartz, D. Genevieve, and B. Sadikovic, "Evaluation of DNA Methylation Episignatures for Diagnosis and Phenotype Correlations in 42 Mendelian Neurodevelopmental Disorders," *The American Journal of Human Genetics*, vol. 106, pp. 356–370, Mar. 2020.

[39] C. Ruizhongtai Qi, L. Yi, H. Su, and L. Guibas, "PointNet++: Deep Hierarchical Feature Learning on Point Sets in a Metric Space," pp. 5099–5108, June 2017.