IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

IEEE ENGINEERING IN MEDICINE AND BIOLOGY SOCIETY SECTION

Received January 27, 2022, accepted February 13, 2022, date of publication February 22, 2022, date of current version March 7, 2022.

*Digital Object Identifier 10.1109/ACCESS.2022.3153357*

# Multi-Scale Part-Based Syndrome Classification of 3D Facial Images

**SOHA SADAT MAHDI[1], HAROLD MATTHEWS[1,2,3], NELE NAUWELAERS[1], MICHIEL VANNESTE[2], SHUNWANG GONG[4], GIORGOS BOURITSAS[4], GARETH S. BAYNAM[5,6], PETER HAMMOND[2], RICHARD SPRITZ[7], OPHIR D. KLEIN[8], BENEDIKT HALLGRÍMSSON[9], HILDE PEETERS[2], MICHAEL BRONSTEIN[4], AND PETER CLAES[1,2]**

[1]MIRC, ESAT/PSI—UZ Leuven, KU Leuven, 3000 Leuven, Belgium
[2]Laboratory for Genetic Epidemiology, Department of Human Genetics, KU Leuven, 3000 Leuven, Belgium
[3]Facial Sciences Research Group, Murdoch Children's Research Institute, Parkville, VIC 3052, Australia
[4]Department of Computing, Imperial College London, London SW7 2AZ, U.K.
[5]School of Earth and Planetary Sciences, Faculty of Science and Engineering, Curtin University, Perth, WA 6845, Australia
[6]Western Australian Register of Developmental Anomalies, King Edward Memorial Hospital, Subiaco, WA 6008, Australia
[7]Human Medical Genetics and Genomics Program, School of Medicine, University of Colorado, Aurora, CO 80045, USA
[8]Department of Orofacial Sciences and Pediatrics, University of California at San Francisco, San Francisco, CA 94143, USA
[9]Department of Cell Biology and Anatomy, Alberta Childre's Hospital Research Institute, Cumming School of Medicine, University of Calgary, Calgary, AB T2N 4N1, Canada

Corresponding author: Soha Sadat Mahdi (sohasadat.mahdi@kuleuven.be)

This work was supported in part by the Research Fund Katholieke Universiteit (KU) Leuven through Bijzonder Onderzoeksfond (BOF-C1) under Grant C14/20/081; and in part by the Research Program of the Research Foundation—Flanders, Belgium, through Fonds Wetenschappelijk Onderzoek (FWO) under Grant G078518N. The work of Peter Hammond, Peter Claes, and Hilde Peeters was supported in part by the KU Leuven Research Team, and in part by the National Institutes of Health under Grant 1-R01-DE027023.

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Ethical Review Board of KU Leuven and University Hospitals Gasthuisberg, Leuven under Application Nos. S56392 and S60568.

**ABSTRACT** Identification and delineation of craniofacial characteristics support the clinical and molecular diagnosis of genetic syndromes. Deep learning (DL) frameworks for syndrome identification from 2D facial images are trained on large clinical datasets using standard convolutional neural networks for classification. In contrast, despite the increased availability of 3D scanners in clinical setups, similar frameworks remain absent for 3D facial photographs. The main challenges involve working with smaller datasets and the need for DL operations applicable to 3D geometric data. Therefore, to date, most 3D methods refrain from working across multiple syndromic groups and/or are solely based on traditional machine learning. The first contribution of this work is the use of geometric deep learning with spiral convolutions in a triplet-loss architecture. This geometric encoding (GE) learns a lower dimensional metric space from 3D facial data that is used as input to linear discriminant analysis (LDA) performing multiclass classification. Benchmarking is done against principal component analysis (PCA), a common technique in 3D facial shape analysis, and related work based on 65 distinct 3D facial landmarks as input to LDA. The second contribution of this work involves a part-based implementation to 3D facial shape analysis and multi-class syndrome classification, and this is applied to both GE and PCA. Based on 1,786 3D facial photographs of controls and individuals from 13 different syndrome classes, a five-fold cross-validation was used to investigate both contributions. Results indicate that GE performs better than PCA as input to LDA, and this especially so for more compact (lower dimensional) spaces. In addition, a part-based approach increases performance significantly for both GE and PCA, with a more significant improvement for the latter. I.e., this contribution enhances the power of the dataset. Finally, and interestingly, according to ablation studies within the part-based approach, the upper lip is the most distinguishing facial segment for classifying genetic syndromes in our dataset, which follows clinical expectation. This work stimulates an enhanced use of advanced part-based geometric deep learning methods for 3D facial imaging in clinical genetics.

The associate editor coordinating the review of this manuscript and approving it for publication was Gina Tourassi.

**INDEX TERMS** Clinical genetics, computer-aided diagnosis, deep phenotyping, 3D shape analysis, geometric deep learning, precision public health, spiral convolutions, syndrome classification.

## I. INTRODUCTION

Genetic conditions frequently present with distinct facial characteristics, which are often the first clue in diagnostics. To this day, a clinical diagnosis relies on an assessment by clinical experts who are trained to recognize facial phenotypes associated with syndroms. However, subtle facial phenotypes may not be obvious to the clinician. Further, it can be difficult for a clinician to keep pace with the ever-expanding catalog of clinical and molecular diagnoses and their associated phenotypes. Therefore, objective facial phenotyping for syndrome identification is needed to assist in clinical diagnosis [1]–[5]. Previous work has mostly focused on 2D facial images [6], with large-scale deep convolutional neural networks being developed for and implemented in the clinic [1]. While 2D photographs are easier to obtain, 3D images capture facial shape and morphology more directly and accurately as they are not subject to distortions due to projections, positional changes, and lighting conditions. Given the increasing accessibility of 3D imaging hardware, including consumer-grade depth sensors in modern smartphones [7], large-scale 3D shape analysis and deep learning for syndrome classification is becoming a practical possibility. Previous work in this domain has used linear dimensionality reduction and classification techniques, or feed-forward neural networks, and has typically focused on discriminating one or a few syndrome groups from controls [8]–[10]. The most comprehensive attempt at 3D multi-syndrome classification deployed linear techniques on 64 syndrome classes, with a sparse configuration of 65 anatomical 3D facial landmarks [3].

Training convolutional neural networks (CNNs) on 3D photographs is a developing field. Some approaches ignore the local connectivity of the interconnected 3D 'mesh' data by transforming it into a 2D UV or 3D voxel representation [11], [12], or by learning only from the point cloud [13]. This may result in a substantial loss of information about the surface geometry. With recently introduced Geometric Deep Learning (GDL) techniques [14], it is now possible to apply deep learning directly on non-Euclidean facial surfaces, which are discretized as graphs or meshes [15]–[18]. Drawing from this literature, we use spiral convolutional operators which apply local anisotropic filters to features given on a non-Euclidean domain, mimicking the classical convolutional filters used in CNNs [15], [19]–[22]. These spiral operators are applicable to meshes that share a common topology, as for the 3D facial data in this work. While a cross-entropy loss is the typical loss used for classification tasks, we use deep metric learning instead to learn similarity measures based on discriminative facial features. More specifically, we implement a geometric encoder (GE), which is a triplet-based Siamese architecture trained with a triplet loss function [23]. The advantage of the triplet loss is that it can efficiently learn many groups, even with a small number of samples per group [24], [25].

The human face is a multipartite morphological shape that expresses both shape integration and modularity [6]. Anatomical structures with different embryological origins and functions are combined in a viable, functional whole. As such, information useful for syndrome classification may occur in individual and different facial regions and/or at the level of the whole face. To exploit this, multi-scale part-based approaches have been employed in literature, for 2D images, in which separate models are trained for local regions such as the eyes, nose, mouth, and chin, along with the entire face [1], [26]. Inspired by this idea, we propose a part-based GE using a data-driven hierarchical facial segmentation, intended to reflect the integrated and modular nature of facial variation. Figure 1 shows the general pipeline. First, a low dimensional embedding vector is learned for each facial segment using a separate encoder block. Then, the embeddings of multiple segments are concatenated to form the final embedding space which is then fed to a linear discriminant analysis (LDA) classifier. Performance of the GE is benchmarked against principal component analysis (PCA), which is a commonly used unsupervised linear dimensionality reduction technique in 3D facial shape analysis [27], [28]. For each embedding type (PCA and GDL), we assess the contribution of the part-based setup by comparing its performance to using the full face as a single segment and investigate the contributions of different facial segments to the performance.

The main contributions of this work can be shortly listed as: 1) combining geometric deep learning with spiral convolutions and a triplet-based architecture, for the first time to learn phenotypic features that discriminate genetic syndromes from 3D facial shape; 2) incorporating a part-based approach to 3D facial shape analysis and multi-class syndrome classification such that the classification performance increases for both baseline and our geometric model; 3) providing visual and quantitative feedback on the classification output, and investigating the most distinguishing facial segment for classifying genetic syndromes by performing ablation studies within the part-based approach.

## II. MATERIALS AND METHODS

### A. DATASET

The dataset comprises 1,786 3D facial images of controls and individuals clinically diagnosed into one of 13 different syndromes. All images were captured using the 3dMD or Vectra H1 3D imaging systems and were sourced from:

1) The FaceBase repository (www.facebase.org) ''Developing 3D Craniofacial Morphometry Data and Tools to Transform Dysmorphology, FB00000861'', collected at patient support groups in the USA, Canada, and the UK [3], [29].
2) The Western Australian Health Department. This collection is from the database of the Health Department

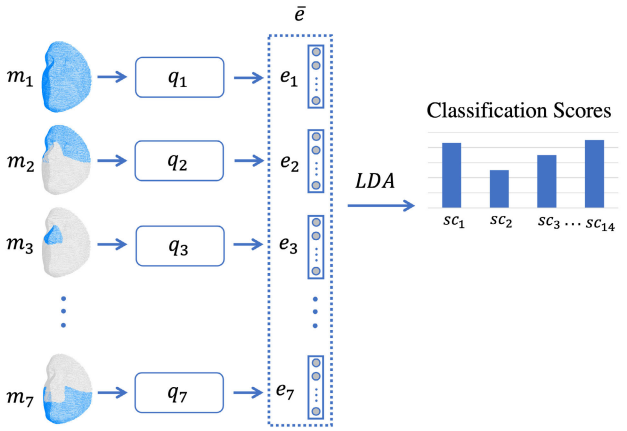



**FIGURE 1.** The syndrome classification scheme using a part-based approach. Lower dimensional embeddings ($e_i$) are learned for each facial segment ($m_i$) via function $q_i$ which can be replaced either by PCA (see Equation 4) or by a geometric encoder using a spiral convolutional operator (see Equation 5). Lower dimensional embeddings are concatenated ($\bar{e}$) and final classification supporting scores ($sc_i$) are obtained from linear discriminant analysis (LDA) classifier. The flow of the pipeline is provided in Algorithm 1.

**Algorithm 1** Part-Based Classification Pipeline

1. pre-process 3D Image: f = P(r) (Equation 1)
2. for i ranging from 1 to 7 (number of segments) do:
   encode input mesh: $e_i \leftarrow q_i(m_i)$
   if Baseline then:
   $q_i \leftarrow g_i(m_i)$ (Equation 4)
   else if GE then:
   $q_i \leftarrow h_i(m_i)$ (Equation 5)
   end if
   end for
3. concatenate embeddings: $\bar{e} \leftarrow (e_i||_{i=1}^{7})$
4. call classifier: $Z = [sc_1, sc_2, \ldots, sc_n] \leftarrow LDA(\bar{e})$

of Western 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. [28]

From these three collections combined, groups with >= 80 individuals were selected and only one image per person was included. Our original dataset is highly imbalanced with many groups containing images of a few individuals. Therefore, looking at the imbalance ratio of the dataset, defined as the ratio of the minimum and the maximum group size within the dataset, we selected the minimum group size such that the imbalance ratio remains below 1:3. To be precise, the imbalance ratio is 1:2.62 with the minimum group size of 80. Note that this selection was done prior to development and testing and was not further optimized based on the results. Approximately, 45%, 54% and <1% of the data used in this work are collected by the first, second, and the third listed source respectively. The demographic characteristics of the groups are shown in Table 1. This study was approved by the ethical review board of KU Leuven and University Hospitals Gasthuisberg, Leuven (S56392, S60568).

**TABLE 1.** Group number and names, sample size (N), mean and standard deviation of age (M±SD), and the female/male ratio (F/M) for each group.

| Synd. Nr. | Synd. Group | N | M ± SD | F/M |
|---|---|---|---|---|
| 1 | Control | 139 | 30.00 ± 10.83 | 0.74 |
| 2 | Angelman | 106 | 10.11 ± 7.62 | 0.48 |
| 3 | Bardet-Biedl (BBS) | 87 | 26.33 ± 14.78 | 0.48 |
| 4 | CHARGE | 87 | 13.73 ± 9.50 | 0.56 |
| 5 | 22q11.2 Del | 170 | 10.63 ± 6.06 | 0.49 |
| 6 | Cornelia de Lange | 182 | 12.15 ± 9.19 | 0.54 |
| 7 | Ehlers Danlos | 90 | 33.95 ± 18.76 | 0.86 |
| 8 | Marfan | 113 | 23.86 ± 16.48 | 0.59 |
| 9 | Noonan | 150 | 14.11 ± 12.67 | 0.46 |
| 10 | Prader Willi | 82 | 21.27 ± 13.30 | 0.48 |
| 11 | Smith Magenis | 128 | 14.14 ± 9.23 | 0.57 |
| 12 | Turner | 86 | 25.8 ± 19.30 | 1 |
| 13 | Williams | 215 | 17.81 ± 14.16 | 0.46 |
| 14 | Wolf Hirschhorn | 151 | 10.89 ± 9.38 | 0.56 |
| | Total | 1786 | 17.87 ± 14.28 | 0.59 |

### B. IMAGE PRE-PROCESSING

To apply spiral convolutional operators, meshes with a fixed topology are required. This fixed topology also allows the removal of extraneous (e.g., non-shape related) variation using techniques from statistical shape analysis, which potentially lowers the learning curve for the network. To accomplish this, a pre-processing routine P was applied to all raw facial scans R to generate meshes with fixed topology F:

$$P : R \rightarrow F \tag{1}$$

R and F are representing the same facial shape, however, while R has a random mesh representation, the output F is the structured mesh with the same topology across all faces. Pre-processing started by removing hair and ears, and indicating five positioning landmarks on each facial scan. These were used as input to a non-rigid 3D surface registration pipeline as implemented in MeshMonk [31] that gradually warped a generic facial template, comprising 7,160 vertices and 14,050 triangles into the shape of each target. This warped template, produced for each image, constitutes a representation of its shape, resampled to a standard topology or canonical mesh. Given this fixed topology, each image can be symmetrized and its position, orientation and size standardized; a generalized Procrustes analysis of all resampled images and their reflected copy was performed, followed by averaging each image and its reflected copy. Note that the latter operation

was facilitated by the bilaterally symmetrical constructed template [32] as provided in MeshMonk.

Meshes were then transferred to another template, which contains 8,321 vertices instead of 7,160 vertices and 16,384 triangles instead of 14,050 triangles. This template was introduced in [33] to facilitate equidistant 3D mesh down- and up-samplings, as mesh pooling operations in combination with the spiral convolutions. The 3D transformation between both templates was performed by an interpolation using 3D thin-plate spline Radial Basis Function . After preprocessing, each 3D face was described as a manifold triangle mesh:

$$F = (V, \mathcal{E}, \Phi) \tag{2}$$

where $V = \{v_i\}_{i=1}^{8,321}$ is a set of 8,321 3D vertices $v_i = (x_i, y_i, z_i)$ defining the mesh geometry, and $\mathcal{E}$ and $\Phi$ are set of edges and faces which define the mesh topology. $\mathcal{E}$ defines edges by an adjacency matrix where:

$$\varepsilon_{i,j} = \begin{cases} 1, & \text{where } v_i \text{ and } v_j \text{ are connected} \\ 0, & \text{otherwise} \end{cases} \tag{3}$$

and faces are defined by $\Phi = \{\varphi\}_{i=1}^{16,384}$ where $\varphi_i = \{v_t, v_p, v_q || \varepsilon_{t,p} = 1, \varepsilon_{t,q} = 1, \varepsilon_{p,q} = 1\}$ declares the vertices of each triangle in the template. Since all our meshes have the same topology as the template, $\mathcal{E}$ and $\Phi$ are fixed.

### C. LEARNING LOWER DIMENSIONAL EMBEDDINGS

#### 1) PRINCIPAL COMPONENT ANALYSIS (PCA)

PCA derives a low-dimensional latent space that is a linear subspace of the space spanned by all features (3D vertices), which optimally preserves the Euclidean distance among all observations (individual faces) [34]. Embeddings into the latent space of 3D vertices were obtained by selecting multiple principal components (PCs). To be more specific, consider $\tilde{F}$ containing all facial data reshaped to form an $n \times k$ matrix. Each row of $\tilde{F}$ is obtained by flattening $V$, and therefore $n = 1,786$ is the number of individuals, and $k = 8,321 \times 3$ . First, $\tilde{F}$ is column-mean centered. Then, PCs are calculated by singular value decomposition of $\tilde{F}$: $US\Lambda^T = \tilde{F}$, where S is a diagonal matrix of singular values $s$ in descending order of magnitude, $U$ contains left singular vectors and $\Lambda$ contains right singular vectors or PCs. Then, a flattened vector $\tilde{f}$ is projected to the space spanned by the PCs, and the corresponding embedding is then calculated by function g:

$$g : F \to E \ , \ e = g(f) = \tilde{f}.\Lambda \tag{4}$$

where $e_i$, the $i^{th}$ row of E, is the embedding of $i^{th}$ individual in the dataset. While the first PCs contain meaningful variations, the latter typically code for the noise in the data, and the variance explained by $n^{th}$ PC is measured by $\sigma_n^2 = \left(\frac{s_n}{\sqrt{n-1}}\right)^2$. When columns of $\Lambda$ are sorted in order of decreasing $\sigma^2$, the columns of $\Lambda$ define mutually orthogonal directions within the data that maximize the variance represented in the linear subspace.

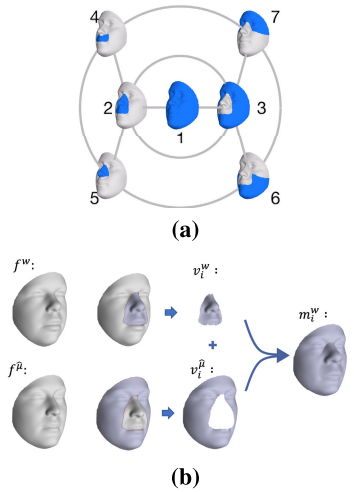



**FIGURE 2.** (a) The first three levels of the hierarchical segmentation of 3D facial shape. [22] (b) Mesh-padding: From the average face of all individuals with (for instance) Williams syndrome ($f^w$), the nose segment is padded with the average face of all individuals in the dataset ($f^{\hat{n}}$). Equation 6 explains the mesh-padding with more details.

#### 2) GEOMETRIC ENCODER

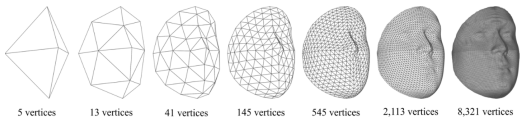
In our syndrome classification scenario, the metric space is learned by our GE such that the feature representations of patients within the same syndrome group are situated closer to each other than patients from a different syndrome group. Once trained, a GE can be represented by function h that maps an input mesh $f \in F$ to a low dimensional embedding $e \in E$:

$$h : F \to E \ , \ e = h(f) \tag{5}$$

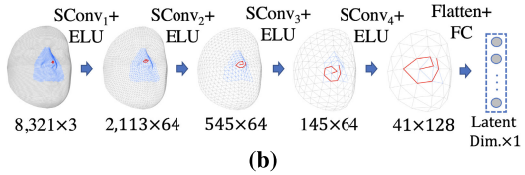
A triplet-loss network is a supervised deep metric learner that relates individuals in terms of such group membership [35], and consists of three identical subnetworks. Triplet networks are trained with triplets of the data comprising an anchor ($f_a$), positive ($f_p$) and negative sample ($f_n$). In 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 $(e_a, e_p, e_n) = (h(f_a), h(f_p), h(f_n))$. The loss function with which the triplet architecture was trained is defined as: $t = max(\| e_a - e_p \|_2^2 - \| e_a - e_n \|_2^2 + \alpha, 0)$ where $\alpha$ is the margin between paired positive and negative samples of the triplet, and according to [35] it is set to 0.2. Changing this parameter did not significantly change the outcomes (data not shown). Triplets were selected by a random triplet mining strategy from all possible triplets within a batch.

Choosing an appropriate architecture of the triplet subnetworks significantly affects the capacity of the embeddings to capture and learn the relevant information to separate individuals according to group membership. We use GDL to learn directly from the 3D facial meshes and efficiently leverage the underlying geometry by using spiral convolution operators, and for this reason we name the triplet subnetworks ''geometric encoders''. For each vertex, we convert its neigh-

borhood to an ordered sequence of vertices by applying a spiral scan as follows: first, an arbitrary direction is chosen in order to indicate the starting point of the spiral. Then, we follow a counterclockwise trajectory adding neighboring vertices sequentially (Figure 3b). Based on the fact that the one-ring neighbors of a vertex in the meshes cover a range from seven to ten vertices and also suggested by [19] and [20], the trajectory is truncated to nine vertices including the central vertex, i.e. the spiral length is set to nine. Larger spiral lengths were initially tested in a geometric autoencoder, and although the computation time was increased, no significant improvement in reconstruction performance was observed. Then, the ordered neighbor vertices are assigned the corresponding weights of the spiral convolution operator, i.e.: $\forall v \in V, h'(v) = \sum w_i^T h(S_i(v))$, 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 [20]. Since all the faces have the same fixed topology, the spirals were determined on the template mesh only once.




(a)




(b)

| Layer no. | Layer Type | Spiral length | In-Out Dimension | Number of Spirals |
|---|---|---|---|---|
| 1 | Sconv+ELU | 9 | 8,321 - 2,113 | 64 |
| 2 | Sconv+ELU | 9 | 2,113 - 545 | 64 |
| 3 | Sconv+ELU | 9 | 545 - 145 | 64 |
| 4 | Sconv+ELU | 9 | 145 - 41 | 128 |
| 5 | flatten+FC | latent dim. | 41-latent dim. | |

(c)

**FIGURE 3. (a) The result of the equidistant mesh sampling scheme introduced in [33]. (b) The visual architecture of a (part-based) geometric encoder. (c) The detailed architecture of a GE. Spiral length indicates the number of vertices defining the spirals. In-Out dimensions indicate the number of mesh vertices in each Sconv layer before and after down-sampling. Number of spirals defines the number of spiral filters in each Sconv layer.**

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. We, however, developed and published a 3D Mesh down- and up-sampling scheme that retains the property of equidistant mesh sampling in [33]. Starting from five initial points shown on the left in Figure 3a, 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 the average resolution of 2mm, meaning that the average edge length is 2mm. For our geometric encoder, the five highest levels of resolution 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. We chose equidistant sampling for the part-based approach, to ensure a consistent number of vertices in segments with similar size. The number of spirals in each layer (last column of the table provided in Figure 3C) was chosen empirically based on the previous and related works [22], [33], as well as in other in-house projects where similar facial data structures are used.

The architecture of our GE is illustrated in Figure 3b and the details of the model are reported in table provided in Figure 3c. The spiral convolutional (Sconv) layer in this table consists of first, convolving spirals on vertices of the mesh in the current layer, and second, down-sampling the current mesh to obtain input for the next layer. Each Sconv layer is followed by an exponential linear unit (ELU).

#### 3) FACIAL SEGMENTATION AND PART-BASED LEARNING

To exploit the multi-scale nature of facial variation, this work employed the hierarchical 3D facial surface segmentation proposed in [36], based on ∼ 8,000 3D facial images of individuals from an unselected/non-clinical European population and that were also processed with MeshMonk. The segmentation sequentially splits the vertices of the facial surface into smaller subsets by spectral clustering such that covariation within subsets is maximized and covariation between subsets is minimized. We refer to these subsets as segments. Here, we adopted the first three levels of this segmentation as first suggested in [22] and shown in Figure 2. Each segment $m$ is a subset of the full face, hence, is defined by $m = \{(v', \varepsilon', \phi') | \varepsilon' \subseteq \mathcal{E}, \phi' \subseteq \Phi\}$, where $\varepsilon'$ and $\phi'$ are fixed and predefined as shown in Figure 2a.

In the part-based approach, each 3D segment is processed separately. This means for every face, different embeddings are learned for each segment. Part-based PCA simply takes only the vertices of each segment into account, while part-based GEs were implemented using 3D mesh-padding. More specifically, the facial data for training a GE on segment $i$, which is noted as $m_i$, consists of the corresponding vertices of the segment $i$ padded with the vertices of the average face of all classes $m^{\hat{\mu}}$. This is shown in Figure 2b and defined as

$$m_i = \begin{cases} (v_i, \ \varepsilon_i, \ \phi_i) \\ \left(v^{\hat{\mu}}, \mathcal{E} \backslash \varepsilon_i, \ \Phi \backslash \phi_i\right) \end{cases} \tag{6}$$

### D. FULL PIPELINE

Starting from a raw image data $r$, the facial scan was pre-processed to generate the structured mesh representation:

$$1. f = P(r) \tag{7}$$

P is the pre-processing function (Equation 1), and f is the output mesh representation expressed in Equation 2. Next, each facial segment $m_i$ obtained from Equation 6 was encoded to a low dimensional embedding space by the encoding function $q : F \rightarrow E$:

$$2.\ e_i = q_i(m_i) \tag{8}$$

where $q$ is derived from Equation 4 for the baseline, or Equation 5 for the GE. Next, embeddings of all segments were concatenated (expressed by the concatenation operator.||) and passed to a syndrome classifier:

$$3.\ \bar{e} = (e_i||)_{i=1}^{7} \tag{9}$$

$$4.\ Z = q(\bar{e}) \tag{10}$$

where $\bar{e}$ is the $m \times d$ dimensional concatenated embeddings where m is the number of segments and d is the dimension of the embedding, Z is the $1 \times n$ dimensional vector containing classification scores for n syndrome groups, and g is the classification function that is replaced by LDA.

### E. TRAINING AND EVALUATION

#### 1) TRAINING

The GE blocks are trained for 600 epochs, when the validation loss plateaus, using the Adam optimizer, with a batch size of 60 (limited by the maximum GPU memory). The initial learning rate is set to 1e-4 according to experiments we ran with a range of (1e-1,1e-8), and a decay rate of 0.99 after each epoch. The models are implemented and trained on an NVIDIA GeForce RTX 2080 Ti using PyTorch 1.1.0.

#### 2) EVALUATION

To assess the non-linear metric learning based on a GDL architecture followed by LDA, we compared our results to those obtained using PCA followed by LDA as a baseline. We also compared classification based on embeddings learned from the full-face to those of the part-based GE and PCA. Taking into account the computation in training for each fold involved, the dataset was divided into a five-fold cross-validation. In each fold, 20% of the data from each group was randomly selected and devoted to the test set, and the remaining 80% was used for training the GE or PCA along with LDA.

One-vs-rest classification performance with LDA was used to evaluate the GE and PCA. For part-based models, LDA was trained on the embedding coordinates obtained from all facial segments concatenated into a single feature vector for each face. LDA was the chosen classifier in this work as it was the best performing model among multiple tested techniques including support vector machines, K- nearest neighbors, and a multi-layer perceptron (data not shown).

Since the group sizes in the test dataset are highly imbalanced, it is important to consider classification metrics that are more robust to the group size imbalance. Therefore, to compare the overall performance of models, the classification measures reported or referred to throughout this paper, are:

- Sensitivity: measure of how well the classifier can identify true positives;
- Specificity: measure of how well the classifier can identify true negatives;
- Balanced accuracy: the mean of sensitivity and specificity;
- Area under the precision-recall curve (PR-AUC): indicator of both recall and precision, where high precision relates to a low false positive rate, and high recall relates to a low false negative rate;
- F1 score: harmonic mean of precision and recall;
- Adjusted rand index (ARI): a measure of agreement between the results obtained by a clustering (or classification) process and the results defined by external criteria (in our case ground truth labels) [37]. As supported in [38], we incorporated ARI as a classification measure in this work.

### F. EXPERIMENTS

#### 1) LATENT SPACE DIMENSIONALITY

The number of dimensions retained in a lower dimensional embedding space can affect the classification accuracy. Therefore, to examine this, we started with a low dimensionality of 4 and increased the dimensions to 14, 24, 35, 47, and 57. These numbers explain up to 68%, 90%, 95%, 97%, 98%, and 98.51%, respectively, of the variation in the full face using PCA, and we compared them to geometric embeddings retaining the same number of dimensions. Furthermore, for consistency, the same numbers of dimensions were used for each part-based encoder in the part-based setup.

#### 2) CONTRIBUTION OF PART-BASED LEARNING

The contribution of part-based learning is assessed in multiple ways. In the first instance, we contrasted classification evaluation metrics based on only the embeddings of the full face, to those based on the embedding from all facial segments. In the second instance, we investigated the contribution of individual segments. For part-based learning to be of value, the metric spaces of individual segments should contain information complementary to that of other segments. We assessed this as follows: the similarity among the 7 different metric spaces was summarized as the $7 \times 7$ similarity matrix of RV coefficients between each pair of distance matrices, transformed into cross-product matrices as per [39]. The eigenvectors of this similarity matrix define a similarity space where distance represents the difference among metric spaces. The first eigenvector represents information that is common among the metric spaces: those spaces with larger projections onto this eigenvector contain more common information and those with smaller projections contain less [39]. We used this to assess the relative amount of information contained in one metric space that is not contained in others. The expectation was that those metric spaces with lower projections are especially valuable to the

**TABLE 2.** One-vs-all classification metrics (balanced accuracy, sensitivity, specificity, and F1 score) for full-face (FF) or part-based (PB) geometric encoder (GE) and principal component analysis (PCA), for increasing dimensions (DIMs). Reported numbers indicate mean ±standard deviation of 14 syndrome groups.

| | BALANCED ACCURACY | | | | SENSITIVITY / SPECIFICITY | | | | F1 SCORE | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| DIM | PCA FF | PCA PB | GE FF | GE PB | PCA FF | PCA PB | GE FF | GE PB | PCA FF | PCA PB | GE FF | GE PB |
| 4 | 0.7 ± 0.047 | 0.82 ± 0.036 | 0.81 ± 0.037 | 0.87 ± 0.033 | 0.71 ± 0.092<br>0.69 ± 0.023 | 0.82 ± 0.075<br>0.83 ± 0.019 | 0.83 ± 0.071<br>0.78 ± 0.036 | 0.86 ± 0.066<br>0.88 ± 0.018 | 0.25 ± 0.03 | 0.42 ± 0.032 | 0.37 ± 0.04 | 0.5 ± 0.038 |
| 14 | 0.7 ± 0.038 | 0.82 ± 0.033 | 0.87 ± 0.035 | 0.87 ± 0.031 | 0.71 ± 0.079<br>0.69 ± 0.022 | 0.82 ± 0.072<br>0.83 ± 0.017 | 0.86 ± 0.066<br>0.88 ± 0.02 | 0.86 ± 0.063<br>0.88 ± 0.014 | 0.25 ± 0.032 | 0.42 ± 0.041 | 0.51 ± 0.049 | 0.5 ± 0.044 |
| 24 | 0.83 ± 0.037 | 0.89 ± 0.035 | 0.87 ± 0.036 | 0.88 ± 0.04 | 0.83 ± 0.075<br>0.84 ± 0.017 | 0.84 ± 0.07<br>0.94 ± 0.016 | 0.85 ± 0.072<br>0.89 ± 0.015 | 0.84 ± 0.082<br>0.93 ± 0.013 | 0.43 ± 0.035 | 0.63 ± 0.053 | 0.53 ± 0.041 | 0.61 ± 0.05 |
| 35 | 0.86 ± 0.029 | 0.88 ± 0.035 | 0.88 ± 0.028 | 0.88 ± 0.036 | 0.85 ± 0.062<br>0.87 ± 0.015 | 0.82 ± 0.068<br>0.94 ± 0.014 | 0.86 ± 0.06<br>0.9 ± 0.016 | 0.82 ± 0.073<br>0.94 ± 0.013 | 0.48 ± 0.031 | 0.64 ± 0.059 | 0.55 ± 0.039 | 0.62 ± 0.045 |
| 47 | 0.87 ± 0.031 | 0.88 ± 0.037 | 0.88 ± 0.036 | 0.87 ± 0.037 | 0.86 ± 0.065<br>0.89 ± 0.017 | 0.81 ± 0.073<br>0.95 ± 0.012 | 0.86 ± 0.068<br>0.91 ± 0.015 | 0.8 ± 0.076<br>0.94 ± 0.011 | 0.52 ± 0.036 | 0.65 ± 0.057 | 0.57 ± 0.047 | 0.63 ± 0.047 |
| 57 | 0.88 ± 0.033 | 0.87 ± 0.037 | 0.88 ± 0.039 | 0.87 ± 0.039 | 0.86 ± 0.067<br>0.9 ± 0.016 | 0.8 ± 0.075<br>0.95 ± 0.013 | 0.85 ± 0.08<br>0.91 ± 0.015 | 0.79 ± 0.078<br>0.95 ± 0.013 | 0.55 ± 0.039 | 0.65 ± 0.053 | 0.58 ± 0.046 | 0.64 ± 0.049 |

part-based classification as they contain information that is not contained in metric spaces derived from other segments.

We further assessed the unique contribution of each segment by removing the embedding derived from the segment and re-training the LDA classifier. We used the ARI to summarize the agreement between true and predicted labels. ARI values close to zero indicate chance performance, and values of one indicate perfect classification.

#### 3) ONE-VS-ALL GEOMETRIC ENCODER

So far, the GEs were trained in a multi-class setup. This means, the network is trained using all group labels at once, and one output vector of various dimensions is generated. In comparison, we also trained a two-class GE for each syndrome group in a one-vs-all instead of a many-vs-many setup. For simplicity, we refer this model as "binary GE". Each binary GE has one dimensional output, thus once concatenated, for each segment, the embedding space of all syndromes will be a 14-dimensional vector. We will then compare the PR-AUC based on this experiment with those from a 14 dimensional multi-class GE and PCA.

## III. RESULTS

### A. CLASSIFICATION PERFORMANCE OF PART-BASED GE AND PCA

The average balanced accuracy, sensitivity, specificity, and F1 score for each dimension are provided in Table 2. This table reports results of the full-face and the part-based models for both PCA and GE. To compare models, the PR-AUC is plotted for increasing dimensions in Figure 4. It indicates that, for the full-face approach, GE outperforms PCA, although the discrepancy becomes smaller as the number of dimensions increases. The part-based approach increases the performances in lower dimensionalities for both GE and PCA while the improvement is the largest for PCA. In higher dimensions, however, the performance gain decreases for both approaches. According to this figure, the optimal classification performance is achived by 24-dimensional part-based approaches. The classification measures per syndrome

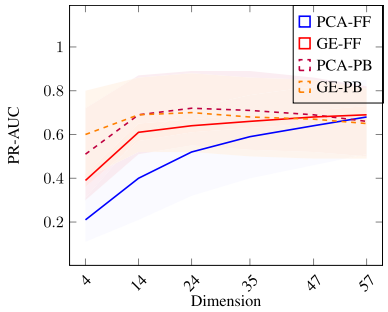



**FIGURE 4.** Average area under precision-recall curve (PR-AUC) as a function of embedding dimensions for full-face (FF) and part-based (PB) principal component analysis (PCA) and geometric encoder (GE). Shadow around each line indicates standard deviation.

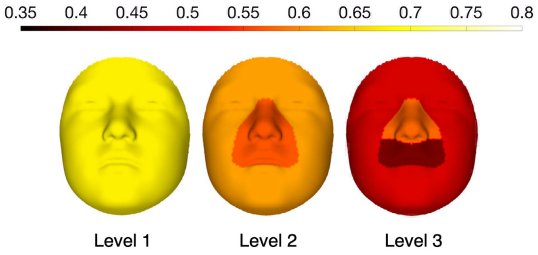



**FIGURE 5.** Part-based heatmap of classifying 22q11_2 deletion syndrome. Each segment in the first, second, and third levels of the hierarchical segmentation is colored by the PR-AUC of classification based on the segment's embedding.

for 24-dimensional part-based approaches are available in Table S.1 in supplementary material.

### B. CLASSIFICATION PERFORMANCE OF BINARY GE

Figure 6 shows the PR-AUC per syndrome obtained from each binary GE, multi-class GE, and PCA, all trained with a part-based setup with the optimal embedding dimension (24) per segment. Differences in performance among the binary implementation and multi-class GE and PCA were compared statistically with pairwise comparisons using a paired two-tailed Wilcoxon signed rank test. The PR-AUC of the

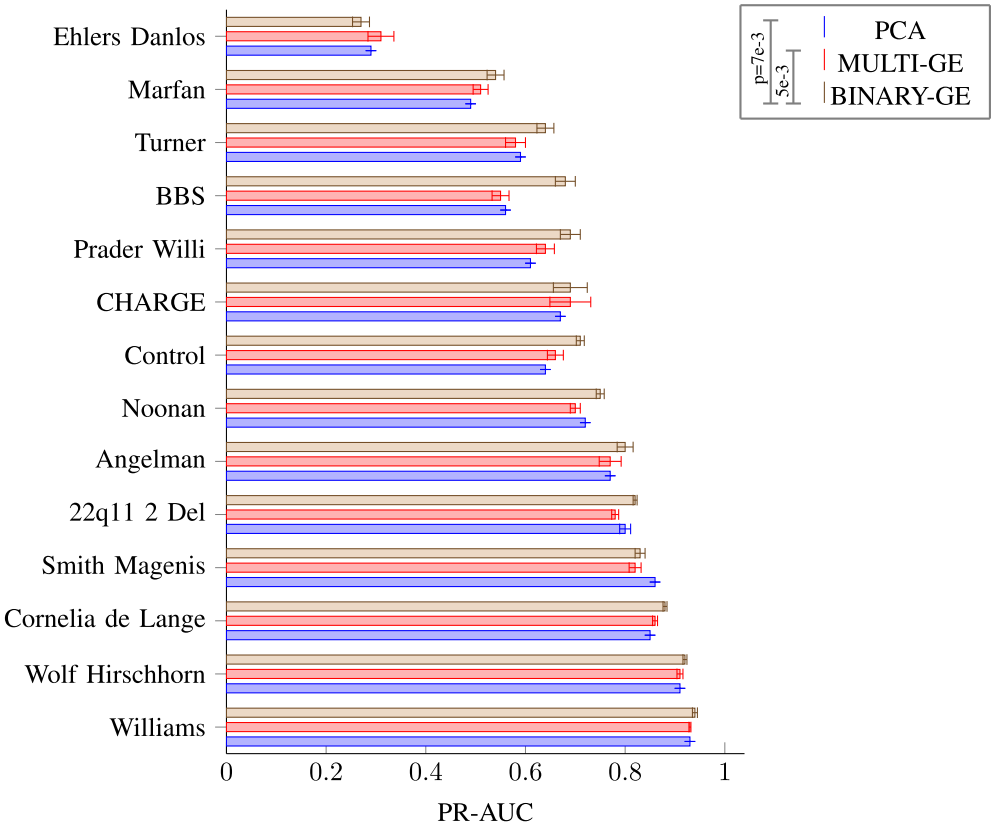



**FIGURE 6.** Comparison between the performance of the Binary GE with multiclass GE and PCA, all in part-based setup with 24 dimensions per segment. Error bars indicate the standard error of the mean.

binary GE is significantly higher than that of the multi-class GE and part-based PCA (p-values $< 0.05$).

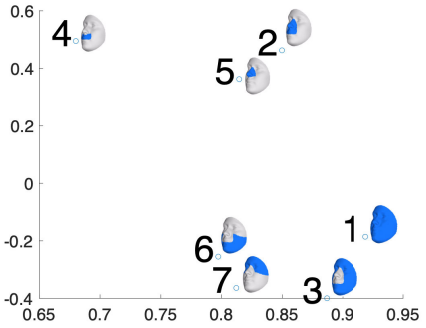
### C. INFLUENCE OF INDIVIDUAL SEGMENTS

Figure 7a displays the first two dimensions of a similarity space comparing the metric spaces derived from individual segments. Each point represents a metric space learned from a particular segment and projections on the first dimension (plotted on the horizontal axis) inversely represents the amount of information represented in that space that is not shared among other spaces [39]. Table 7b shows the result of an ablation study as ARI computed from predicted labels and the ground truth labels. Both of these analyses show that in the context of the hieracrchical segmentation, the metric space of the full face contains the least unique information for discriminating syndromes in this dataset while segment 4 comprising the vermillion and cutaneous upper lip contains the most. This is expected since the full face overlaps the most with all the other segments, and hence has little extra to contribute.

Aside from an increased classification performance, another advantage of the part-based classification is gaining the opportunity to quantitatively and qualitatively measure the contribution of each facial segment in the final classification of each syndrome. The latter is highly interesting since it allows us to compare the facial segments that contribute to the automated classification with the clinical features that are used by expert clinicians to recognize a particular syndrome. E.g., Figure 5 illustrates the clinically-relevant visual representation of the relative contribution of different facial parts for the first three levels of the hierarchical segmentation. The color of each segment is indexed to the PR-AUC computed from the one-vs-all classification of 22q11_2 deletion syndrome. While there is wide inter-individual variability in the presence of mild dysmorphic features in 22q11_2 deletion, the most consistent feature is a tubular nose with underdeveloped alae nasi and a round nasal tip. In alignment with this, the heatmaps in the first, second, and third levels of Fig. 5 indicate the relatively higher contribution of the nasal segment in the automated classification.

## IV. DISCUSSION

Deep CNNs for multi-class syndrome classification based on 2D photographs are now being routinely used in many clinics [1], [2]. 2D images are readily available but encode only indirectly information about 3D shape. By now, large-scale databases of 3D photographs of clinical populations have been collected, although compared to the 2D datasets of genetic syndromes used in the literature [1], available 3D datasets are still considerably (about ten times) smaller in size and diversity. With increasingly inexpensive and portable 3D imaging hardwares, this imaging modality is expected to grow in popularity and use. Furthermore, recently developed GDL techniques now allow CNNs to be deployed




(a)

| Removed segment | Fold 1 | Fold 2 | Fold 3 | Fold 4 | Fold 5 | **Mean** | STD |
|---|---|---|---|---|---|---|---|
| / | 0.652 | 0.629 | 0.688 | 0.635 | 0.664 | **0.654** | 0.0244 |
| seg 1 | 0.525 | 0.568 | 0.567 | 0.572 | 0.588 | **0.564** | 0.0234 |
| seg 2 | 0.397 | 0.396 | 0.38 | 0.354 | 0.407 | **0.387** | 0.021 |
| seg 3 | 0.489 | 0.516 | 0.525 | 0.473 | 0.529 | **0.506** | 0.024 |
| seg 4 | 0.279 | 0.245 | 0.255 | 0.276 | 0.287 | **0.268** | 0.018 |
| seg 5 | 0.336 | 0.377 | 0.347 | 0.322 | 0.378 | **0.352** | 0.025 |
| seg 6 | 0.345 | 0.317 | 0.341 | 0.325 | 0.352 | **0.336** | 0.015 |
| seg 7 | 0.417 | 0.386 | 0.403 | 0.393 | 0.398 | **0.399** | 0.012 |

(b)

**FIGURE 7.** (a) The projections of the seven segments onto the first and second eigenvectors of the similarity matrix. This compares the metric spaces learned from different facial segments. Projections on the first eigenvector inversely index the agreement between the metric space of the segment and the other metric spaces. (b) Adjusted rand index (ARI) between the ground truth and obtained labels from the LDA classifier. Results indicate ARI based on the concatenated embeddings (24-dimensional embedding per segment). In each row, the embedding that is referred to is excluded.

directly onto 3D images. To tackle the data size limitation, we 1) pre-processed and standardized the data using MeshMonk, 2) Implemented the part-based approach to divide the learning between multiple GEs, and 3) trained our GEs in a triplet-based setup, so that the models learn from different combinations of data samples.

In this study we apply GDL on 3D facial meshes for a multi-class syndrome classification. We developed a GDL-based metric learner to learn metric spaces that separate syndrome classes. We compared classification performance to that obtained using latent spaces defined by PCA. To exploit the multi-scale nature of facial variation, this was further expanded into a part-based framework where metric spaces are defined for each facial segment of the hierarchical global-to-local facial segmentation. We compared classification performance from embeddings on the full-face only to those using all segments and further assessed the individual contribution of embeddings defined for each facial segment. Additionally, we assessed the performance of the geometric part-based setup when trained from labels of all classes in a single part-based metric learner (Multiclass-GE) compared to training a separate part-based metric learner for each syndrome using positive (belonging to the class) vs all other (Binary-GE). In all classification experiments, LDA was used as a classifier for a one-vs-all classification.

GE significantly outperforms PCA in lower dimensonality, which was expected. This discrepancy becomes smaller as the number of dimensions increases. This is mainly because GEs are trained in a supervised way (in contrast to PCA), therefore, the embeddings are trained to preserve the most relevant information to the discrimination task. This is likely to include localized facial features only common to a small proportion of the training set (e.g., to one particular syndrome). PCA ultimately learns these also, but it takes additional dimensions to do so, because it only aims to preserve the maximum amount of variation across the entire training dataset. Therefore, its first dimensions will capture broad patterns of facial variation and population structure and not specific discriminative facial features. It is important to note that for static facial shape analysis, PCA was and still is a strong and popular approach. The part-based implementation clearly improved the performance of PCA, and this close to the level of the GE part-based equivalent. In applications where clear non-rigid variations are concerned (e.g. facial expressions and body movements), however, illustrating the superiority of neural models is more straightforward [19], [40]. Moreover, training a single full-face GE, trained on an NVIDIA GeForce RTX 2080 Ti using PyTorch 1.1.0 takes 18 minutes and 25 seconds, while the PCA method takes 13 seconds in MATLAB 2020, on the same machine. Although the time complexity is significantly higher for the deep learning approach, we should keep in mind that this complexity is reduced to training time only and once a trained GE is loaded, it takes less than 3 seconds to embed all the faces. Figure S.2 of supplementary material depicts a 2D visualization of the 14 dimensional part-based PCA and GE projections of the train set (smaller dots) and test set (larger dots) using t-SNE [41]. This plot demonstrates that the GE learns a better structured metric space in which distinct syndromes reside in clusters.

The part-based approach yields better performance than the full face in spaces of smaller dimensionality for both GE and PCA while the improvement is the largest for PCA. With higher dimensionalities, however, the performance of the part-based approach decreases for both GE and PCA. This suggests that with a part-based scheme at 24 dimensions per segment, performance is ceiling and models with more dimensionality are overfitting. Using the part-based model, localized information of the face is forcefully added into the model. For PCA, this means that, even with low dimensionality, localized shape variation remains available, that is otherwise lost in the context of the full face only. For the PCA model, adding segments to the analysis has a similar effect as increasing the dimensions of the full-face model. Although the performance of the part-based PCA and the part-based GE are close, and that the improvement of the latter is marginal, the gain in using the part-based approach is still notable for the GE, suggesting that discriminative facial

variation is represented compactly by both approaches when using a part-based scheme.

The part-based Binary GEs trained for individual syndromes outperformed both the part-based PCAs and multi-class GEs. With more data and resources available, this model can potentially be prioritized over multi-class GE. However, it is important to note that for a binary GE with $d$ dimensional embedding size, the final concatenated embedding dimension of each input would be $d \times 7 \times 14$ (for 7 segments and 14 classes). In contrast, in a multi-class GE, the final concatenated embedding dimension for 7 segments would be $d \times 7$. This means that the binary GE is less flexible in terms of the embedding compactness. This constraint together with the longer training time and more memory usage, make binary GEs more difficult to implement and to use in practice, particularly for a larger-scale syndrome classification problem.

We assessed the contribution of metric spaces of different facial segments learned by the GE. Segment 1, comprising the full face, and segment 3, comprising all of the face but the nose and upper lip have the smallest unique contribution. This is not surprising as, given these are large facial regions and at the top level of the hierarchy, much of the discriminating information they contain is also contained in their sub-segments at lower levels of the hierarchy. However, each segment contains some unique information that is not contained in their sub-segments, demonstrating the value of a multi-scale, hierarchical approach to facial segmentation, as opposed to simply dividing the face into mutually exclusive parts (e.g., at a single level of the hierarchy). Segment 4 has the largest unique contribution to classification. This comprises the cutaneous and vermillion upper lip. The upper lip develops early in embryogenesis and abnormalities in this area are associated with a wide variety of developmental disorders. Further, as shown in [42], the correlation between genomic signals of the upper lip with other segments on the third level of hierarchical segmentation is relatively small. As such this region appears relatively independent and its development is especially sensitive to disturbances caused by genetic anomalies, explaining why it may be especially useful in discriminating among syndromes. Using the part-based approach, it is also possible to give feedback to clinicians about the relative contribution of different facial segments to the classifications of a patient into a syndrome, which helps to explain the decision making of the network.

The age and sex imbalance of the dataset may have induced biases in the learning. However, our embedding spaces are learned by random combinations of (anchor, positive, negative) triplets. As such, during training, anchors will be exposed to negative instances with the same demographic characteristics and positive instances with different demographic characteristics, reducing the chance that the network will incorrectly learn facial features associated with demographics as characteristic features of the syndrome. For further assurance, we trained a full-face GE with an age- and sex-corrected dataset using linear regression methods to compare the results, and no significant changes were observed (Table S.3 of supplementary material).

The clinical state-of-the art is Face2Gene-Clinic (F2G-C) which is pretrained on hundreds of syndromes and thousands of images. To compare the 3D approach with F2G-C, computer graphics softwares are used to render 2D images from the 3D facial surfaces. However, given the low-resolution and inferior texture quality of the 3D facial images available, these rendered images remained far from photorealistic. In addition, for many of the 3D images available, only shape without texture or color information was available. Considering these limitations, such a comparison is not fair towards the 2D approach. However, a direct comparison with the 3D state-of-the-art technique published in [3] is feasible. To do so, 65 landmarks were selected on all our faces, as done in [3]. Then, regularized LDA is performed in a binary (one-vs-rest) set up, on the vector of $65 \times 3$ dimensional input data. The average over five data folds of balanced accuracy, sensitivity, specificity and F1 score are $0.78 \pm 0.069$, $0.58 \pm 0.135$,$0.98 \pm 0.006$, and $0.62 \pm 0.135$ respectively. In comparison, we look at the highest dimension of the full-face analysis that we performed (57 dimensions). The results show that the balanced accuracy for 57 dimensional PCA as well as GE (both as input to LDA) which are based on the dense facial meshes are 10% higher than the 3D state-of-the-art with 65 facial landmarks.

## V. CONCLUSION

In this work, we proposed a 3D part-based GDL model as an assisting tool for identifying candidate disorders based on facial shape from 3D surface images. First, we introduced a geometric encoder (GE) compared to PCA as input to LDA, where the former generated a clear improvement. Second, we proposed a part-based implementation to 3D facial shape analysis and multiclass syndrome classification, and this applied to both GE and PCA. The comparison between part-based versus holistic (or full face) approaches indicated substantial improvements. In addition, we are able to provide localized feedback on the contribution of each facial segment in the syndrome classification of a patient's image which aids clinicians in their assessment of the result. Lastly, based on ablation studies within the part-based approach, we investigated which facial segment stored the most unique information. This work is a collection of techniques found in the literature, and therefore, individual contributions to each of the components can further increase the work. To be more specific, future work includes optimizing the hierarchical facial segmentation using multi-task learning methods, instead of incorporating a previously obtained data-driven segmentation. Moreover, since many syndromes are extremely rare and hence their group size will always remain small, learning from highly imbalanced data techniques will become important. Lastly, other existing geometric deep learning methods, aside from the spiral convolutions used in this work, such as PointNet++ [43] are of interest to explore since they 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.

## VI. CODE AVAILABILITY

The necessary code for reproducing the results based on the low dimensional embeddings and the trained models are available at https://github.com/sohamh/Multi-Scale-Part-Based-Syndrome-Classification.git. Low dimensional embeddings are reproducible for the portion of the data that is publicly available (The FaceBase repository (www.facebase.org)).

## REFERENCES

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

[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, Jun. 2014, Art. no. e02020.

[3] B. Hallgrímsson *et al.*, "Automated syndrome diagnosis by three-dimensional facial imaging," *Genet. Med.*, vol. 22, pp. 1682–1693, Oct. 2020.

[4] D. L. Narayanan, P. Ranganath, S. Aggarwal, A. Dalal, S. R. Phadke, and K. Mandal, "Computer-aided facial analysis in diagnosing dysmorphic syndromes in Indian children," *Indian Pediatrics*, vol. 56, no. 12, pp. 1017–1019, Dec. 2019.

[5] J. T. Pantel, N. Hajjir, M. Danyel, J. Elsner, A. T. Abad-Perez, P. Hansen, S. Mundlos, M. Spielmann, D. Horn, C.-E. Ott, and M. A. Mensah, "Efficiency of computer-aided facial phenotyping (DeepGestalt) in individuals with and without a genetic syndrome: Diagnostic accuracy study," *J. Med. Internet Res.*, vol. 22, no. 10, Oct. 2020, Art. no. e19263.

[6] M. Quinto-Sánchez *et al.*, "Developmental pathways inferred from modularity, morphological integration and fluctuating asymmetry patterns in the human face," *Sci. Rep.*, vol. 8, no. 1, pp. 963–978, Dec. 2018.

[7] 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 Reconstructive Surg.*, vol. 146, no. 6, pp. 1407–1417, Dec. 2020.

[8] S. Fang, J. McLaughlin, J. Fang, J. Huang, I. Autti-Rämö, Å. Fagerlund, S. Jacobson, L. Robinson, H. Hoyme, S. Mattson, E. Riley, F. Zhou, R. Ward, E. Moore, and T. Foroud, "Automated diagnosis of fetal alcohol syndrome using 3D facial image analysis," *Orthodontics Craniofacial Res.*, vol. 11, no. 3, pp. 162–171, Aug. 2008.

[9] P. Hammond, T. J. Hutton, J. E. Allanson, B. Buxton, L. E. Campbell, J. Clayton-Smith, D. Donnai, A. Karmiloff-Smith, K. Metcalfe, K. C. Murphy, M. Patton, B. Pober, K. Prescott, P. Scambler, A. Shaw, A. C. M. Smith, A. F. Stevens, I. K. Temple, R. Hennekam, and M. Tassabehji, "Discriminating power of localized three-dimensional facial morphology," *Amer. J. Hum. Genet.*, vol. 77, no. 6, pp. 999–1010, Dec. 2005.

[10] G. de Jong, E. Bijlsma, J. Meulstee, M. Wennen, E. van Lindert, T. Maal, R. Aquarius, and H. Delye, "Combining deep learning with 3D stereophotogrammetry for craniosynostosis diagnosis," *Sci. Rep.*, vol. 10, no. 1, pp. 15346–15352, Dec. 2020.

[11] Z. Wu, S. Song, A. Khosla, F. Yu, L. Zhang, X. Tang, and J. Xiao, "3D ShapeNets: A deep representation for volumetric shapes," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Boston, MA, USA, Jun. 2015, pp. 1912–1920.

[12] D. Maturana and S. Scherer, "VoxNet: A 3D convolutional neural network for real-time object recognition," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Hamburg, Germany, Sep./Oct. 2015, pp. 922–928.

[13] R. Q. Charles, H. Su, M. Kaichun, and L. J. Guibas, "PointNet: Deep learning on point sets for 3D classification and segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Honolulu, HI, USA, Jul. 2017, pp. 77–85.

[14] M. M. Bronstein, J. Bruna, Y. LeCun, A. Szlam, and P. Vandergheynst, "Geometric deep learning: Going beyond Euclidean data," *IEEE Signal Process. Mag.*, vol. 34, no. 4, pp. 18–42, Jul. 2017.

[15] I. Lim, A. Dielen, M. Campen, and L. Kobbelt, "A simple approach to intrinsic correspondence learning on unstructured 3D meshes," in *Computer Vision—ECCV 2018 Workshops* (Lecture Notes in Computer Science), vol. 11131, L. Leal-Taixé and S. Roth, Eds. Cham, Switzerland: Springer, 2019, pp. 349–362.

[16] J. Bruna, W. Zaremba, A. D. Szlam, and Y. LeCun, "Spectral networks and locally connected networks on graphs," in *Proc. ICLR*, 2014, pp. 1–14.

[17] M. Defferrard, X. Bresson, and P. Vandergheynst, "Convolutional neural networks on graphs with fast localized spectral filtering," in *Proc. 30th Int. Conf. Neural Inf. Process. Syst. (NIPS)*, Barcelona, Spain, Dec. 2016, pp. 3844–3852.

[18] F. Monti, D. Boscaini, J. Masci, E. Rodola, J. Svoboda, and M. M. Bronstein, "Geometric deep learning on graphs and manifolds using mixture model CNNs," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Honolulu, HI, USA, Jul. 2017, pp. 5425–5434.

[19] 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 *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Seoul, South Korea, Oct. 2019, pp. 7212–7221.

[20] S. Gong, L. Chen, M. Bronstein, and S. Zafeiriou, "SpiralNet++: A fast and highly efficient mesh convolution operator," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshop (ICCVW)*, Seoul, South Korea, Oct. 2019, pp. 4141–4148.

[21] D. Kulon, R. A. Guler, I. Kokkinos, M. M. Bronstein, and S. Zafeiriou, "Weakly-supervised mesh-convolutional hand reconstruction in the wild," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Seattle, WA, USA, Jun. 2020, pp. 4989–4999.

[22] S. S. Mahdi, N. Nauwelaers, P. Joris, G. Bouritsas, S. Gong, S. Walsh, M. D. Shriver, M. Bronstein, and P. Claes, "Matching 3D facial shape to demographic properties by geometric metric learning: A part-based approach," *IEEE Trans. Biometrics, Behav., Identity Sci.*, early access, Jun. 29, 2021, doi: 10.1109/TBIOM.2021.3092564.

[23] E. Hoffer and N. Ailon, "Deep metric learning using triplet network," in *Similarity-Based Pattern Recognition* (Lecture Notes in Computer Science), A. Feragen, M. Pelillo, and M. Loog, Eds. Cham, Switzerland: Springer, 2015, pp. 84–92.

[24] A. Hermans, L. Beyer, and B. Leibe, "In defense of the triplet loss for person re-identification," Nov. 2017, *arXiv:1703.07737*.

[25] J. Lee, N. J. Bryan, J. Salamon, Z. Jin, and J. Nam, "Metric learning vs classification for disentangled music representation learning," Aug. 2020, *arXiv:2008.03729*.

[26] O. Ocegueda, S. K. Shah, and I. A. Kakadiaris, "Which parts of the face give out your identity?" in *Proc. CVPR*, Colorado Springs, CO, USA, Jun. 2011, pp. 641–648.

[27] V. Blanz and T. Vetter, "A morphable model for the synthesis of 3D faces," in *Proc. 26th Annu. Conf. Comput. Graph. Interact. Techn. (SIGGRAP)*, 1999, pp. 187–194.

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

[29] O. Klein, W. Mio, R. Spritz, and B. Hallgrimsson, "Developing 3D craniofacial morphometry data and tools to transform dysmorphology," FaceBase, Dataset FB00000861, 2019. [Online]. Available: http://www.facebase.org, doi: 10.25550/TJ0.

[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*, vol. 22, J. Zschocke, M. Baumgartner, E. Morava, M. Patterson, S. Rahman, and V. Peters, Eds. Berlin, Germany: Springer, 2015, pp. 99–106.

[31] J. D. White, A. Ortega-Castrillón, H. Matthews, A. A. Zaidi, O. Ekrami, J. Snyders, Y. Fan, T. Penington, S. Van Dongen, M. D. Shriver, and P. Claes, "MeshMonk: Open-source large-scale intensive 3D phenotyping," *Sci. Rep.*, vol. 9, no. 1, pp. 6085–6096, Dec. 2019.

[32] O. Ekrami, P. Claes, J. D. White, A. A. Zaidi, M. D. Shriver, and S. Van Dongen, "Measuring asymmetry from high-density 3D surface scans: An application to human faces," *PLoS ONE*, vol. 13, no. 12, Dec. 2018, Art. no. e0207895.

[33] 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 *Proc. 25th Int. Conf. Pattern Recognit. (ICPR)*, Milan, Italy, Jan. 2021, pp. 1757–1764.

[34] K. Pearson, "LIII. On lines and planes of closest fit to systems of points in space," *London, Edinburgh, Dublin Phil. Mag. J. Sci.*, vol. 2, no. 11, pp. 559–572, Nov. 1901.
[35] F. Schroff, D. Kalenichenko, and J. Philbin, "FaceNet: A unified embedding for face recognition and clustering," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Boston, MA, USA, Jun. 2015, pp. 815–823.
[36] P. Claes *et al.*, "Genome-wide mapping of global-to-local genetic effects on human facial shape," *Nature Genet.*, vol. 50, no. 3, pp. 414–423, Mar. 2018.
[37] N. X. Vinh, J. Epps, and J. Bailey, "Information theoretic measures for clusterings comparison: Variants, properties, normalization and correction for chance," *J. Mach. Learn. Res.*, vol. 11, no. 95, pp. 2837–2854, 2010.
[38] J. M. Santos and M. Embrechts, "On the use of the adjusted RAND index as a metric for evaluating supervised classification," in *Artificial Neural Networks—ICANN 2009* (Lecture Notes in Computer Science), vol. 5769, C. Alippi, M. Polycarpou, C. Panayiotou, and G. Ellinas, Eds. Berlin, Germany: Springer, 2009, pp. 175–184.
[39] H. Abdi, A. J. O'Toole, D. Valentin, and B. Edelman, "DISTATIS: The analysis of multiple distance matrices," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, San Diego, CA, USA, Sep. 2005, p. 42.
[40] R. A. Potamias, J. Zheng, S. Ploumpis, G. Bouritsas, E. Ververas, and S. Zafeiriou, "Learning to generate customized dynamic 3D facial expressions," in *Computer Vision—ECCV 2020* (Lecture Notes in Computer Science), vol. 12374, A. Vedaldi, H. Bischof, T. Brox, and J.-M. Frahm, Eds. Cham, Switzerland: Springer, 2020, pp. 278–294.
[41] L Van der Maaten and G. Hinton, "Visualizing data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, no. 86, pp. 2579–2605, 2008.
[42] J. D. White *et al.*, "Insights into the genetic architecture of the human face," *Nature Genet.*, vol. 53, pp. 45–53, Jan. 2021.
[43] C. R. Qi, L. Yi, H. Su, and L. J. Guibas, "PointNet++: Deep hierarchical feature learning on point sets in a metric space," in *Proc. 31st Int. Conf. Neural Inf. Process. Syst.*, Red Hook, NY, USA, Dec. 2017, pp. 5105–5114.

**SOHA SADAT MAHDI** received the M.S. degree in artificial intelligence from KU Leuven, Belgium. She is currently a Doctoral Researcher with KU Leuven. Her research interests include computer-aided diagnosis of genetic syndromes, and more specifically applications of geometric deep learning in analyzing craniofacial malformations associated with genetic syndromes.

**HAROLD MATTHEWS** received the Ph.D. degree from the Department of Pediatrics, University of Melbourne, Australia, in 2018. As a Postdoctoral Researcher at KU Leuven, he applies statistical shape analysis to 3-D meshes for assessing facial abnormality and understanding diseases affecting craniofacial development.

**NELE NAUWELAERS** received the master's degree in mathematical engineering from KU Leuven, in 2018. After her studies, she started a Ph.D. research with the Medical Imaging Research Center located at the university hospital. In her research, she applies state-of-the-art geometric deep learning techniques to shapes represented on 3-D meshes, with a highlighted interest in encoding facial shape in a biologically meaningful way.

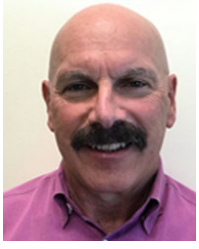
**MICHIEL VANNESTE** received the Graduate degree from the Medical School, KU Leuven, Belgium, in 2020, where he is currently pursuing the Ph.D. degree with the Laboratory for Genetic Epidemiology. He combines a residency in clinical genetics at UZ Leuven, Belgium. His research interest includes major-gene effects on facial variation in both health and disease.

**SHUNWANG GONG** received the M.Sc. degree in advanced computing from the Imperial College London, U.K., in 2018, where he is currently pursuing the Ph.D. degree with the Department of Computing. He is also a Research Assistant with the Department of Computing, Imperial College London. He was awarded the Distinguished Project for his M.Sc. degree. His research interests include graph neural networks, 3-D computer vision, generative models, and machine learning.

**GIORGOS BOURITSAS** received the M.Eng. (Diploma) degree in electrical and computer engineering from the National Technical University of Athens (NTUA), in 2017. He is currently pursuing the Ph.D. degree in machine learning with the Department of Computing, Imperial College London. His current research interests include the intersection of machine learning with graph theory and network science. He is primarily focused on the theoretical underpinnings of graph neural networks, on deep probabilistic models for graph generation and compression, on relevant applications to network analysis, bioinformatics, and computer vision.

**GARETH S. BAYNAM** received the Ph.D. degree in vaccine immunogenetics from The University of Western Australia. He is currently a Clinical Geneticist, a Genomic Policy Advisor, a Patient Advocate, a Clinician Scientist, and an Intrapreneur. He equitably implements innovations through multi-stakeholder partnerships. He directs, chairs, or is on the executive for international initiatives to improve the lives of children and youth living with genetic, rare, and undiagnosed diseases. He is currently the Chair of the Diagnostics Scientific Committee of the International Rare Diseases Research Consortium (IRDiRC). He has clinically led the state-wide implementation of genomic and phenotypic technologies, artificial intelligence and digital health platforms, and omics-associated policy. He initiated the Undiagnosed Diseases Program, WA—an interdisciplinary approach for the most challenging medical mysteries. He is also the Head of the Western Australian Registers for Birth Defects and Cerebral Palsy (Western Australian Register of Developmental Anomalies) and the Director of the Academy of Child and Adolescent Health. He is a Clinical Professor or an Adjunct Associate Professor at multiple universities in WA and Victoria. He is a Board Member of the Genetic and Rare Diseases Network, WA, and a member of the Orphanet Australia National Advisory Body and the Rare Voices Australia Scientific and Medical Advisory Committee.

**PETER HAMMOND** was trained in mathematics at Oxford University and in computer science and artificial intelligence with the Imperial College London. His last posts in a 40-year career were a Professor in computational biology with the Institute of Child Health, UCL, and a Senior Fellow in medical image analysis with the Big Data Institute, Oxford University, analyzing neurofacial anatomy for use in epilepsy, medical genetics, and teratology. His recent visiting research posts were with the Department of Human Genetics, Leuven University, Belgium (2017–2021), and the U.K. Centre for Ecology & Hydrology (2018–2020), and the latter arising from an interest in river pollution and using artificial intelligence to detect sewage spills.



**RICHARD SPRITZ** received the B.S. degree in zoology from the University of Wisconsin-Madison, and the M.D. degree from Pennsylvania State University. Previously, he was the Director of the Human Medical Genetics and Genomics Program. In the 1970s, he was a part of the team that characterized the first human genes and discovered the first human gene mutation, in beta-thalassemia. For over four decades, his lab studied the molecular basis of human genetic diseases, including mapping, discovery, and mutational and functional analysis of many different human disease genes. He has received numerous awards for his research and has over 280 publications, on many different topics relating to genetic disorders and birth defects. He was an intern and a resident in pediatrics at the Children's Hospital of Philadelphia. He is currently an Emeritus Professor in pediatrics with the School of Medicine, University of Colorado. He is a fellow in human genetics with the Yale School of Medicine.

**OPHIR D. KLEIN** received the B.A. degree in Spanish literature from the University of California at Berkeley, Berkeley, and the Ph.D. degree in genetics and the M.D. degree from the Yale School of Medicine. He is currently a Professor in orofacial sciences and pediatrics, the Larry L. Hillblom Distinguished Professor in craniofacial anomalies, and the Charles J. Epstein Professor in human genetics with the University of California at San Francisco (UCSF). He works as the Director of the Institute for Human Genetics, the Chief of the Division of Medical Genetics, the Chair of the Division of Craniofacial Anomalies, and the Director of the Program in Craniofacial Biology. His research interests include understanding how organs form in the embryo and how they regenerate in the adult, with a particular emphasis on the processes underlying craniofacial and dental development and renewal and understanding how stem cells in the intestinal epithelium enable renewal and regeneration. He has received several honors, including a New Innovator Award from NIH and the E. Mead Johnson Award from the Society for Pediatric Research. He was elected to the American Society for Clinical Investigation, the American Association of Physicians, and the National Academy of Medicine. He is a fellow of the American Association for the Advancement of Science.

**BENEDIKT HALLGRÍMSSON** received the B.A. degree (Hons.) from the University of Alberta and the M.A. and Ph.D. degrees in biological anthropology from The University of Chicago. He is currently an International Leader in the quantitative analysis of anatomical variation. He is also the Scientific Director of Basic Science with the Alberta Children's Hospital Research Institute and the Head of the Department of Cell Biology and Anatomy. His work focuses on structural birth defects and the developmental genetics of complex traits. He integrates 3-D imaging and morphometry with genetics and developmental biology. He has published more than 150 journal articles, 32 chapters, three edited volumes, and a textbook. He is a fellow of the American Association for the Advancement of Science and the Canadian Academy of Health Sciences. He was awarded the Rohlf Medal for Excellence in Morphometrics, in 2015.

**HILDE PEETERS** received the master's degree in genetic epidemiology from the Netherlands Institute for Health Sciences. She was a Student Researcher with the Center for Human Genetics Leuven working in the field of quantitative genetics and twin studies during her medical training. Subsequently, she specialized in pediatrics and did a Ph.D. degree supported by the FWO within the core domains of human molecular genetics and developmental biology. She was granted an FWO Postdoctoral Fellowship to join the Research Group of Genetic Epidemiology and Statistical Genetics of Prof. C. Van Duijn in Rotterdam for two years. In 2010, she was appointed with an academic position at KU Leuven as a part-time Assistant Professor. The clinical duties accounted for a full-time position mainly on monogenic disorders with a focus on developmental disorders and dysmorpohology. She was granted a Senior Clinical Investigator Fellowship by the FWO for 50% research activities for the project of "Improving counseling for autism and neurodevelopmental disorders through gene mapping, risk variants, and advanced methods in diagnostics." Furthermore, in line with her research and clinical interest, she became the Clinical Laboratory Supervisor for the Diagnostic Laboratory of Congenital and Developmental Disorders, University Hospitals Leuven. She has authored over 80 peer-reviewed journal publications, 122 conference proceedings, and two book chapters (H-index of 22).

**MICHAEL BRONSTEIN** received the Ph.D. degree from Technion, in 2007. He has been working as a Professor with USI Lugano, Switzerland, since 2010. He held visiting positions at Stanford, Harvard, MIT, and TUM. In 2018, he joined the Department of Computing, Imperial College London, London, as a Professor. His main expertise is in theoretical and computational geometric methods for machine learning and data science.

**PETER CLAES** received the Graduate degree from the Department of Electrical engineering (ESAT), KU Leuven, with a major in multimedia and signal processing, in 2002, and the Ph.D. degree in engineering from KU Leuven, in 2007. He continued into a postdoctoral research with the Melbourne Dental School, University of Melbourne, from 2007 to 2011. In 2018, he was a Visiting Scholar with the Biomedical Engineering Department, University of Oxford, U.K. Since 2014, he has been an Honorary Research Fellow with the Murdoch Children's Research Institute, Melbourne, Australia. Since October 2019, he has been a Research Associate Professor in a joint appointment with the Department of ESAT-PSI and the Department of Human Genetics with KU Leuven.