



Article

A Novel Approach Based on Hypergraph Convolutional Neural Networks for Cartilage Shape Description and Longitudinal Prediction of Knee Osteoarthritis Progression

John B. Theocharis [†], Christos G. Chadoulos ^{*,†} and Andreas L. Symeonidis

Department of Electrical & Computer Engineering, Aristotle University of Thessaloniki, 54124 Thessaloniki, Greece; theochar@ece.auth.gr (J.B.T.); symeonid@ece.auth.gr (A.L.S.)

* Correspondence: christgc@ece.auth.gr

[†] These authors contributed equally to this work.

Abstract: Knee osteoarthritis (KOA) is a highly prevalent musculoskeletal joint disorder affecting a significant portion of the population worldwide. Accurate predictions of KOA progression can assist clinicians in drawing preventive strategies for patients. In this paper, we present an integrated approach based on hypergraph convolutional networks (HGCNs) for longitudinal predictions of KOA grades and progressions from MRI images. We propose two novel models, namely, the *C_Shape.Net* and the predictor network. The *C_Shape.Net* operates on a hypergraph of volumetric nodes, especially designed to represent the surface and volumetric features of the cartilage. It encompasses deep HGCN convolutions, graph pooling, and readout operations in a hierarchy of layers, providing, at the output, expressive 3D shape descriptors of the cartilage volume. The predictor is a spatio-temporal HGCN network (*ST_HGCN*), following the sequence-to-sequence learning scheme. Concretely, it transforms sequences of knee representations at the historical stage into sequences of KOA predictions at the prediction stage. The predictor includes spatial HGCN convolutions, attention-based temporal fusion of feature embeddings at multiple layers, and a transformer module that generates longitudinal predictions at follow-up times. We present comprehensive experiments on the Osteoarthritis Initiative (OAI) cohort to evaluate the performance of our methodology for various tasks, including node classification, longitudinal KL grading, and progression. The basic finding of the experiments is that the larger the depth of the historical stage, the higher the accuracy of the obtained predictions in all tasks. For the maximum historic depth of four years, our method yielded an average balanced accuracy (BA) of 85.94% in KOA grading, and accuracies of 91.89% (+1), 88.11% (+2), 84.35% (+3), and 79.41% (+4) for the four consecutive follow-up visits. Under the same setting, we also achieved an average value of Area Under Curve (AUC) of 0.94 for the prediction of progression incidence, and follow-up AUC values of 0.81 (+1), 0.77 (+2), 0.73 (+3), and 0.68 (+4), respectively.

Keywords: knee cartilage osteoarthritis (KOA); deep learning; hypergraph convolutional networks (HGCNs); sequence-to-sequence learning; 3D shape descriptors; Kellgren–Lawrence (KL) grading; KOA progression



Academic Editor: Karin Verspoor

Received: 16 February 2025

Revised: 8 April 2025

Accepted: 16 April 2025

Published: 26 April 2025

Citation: Theocharis, J.B.; Chadoulos, C.G.; Symeonidis, A.L. A Novel Approach Based on Hypergraph Convolutional Neural Networks for Cartilage Shape Description and Longitudinal Prediction of Knee Osteoarthritis Progression. *Mach. Learn. Knowl. Extr.* **2025**, *7*, 40. <https://doi.org/10.3390/make7020040>

Copyright: © 2025 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and

conditions of the Creative Commons Attribution (CC BY) license

(<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Knee Osteoarthritis (KOA) is one of the most prominent joint diseases affecting people around the world, with the majority of patients experiencing its onset at age 60 or greater. The generative causes that are thought to contribute the most in the onset and progression

of the disease are (1) misalignments in bone structure and anatomy, either inherited (congenital) or developed throughout one's life, (2) excess body mass that chronically places considerable mechanical stress on the knee joint, coupled with a general state of inactivity which, in time, weakens and diminishes the supporting musculature (3), and (4) advanced age [1]. The ensuing cartilage loss induced by the condition exposes the femoral and tibial bones to high friction forces during any type of movement, leading to a gradual denudation of the cartilage volume, and in the more extreme scenarios, to complete exposure of the respective bone surfaces. The above consequences, in time, result in substantial loss of quality of life in the more severe cases, limiting the patients' mobility in everyday tasks, and overall resulting in a state of diminished general well-being [2].

1.1. KOA Detection and Progression

Imaging modalities such as X-rays and MRIs are invaluable tools in the hands of experts, with regards to diagnosing KOA and assessing its future trajectory. Data repositories that host large collections of imaging data provide unique opportunities to the scientific community, towards the development of methods that can adequately capture the intricacies of the incidence of the disease. Over the years, there has been a considerable body of research making use of such resources to automatically detect the presence of KOA in patients. Recent advancements in both hardware and software technologies are offering new possibilities with regards to the development of large-scale models that can harness the full wealth of the available data.

One of the most prominent scoring systems used for the evaluation of the severity of KOA is the Kellgren–Lawrence (KL) grading scale [3]. This scale is used by experts to evaluate the severity of the condition based on findings from imaging (predominantly X-ray) data. The scale extends from 0 up to 4, with each increasing grade signifying a progressively more severe condition:

1. **KL0:** Definite absence of KOA;
2. **KL1:** Doubtful Joint-Space-Narrowing (*JSN*);
3. **KL2:** Definite presence of osteophytes and indications of *JSN*;
4. **KL3:** Presence of multiple moderate osteophytes, and possible detection of bone deformities;
5. **KL4:** Presence of large osteophytes, severe narrowing of joint space, and definite bone deformities.

Apart from imaging data, the additional incorporation of generic demographic information may potentially yield a net positive effect in the predictive capabilities of the aforementioned models, since all the above variables constitute known risk factors that can influence the future trajectory of KOA. Variables such as age, Body-Mass-Index (BMI) and gender are primary examples of such features, and may be harnessed to steer the diagnostic models towards more robust performance. A comprehensive literature review of the research mentioned above can be found in [4].

In addition to the efforts towards the development of automatic methods that can identify the presence of KOA, an equally pressing task is the exploitation of the available longitudinal data repositories for the advancement of existing methodologies, towards the detection of progression trajectories in the temporal dimension. Deep learning models, making use of high-dimensional representations of the raw data, can potentially identify progression patterns that are impossible for a human expert to detect with consistency. Enabling clinicians to confidently predict such a pattern in a patient will enhance their ability to intervene in the early stages of KOA development. This, in turn, may allow for timely treatment of the condition, ultimately improving the future outlook for the afflicted person by dampening the more severe symptoms associated with it, leading to better

outcomes overall. Several studies reviewing recent developments towards this direction can be found in [5,6].

1.2. Proposed Methodology

In this paper, we propose an integrated approach to acquire longitudinal grading of KOA severity. Figure 1 shows the pipeline of our method. It comprises a preliminary segmentation stage and two main networks, namely, the cartilage 3D shape network (*C_Shape.Net*) and the predictor network (*ST_HGCN*). Both networks exploit the higher order hyperedges and the multi-view fusion of *HGCN* convolutions to achieve their tasks.

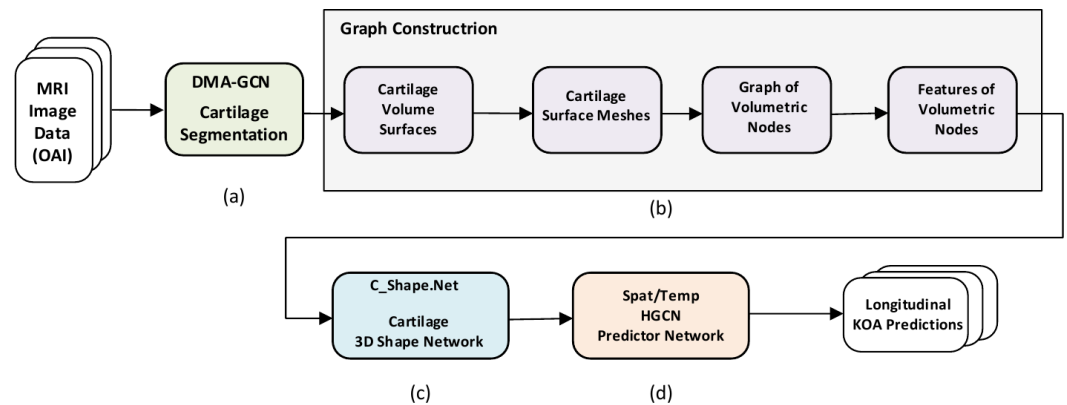


Figure 1. Overview of the proposed methodology. (a) Preliminary stage of 3D cartilage segmentation. (b) Graph construction of volumetric nodes. (c) *C_Shape.Net* providing 3D shape descriptors of the cartilage volume. (d) The *ST_HGCN* predictor network providing longitudinal predictions of KOA grades and progression.

Most existing methods directly use radiology scans (usually X-ray) as an input source to produce KOA predictions. This suggests that the capturing of the cartilage characteristics and the assessment of KOA grades are blended, thus increasing the difficulties of the combined task at hand. In our perspective, we decouple the two procedures. Concretely, we deploy the recently proposed *DMA-GCN* [7] method to produce the 3D cartilage segmentation maps from MRI scans at the preliminary stage (Figure 1a). The model integrates local learning convolutional units that aggregate spatial contextual information at multiple scales and global learning convolutional units exploring the global relationship among nodes in the image. These units are suitably interconnected within a dense convolutional architecture that allows the generation of comprehensive feature representations and hence, accurate classification results.

The graph construction task (Figure 1b) aims to transform the 3D cartilage volume (CV) into a graph of volumetric nodes (VNs). The process is especially designed to describe the specific structural properties of the knee CV. VNs are formed as 3D structures between two corresponding faces (upper/bottom), belonging to triangular meshes established on the different surfaces of the CV. The nodes are equipped with a set of local features, including spatial and structural geometric features of the pertaining faces, as well as volumetric measures, such as thickness and volume values. To further enhance the representation capabilities of nodes, we also exploit local spatial information by incorporating the features of neighboring nodes.

The *C_Shape.Net* (Figure 1c) is a deep hypergraph convolutional network with a hierarchical structure. This network operates on a patient's level, i.e., acts on a hypergraph associated with each knee's CV individually. Given the VN representations, it aims to generate a global 3D shape descriptor of a knee, describing the structural properties of the entire 3D cartilage volume. At each layer, *C_Shape.Net* encompasses three main operations,

specifically, the densely connected *HGCN* block (*DHGCN*), the graph coarsening, and the readout layers. The *DHGCN* block conducts multi-view *HGCN* convolutions, with the different views corresponding to the surface and the volumetric features, respectively. In addition, *DHGCN* also incorporates local and global multi-view modules. The former implements aggregation of neighboring *VNs* in a local 3-hop neighborhood, while the latter ones explore global pairwise affinities between distant nodes in *CV*. The graph coarsening layer reduces the size of the incoming graph by retaining a portion of representative *VNs*, while the readout unit extracts global features from the graph at the different layers in the hierarchy. To acquire a better shape representation of the entire knee, the final shape descriptor is formed as a concatenation of the respective shapes corresponding to the medial and lateral compartments of the knee, respectively.

As the final stage of our methodology, the predictor model is a *HGCN* network with spatio-temporal processing properties (*ST_HGCN*) (Figure 1d). Contrary to the *C_Shape.Net*, the *ST_HGCN* network now operates on temporally interconnected hypergraphs of knees (nodes) in the *OAI* repository. These hypergraphs are considered at different time steps, including the baseline visit and the following eight yearly follow-up visits. *ST_HGCN* aims at generating multi-step ahead predictions of *KOA* grades at the prediction stage, based on knee data at the historical stage. The model undertakes three main tasks: (a) the spatial convolutions, (b) the attention-based temporal fusion, and (c) the sequence transformation. The former task performs multi-view *HGCN* convolutions, consecutively on every graph of knees, at the different time slices of the historical stage. This operation explores the pairwise relations among the knees, with the scope to acquire a more comprehensive feature representation. The knee relationships are encoded by considering four different hyperedges (views): the global shape descriptors created by the *C_Shape.Net* and the three demographic features (age, BMI, gender). The second task aggregates the feature embeddings across the different branches at the historical stage. Its goal is to discern important temporal changes within the knee's shape sequence, which may be valuable for assessing the *KOA* incidence and progression. Finally, for each node, the transformer network transforms the sequence of fused embeddings from the historical stage to a sequence of *KOA* grades at the prediction stage.

Our methodology is comprehensively evaluated on the *OAI* [8] dataset. In the experimental setup, we examine the accuracy of longitudinal predictions for different sizes of the historical data. In addition, we investigate the demographic-based performance, i.e., the accuracies attained on different patients' groups in terms of age, BMI, and gender. Finally, we provide experimental results to assess the efficacy of the proposed approach on the issue of *KOA* progression.

In summary, the main contributions of this paper are described as follows:

- (I) *A process to construct a graph of volumetric nodes to represent a 3D cartilage volume*: The nodes are described by a rich set of features, including local spatial and geometric features of the surfaces, along with descriptive volumetric measures.
- (II) *A novel deep hypergraph convolutional neural network (C_Shape.Net)*: The shape network operates on volumetric graphs, aiming to generate a comprehensive 3D shape descriptor of a knee's cartilage volume.
- (III) *A novel HGCN-based predictor network (ST_HGCN)*: The model exhibits spatio-temporal processing capabilities and adopts the sequence-to-sequence learning approach. The predictor transforms the shape data sequences at the historical stage to sequences of *KOA* grades at the prediction stage.
- (IV) *Multi-view convolutions and adaptive hyperedge learning*: We apply multi-view *HGCN* convolutions to fuse and weight different data sources. In the *C_Shape.Net*, the different views correspond to the face and volumetric features of nodes, where in *ST_HGCN*,

the views are associated with the shape features and the demographic risk factors of patients. Further, in all *HGCN* convolutions, we perform adaptive hyperedge learning at each layer.

- (V) *Semi-supervised learning (SSL)*: The *ST_HGCN* network is trained under the *SSL* approach, which allows the leverage of both the training and testing shape data during learning.
- (VI) *Experimental evaluation*: The combined pipeline of the proposed methodology is extensively tested on the *OAI* cohort. We have devised various experimental scenarios, to assess the accuracy of the multi-step ahead predictions of *KOA* at future visits, the demographic-based accuracies, the early-stage predictions of *KOA*, as well as the *KOA* progression at the prediction stage.

2. Related Works

In this section, we briefly review related works on 3D shape descriptors and *KOA* risk models.

2.1. 3D Shape Descriptors

With the advancements in the field of 3D computer vision, the 3D shape analysis based on deep learning has been gaining increased attention in recent years, for object recognition, segmentation, and retrieval tasks. A thorough review of the 3D shape representation methods can be found in [9]. Broadly, the plethora of existing approaches can be distinguished into five categories, namely, the view-based, the point-cloud-based, the voxel-based, the mesh-based, and the graph convolutional network *GCN*-based methods.

Exploiting the advanced processing capabilities of *CNNs* on 2D images, the view-based methods combine a collection of different 2D views of a 3D object. Some representative models in this category are the multi-view *CNN* (*MVCNN*) [10], the group-view *CNN* (*GVCNN*) [11], the view N-gram Network (*VNN*) [12], and the multi-view harmonized bilinear network (*MHBN*) [13]. *MVCNN* uses *CNNs* to extract multi-view features, which are then aggregated via max-pooling to create compact 3D shape descriptors. *GVCNN* performs advanced feature aggregation by considering group multi-view features and then devises a feature pooling scheme on group views. *MHBN* aggregates local convolutional multi-view features through harmonized bilinear pooling to obtain more discriminative 3D shape representations.

The methods of the second category operate on point clouds, which comprise a group of irregular and unordered points distributed in 3D space. They are striving to tackle the inherent difficulties of convolutional operations in handling directly a point cloud. *PointNet* [14] learns to encode a point cloud and generates its global descriptor for segmentation and classification tasks. *PointNet++* [15] confronts the inability of *PointNet* to identify neighborhood structures by aggregating local features of a point cloud. The *PointWeb* [16] densely interconnects all points in a local neighborhood in a point cloud, aiming to define point features representing its local region characteristics. The proposed adaptive feature adjustment *AFA* module is then used to identify interactions between points.

Voxel-based methods deploy *CNNs* to extract 3D shape features from dense 3D volumetric grids of voxels (3D cubes). Popular networks of this category are the *ShapeNet* [17], the *VoxNet* [18], and the *MO-VCNN* [19]. *ShapeNet* generates 3D shape descriptors directly from volumetric grids using a convolutional Deep Belief Network (*DBN*), while *VoxNet* extracts features using a 3D *CNN*. *MO-VCNN* considers volumetric data of different orientations. The authors use 3D *CNNs* to extract orientation specific features. These features are then aggregated via pooling and passed through another 3D *CNN* to make predictions.

The 3D shape analysis based on 3D meshes has also attracted great interest in the literature. Recent methods in this category are the *MeshNet* [20] and the *MeshCNN* [21]. In *MeshNet*, the faces are first passed through the *NN*-based spatial and structural descriptors to generate their respective features. Next, the individual face features along with the ones of its neighboring faces are aggregated via mesh convolutional blocks to obtain the shape descriptor after pooling. *MeshNet* uses a suitably designed *CNN* to generate shape descriptors from 3D triangular meshes. The model utilizes specialized convolutional and pooling layers that operate on the mesh edges, leveraging their intrinsic geodesic connections.

The *GCN*-based methods leverage the ability of *GCNs* to conduct node convolutions on graphs. Three interesting contributions in this field are the *View-GCN* [22], the *MHGCN* [9], and the *MDC-GCN* [23]. In *View-GCN*, a view-graph is first constructed with multiple views from different orientations, considered as graph nodes. Then, hierarchical *GCN* convolutional networks are designed to learn discriminating shape descriptors, considering relations among the views. *MHGCN* suggests a multi-scale representation scheme on *HGCN* for 3D shape retrieval and recognition. The method integrates multiple 3D shape descriptors (views) obtained from different feature extractors, each view described by a separate hypergraph. High-order relationships among the views are acquired by concatenating the view-specific incident matrices. In addition, multiple feature representations across the convolution layers are fused for more robust results. The *MDC-GCN* model introduces densely connected *GCNs* on 3D meshes to extract shape descriptors for object classification and segmentation. Initially, the 3D mesh is converted into a graph structure, with each face corresponding to a node in the graph. To enhance the expressive power of nodes, the authors cast a feature set incorporating local geometric characteristics of a face and its 1-ring face neighbors. For a better fusion of local and non-local features, convolutions in the *MDC-GCN* are conducted using densely connected *GCN* blocks.

In this work, we propose the *C_Shape.Net* to extract 3D shape descriptors, especially cast according to the requirements of the 3D knee cartilage volumes. Specifically, we seek to identify abnormalities of the bone surfaces as indicators of osteophytes and primarily represent the volumetric characteristics of the joint space of the cartilage. Both these issues are essential for accurate assessments in *KOA* grading and progression. To this end, we first convert the cartilage shape into a graph of *VNs*. Contrary to the 3D cube grid used in voxel-based approaches, the *VNs* here are created as 3D structures of prism type, formed between corresponding faces. Hence, regarding shape analysis, our *C_Shape.Net* incorporates the principles from *GCN*-based, voxel-based, and mesh-based methods, simultaneously.

2.2. *KOA* Prediction Models

With regards to *KOA* severity estimation, most works in the established literature approach the *KL* assignment problem as a standard classification task, whereby the proposed models are trained to derive an association between image representations and *KOA* quantification. Conventional machine learning methods rely on an initial feature extraction step, while deep learning ones streamline the process of feature engineering and *KL* prediction into a single pipeline. The authors in [24] utilize a variant of the Siamese *CNN* architecture that scores *KOA* severity in the *KL* scale. The same task is tackled in [25] via the joint use of a Residual Neural Network (*ResNet*) and a Convolutional Attention Block Module (*CBAM*) [26]. In [27], the problem of *KL* grading is treated as an ordinal regression task, with a similar path being adopted in [28,29], whereby the authors in both works fine-tune several popular *CNN* models to obtain the *KOA* scores. Finally, in [30], a multi-task transfer learning approach is proposed for the joint prediction of *KL* grade and likelihood of Total Knee Replacement (*TKR*).

Existing methods in *KOA* progression prediction usually consider the patient's data at the baseline and predict its future progression up to a specified follow-up time. For instance, the work in [31] predicts early *KOA* incidence within 24 months, while the authors in [32,33] predict *KOA* progression within a middle-term period of 48 months ahead. In [34], following an initial quantification of *KOA* severity based on the Cartilage Damage Index (*CDI*), the authors evaluate four standard machine learning algorithms on the prediction of *KOA* progression, as measured by *KL* grade and Joint Space Narrowing at the medial and lateral compartments (*JSM*, *JSL*), respectively. In [35], a mixed-effects mixture model was employed to initially form two clusters of *KOA* progression vs. non-progression, and several regression models were subsequently trained to estimate the probabilities of subjects belonging to each of the above clusters. The risk assessment models in [36] provide longitudinal predictions of progression, spanning all future time steps of the prediction stage, while the same task is tackled in [37], via the use of a Transformer module that performs multi-modal data fusion. A common feature of these methods is the lack of historic data of patients used in the learning process. While all of the aforementioned works constitute interesting approaches in the *KOA* progression modeling task, they nevertheless share the same limitation of not incorporating the temporal dimension of the available repositories. Instead, these models are trained on images and clinical variables from the baseline visit alone, and treat the progression estimation as a series of independent classification tasks, disregarding the inherent sequential aspect of the data.

A primary goal of this work is to investigate the effect of previous historical data in future *KOA* predictions. Therefore, the data are distinguished here into a historic stage of varying depth and a prediction stage. Accordingly, each knee is associated with a sequence of shape/clinical data and a sequence of grades/progressions at future follow-up times. Further, an important innovation of our approach is that the pool of patients is modeled as a hypergraph of nodes with spatio-temporal characteristics, i.e., temporally interconnected hyperedges evolving over time. Our predictor performs spatial *HGCN* convolutions to exploit the relationships between the patients, temporal fusions to capture the evolution trends, and sequence-to-sequence transformation to obtain longitudinal predictions ahead. A similar framework can be found in the field of traffic flow predictions in road networks [38–40]. The efficacy of spatio-temporal techniques on graph representations has been thoroughly demonstrated by extensive research over the last several years.

3. Hypergraph Convolutional Networks

3.1. Hypergraph Definition

A hypergraph is defined as $\mathcal{G} = \{\mathcal{V}, \mathcal{E}, \mathbf{H}, \mathbf{X}\}$, where $\mathcal{V} = \{v_i\}_{i=1}^N$ is the node set of N nodes ($|\mathcal{V}| = N$), $\mathcal{E} = \{e_j\}_{j=1}^M$ denotes the hyperedge set comprising M hyperedges ($|\mathcal{E}| = M$), and $\mathbf{H} \in \mathbb{R}^{N \times M}$ is the incidence matrix. $\mathbf{X} \in \mathbb{R}^{N \times F}$ is the data matrix containing the feature descriptors of nodes $\mathbf{x}_i \in \mathbb{R}^F, i = 1, \dots, N$ where F denotes the feature dimensionality. The hypergraph structure is represented by an incidence matrix $\mathbf{H} \in \mathbb{R}^{N \times M}$ with its rows and columns corresponding to the nodes and hyperedges, respectively. The entries of $\mathbf{H}$ are defined as follows:

$$\mathbf{H}(v_i, e_j) = \mathbf{H}_{ij} = \begin{cases} 1, & \text{if } v_i \in e_j \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

Each hyperedge e_j is assigned a learnable weight w_j stored in a diagonal weight matrix $\mathbf{W} = \text{diag}(w_1, \dots, w_j, \dots, w_M) \in \mathbb{R}^{M \times M}$. The node degree of $v_i \in \mathcal{V}$ is given by $D_n(i) = \sum_{j=1}^M w_j h_{ij}$. Further, the degree of each hyperedge $e_j \in \mathcal{E}$ is defined as $D_e(i) = \sum_{i=1}^N h_{ij}$. The diagonal matrices $\mathbf{D}_n = \text{diag}(D_n(1), \dots, D_n(N)) \in \mathbb{R}^{N \times N}$ and

$\mathbf{D}_e = \text{diag}(D_e(1), \dots, D_e(M)) \in \mathbb{R}^{M \times M}$ store all the node and hyperedge degrees, respectively. It should be stressed that, unlike the traditional graphs where each edge links two vertices, the hyperedges in a hypergraph are degree-free, which implies that a given hyperedge can involve more than two vertices, thus providing a more flexible description of high-order relationships among nodes.

3.2. Hypergraph Spectral Convolution

Spectral hypergraph convolution relies on the principles of spectral graph theory [41]. In this context, graph convolution is realized using the graph Laplacian matrix, whereby the eigenvectors and eigenvalues are regarded as the Fourier basis and frequencies, respectively. To account for the high computational cost involved in hypergraph convolutions, the authors in [42] suggested the use of K -order Chebyshev polynomials:

$$g * x = \sum_{k=1}^K \theta_k T_k(\tilde{\mathbf{L}})(x) \quad (2)$$

where $\tilde{\mathbf{L}} = (2/\lambda_{\max})\mathbf{L} - \mathbf{I}$ is a rescaled form of the Laplacian matrix $\mathbf{L} \in \mathbb{R}^{N \times N}$ defined as

$$\mathbf{L} = \mathbf{I} - \mathbf{D}_n^{-1/2} \mathbf{H} \mathbf{W} \mathbf{D}_e^{-1} \mathbf{H}^T \mathbf{D}_n^{-1/2} \quad (3)$$

The Chebyshev polynomials are recursively computed via $T_k(x) = 2xT_{k-1}(x) - T_{k-2}(x)$ with $T_0(x) = 1$ and $T_1(x) = x$. The expression in Equation (2) is K -localized, which suggests that the convolution result at a specific node is derived by aggregating the feature representations of nodes belonging to K -hop neighborhoods around the central node. Setting $K = 1$ and $\lambda_{\max} \approx 2$ [42], we are led to a hypergraph convolution, defined as follows:

$$\mathcal{X}^{(l+1)} = \sigma \left(\mathbf{D}_n^{-\frac{1}{2}} \mathbf{H} \mathbf{W} \mathbf{D}_e^{-1} \mathbf{H}^T \mathbf{D}_n^{-\frac{1}{2}} \mathbf{X}^{(l)} \boldsymbol{\Theta}^{(l)} \right) \quad (4)$$

where $\mathcal{X}^{(l+1)}$ are the embeddings at the output of the l -th layer, $\sigma(\cdot)$ denotes the *ReLU* activation function, $\mathbf{H}$ represents the structure of the hypergraph, $\mathbf{W}$ contains the hyper-edge weights, $\mathbf{X}^{(l)} \in \mathbb{R}^{N \times F^{(l)}}$ stands for the data matrix with N nodes and $F^{(l)}$ channels, and $\boldsymbol{\Theta}^{(l)} \in \mathbb{R}^{F^{(l)} \times F^{(l+1)}}$ is the filter matrix of trainable parameters. Stacking multiple convolutional layers, we can build *HGCN* models with enhances learning capabilities.

3.3. Adaptive Hypergraph Learning

A crucial issue in *HGCN* learning to be properly addressed is how the nodes are organized in hyperedges to construct $\mathbf{H}$. In some previous works [43], the hyperedges are attached to each node individually, i.e., $|\mathcal{E}| = M = N$, which leads to an incidence matrix $\mathbf{H} \in \mathbb{R}^{N \times N}$. Specifically, the hyperedge $e_j \in \mathcal{E}$ corresponding to the node $\mathbf{x}_j$, is formed by considering the k -nearest neighbors of $\mathbf{x}_j$, i.e., $e_j = \mathcal{N}(\mathbf{x}_j)$. In this context, we apply an adaptive hypergraph learning (*AHL*) mechanism based on the *GAT*-based attention [44], to automatically learn the hyperedges' structure, using the current layer inputs at each layer.

Given the embedding matrix $\boldsymbol{\Theta}^{(l)}$ in Equation (4), we perform a shared attention mechanism on all node pairs:

$$g_{ij} = \boldsymbol{\alpha}^T \left(\mathbf{x}_i^{(l)} \boldsymbol{\Theta}^{(l)} \parallel \mathbf{x}_j^{(l)} \boldsymbol{\Theta}^{(l)} \right) \quad (5)$$

where $\boldsymbol{\alpha}$ is a learnable weight vector and g_{ij} indicates the pairwise similarity between nodes i and j . The normalized attention coefficients are obtained using a single-layer feedforward neural network, parameterized by *LeakyReLU* nonlinearities, as follows:

$$\mathbf{H}_{(i,j)} = \frac{\exp(\text{LeakyReLU}(g_{ij}))}{\sum_{k \in \mathcal{N}(\mathbf{x}_j)} \exp(\text{LeakyReLU}(g_{kj}))} \quad (6)$$

The attention coefficient $H_{(i,j)}$ provides the membership degree of node- i to hyperedge $e_j, j = 1, \dots, N$. For efficiency reasons, the computation of attention coefficients above is confined to the nodes in $\mathcal{N}(\mathbf{x}_j)$. To stabilize the learning procedure, we consider multiple attention heads. The hyperedge convolution is now obtained by the following:

$$\mathbf{X}^{(l+1)} = \left\| \bigg\|_{k=1}^K \mathbf{X}_k^{(l+1)} \right\| = \left\| \bigg\|_{k=1}^K f\left(\mathbf{X}^{(l)}, \mathbf{H}^{(k)}; \mathbf{W}, \mathbf{\Theta}_k^{(l)}\right) \right\| \quad (7)$$

where K is the number of attention heads, $\mathbf{H}_k^{(l)}$ and $\mathbf{\Theta}_k^{(l)}$ are the produced incidence matrix and the embedding matrix associated with the k -th attention head, $\|$ denotes the concatenation operator, and the function $f(\cdot)$ is used to implement the convolution operator in Equation (4).

4. Materials

Knee data in this study are selected from the *OAI* cohort, which contains longitudinal clinical and radiographic data for 4096 subjects between 45 and 79 years old. The data spread over a 9-year follow-up period, starting from the baseline visit and extending to the next eight yearly follow-up visits. Each subject is attached with a record comprising, among others, the following data: (1) 3D *MRI* images according to the *DESS* acquisition protocol, formed as a collection of consecutive 2D slices. (2) The respective segmentation masks [45], including labels for the following knee-joint structures (classes): Background tissue, Femoral Bone (*FB*), Femoral Cartilage (*FC*), Tibial Bone (*TB*), and Tibial Cartilage (*TC*). (3) Demographic features of {age, BMI, gender}, and a history of knee injury, according to an *OAI* questionnaire at baseline. (4) The Kellgren–Lawrence (*KL*) grades, quantifying the *KOA* severity into five *KL* classes {0, 1, 2, 3, 4}, where the classes are interpreted as {0, 1} = No_OA, {2} = Doubtful_OA, {3} = Moderate_OA and {4} = Severe_OA. (5) Anatomic axis alignment (tibiofemoral angle) and minimum medial joint space width measurements.

In our experiments, we have collected 3114 subjects ($2 \times = 6228$ knees) with complete record data along the follow-up times. As input sources, we deploy the *MRI* scans and the demographic features.

1. With regards to the *MRI* scans, in order to obtain the full segmentation maps for all the knees in the historical stage (which will subsequently serve to generate an initial set of volumetric and geometric features to be processed by the *C_Shape.Net*), we deploy our previously proposed *DMA-GCN* [7] cartilage segmentation model. The segmentation maps generated by the aforementioned process serve the input to the overall *ST_HGCN* *KOA* prediction network presented in this study.
2. Complementing the above imaging data, our approach makes use of additional demographic features for the determination of the various types of hyperedges that constitute the hypergraphs that are utilized by the *ST_HGCN* predictor network. Concretely, we utilize the available information about the age, gender, and Body-Mass-Index (*BMI*) of the subjects whose knees are processed in order to define clusters that can implicitly define the corresponding hyperedges in *ST_HGCN*.

Table 1 presents a *KL*-grade distribution over time. As can be seen, knees are adequately dispersed across the *KL* classes at baseline. Further, a sufficient portion of knees exhibits a trend of *KOA* progression towards more severe stages. Figure 2a depicts a

typical *DESS* scan in the sagittal, coronal, and axial planes showing the different bone and cartilage constituents.

Table 1. *KL*-grade distribution at the multiple historical time points considered in this study.

	Baseline	12-Month	24-Month	36-Month
<i>KL0</i>	2500	2401	2369	2336
<i>KL1</i>	1206	1153	1131	1082
<i>KL2</i>	1610	1645	1640	1636
<i>KL3</i>	774	841	863	910
<i>KL4</i>	138	188	225	264

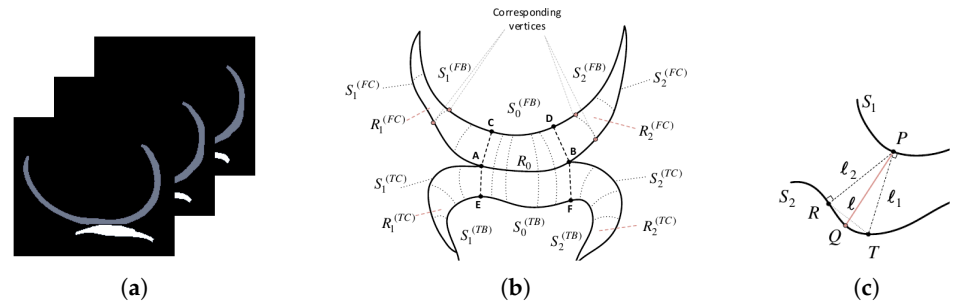


Figure 2. (a) Slices of an MRI segmentation map, showcasing the femoral (gray) and tibial (white) cartilages in the sagittal plane. (b) A 2D illustrative drawing of a particular slice showing the cartilage regions and surfaces involved. (c) Explanatory scheme describing the formation of corresponding vertices between two surfaces.

5. Graph Construction of Volumetric Nodes

Given the overall 3D cartilage volume $CV = FC \cup TC$ submitted by *DMA-GCN*, we develop in this section a four-stage process for constructing a suitable volumetric graph, with the aim to acquire an effective 3D shape representation of the *CV*.

5.1. Cartilage Volume Surfaces

Figure 2b shows an illustrative 2D scheme corresponding to a specific slide along the sagittal direction, which depicts the different cartilage bodies and surfaces involved. Generally, *CV* is delimited by the two main bone surfaces, namely, the *FB* surface $S^{(FB)}$ and the *TB* surface $S^{(TB)}$, respectively. As a first step, for each slide, we identify the characteristic inflection points *A* and *B*, where *FC* intersects with the *TC* part. Next, we draw the lines *AC* and *BD* perpendicular to the $S^{(FB)}$, i.e., $AC \perp S^{(FB)}$ and $BD \perp S^{(FB)}$. Similarly, we draw the lines *AE* and *BF* perpendicular to the $S^{(TB)}$, i.e., $AE \perp S^{(TB)}$ and $BF \perp S^{(TB)}$, respectively. The location of these points divides $S^{(FB)}$ into three parts, $\{S_0^{(FB)}, S_1^{(FB)}, S_2^{(FB)}\}$, where $S_0^{(FB)}$ denotes the central *FB* surface, while the latter ones correspond to the lateral *FB* surfaces. Similarly, $S^{(TB)}$ is split into three parts, $\{S_0^{(TB)}, S_1^{(TB)}, S_2^{(TB)}\}$. Further, we can distinguish the outer *FC* surfaces $\{S_1^{(FC)}, S_2^{(FC)}\}$ along the frontal and lateral view of the knee, and the respective outer surfaces of *TC* denoted as $\{S_1^{(TC)}, S_2^{(TC)}\}$.

In view of the above surface definitions, the entire *CV* area can be decomposed into five regions, circumscribed by their respective surfaces:

$$\begin{aligned}
 \mathcal{R}_0 &= \mathcal{JS} = \{S_0^{(FB)}, S_0^{(TB)}\} \\
 \mathcal{R}_1^{(FC)} &= \{S_1^{(FB)}, S_1^{(FC)}\} & \mathcal{R}_2^{(FC)} &= \{S_2^{(FB)}, S_2^{(FC)}\} \\
 \mathcal{R}_1^{(TC)} &= \{S_1^{(TB)}, S_1^{(TC)}\} & \mathcal{R}_2^{(TC)} &= \{S_2^{(TB)}, S_2^{(TC)}\}
 \end{aligned} \tag{8}$$

It should be stressed that although illustrations are given on a slide basis, the above regions correspond to 3D cartilage volumes, when considering the succession of all slides in the knee MRI. The region $\mathcal{R}_0$ corresponds to the joint space of the knee cartilage and constitutes the crucial volume of interest. The joint space width along with the abnormalities of its embracing bone surfaces (osteophytes) play a pivotal role in assessing the severity grade of KOA. The remaining regions are included in the analysis with the goal of detecting possible appearance of denudation. This also constitutes a significant cause of pain that affects KOA grading seriously.

5.2. Cartilage Surface Meshes

The goal in this section is to construct a triangular 3D mesh for each of the surfaces involved in the cartilage regions. The surface mesh takes the form $\mathcal{M} = \{\mathcal{V}^{(\mathcal{M})}, \mathcal{E}^{(\mathcal{M})}, \mathcal{F}^{(\mathcal{M})}, \mathbf{XF}^{(\mathcal{M})}\}$, where $\mathcal{V}^{(\mathcal{M})} = \{v_i\}$ denotes the set of vertices, $\mathcal{F}^{(\mathcal{M})} = \{f_k\}$ represents the set of triangular faces, while $\mathbf{XF}^{(\mathcal{M})}$ is the vector of face features. Further, $\mathcal{E}^{(\mathcal{M})} = \{E_j\}$ denotes the set of edges on the mesh, each edge representing the connectivity of a face in regard to its 1-hop neighboring faces (Figure 3a). The mesh construction proceeds along the following steps:

- (I) Initially, we perform sufficiently dense sampling along each slide of the *FB* surface to create the set of vertices $\{\mathcal{V}_0^{(FB)}, \mathcal{V}_1^{(FB)}, \mathcal{V}_2^{(FB)}\}$, where $\mathcal{V}_i^{(FB)} \in \mathcal{S}_i^{(FB)}$, $i = 0, 1, 2$. Next, we apply the Delaunay triangularization method [46] to construct the 3D meshes of the *FB* surfaces $\{\mathcal{M}_0^{(FB)}, \mathcal{M}_1^{(FB)}, \mathcal{M}_2^{(FB)}\}$.
- (II) In this step, we establish a one-to-one correspondence (CR) between the vertices on the pairwise surfaces of the cartilage regions defined in Equation (8). Specifically, we apply a CR mapping between the vertices of a reference surface (first argument) and the ones on the target surface (second argument). The vertex mapping is implemented on a per-slide basis. The definition of the corresponding vertices and the computation of their respective thickness values is described in the Appendix A (Figure 3c). For the main region $\mathcal{R}_0$, we consider $\mathcal{V}_0^{(TB)} = CR\{\mathcal{V}_0^{(FB)}\}$, where $\mathcal{V}_0^{(FB)} \in \mathcal{S}_0^{(FB)}$ and $\mathcal{V}_0^{(TB)} \in \mathcal{S}_0^{(TB)}$, indicating the one-to-one correspondence between the vertices on $\mathcal{S}_0^{(FB)}$ (reference) and those on $\mathcal{S}_0^{(TB)}$ (target) surfaces. Proceeding in a similar manner, we can create the CR vertex sets $\mathcal{V}_1^{(FC)} \in CR\{\mathcal{V}_1^{(FB)}\}$ and $\mathcal{V}_2^{(FC)} \in CR\{\mathcal{V}_2^{(FB)}\}$, lying on the outer *FC* surfaces.
- (III) In addition to the previously obtained $\mathcal{V}_0^{(TB)}$, we generate the vertex sets $\{\mathcal{V}_1^{(TB)}, \mathcal{V}_2^{(TB)}\}$ via dense sampling to cover the respective *TB* surfaces, and after triangulation we obtain the meshes $\{\mathcal{M}_1^{(TB)}, \mathcal{M}_2^{(TB)}\}$. Then, the CR mapping completes by creating the sets $\mathcal{V}_1^{(TC)} = CR\{\mathcal{V}_1^{(TB)}\}$ and $\mathcal{V}_2^{(TC)} = CR\{\mathcal{V}_2^{(TB)}\}$ on the outer *TC* surfaces.
- (IV) The one-to-one correspondence between the vertices establishes a similar correspondence between the faces of a reference surface and those of its respective CR surface. For example, we consider that $\mathcal{F}_0^{(TB)} = CR\{\mathcal{F}_0^{(FB)}\}$ to indicate that there is one-to-one mapping in $\mathcal{R}_0$ between the faces on *FB* and *TB*, respectively. The above finding can be further extended to observe that CR assures that the spatial topology between the reference and target meshes is also preserved, which suggests that all CR faces retain a similar neighborhood structure. In view of the above, we can create the 3D meshes of the remaining surfaces of the CV:

$$\begin{aligned}
\mathcal{M}_0^{(TB)} &= CR(\mathcal{M}_0^{(FB)}) \\
\mathcal{M}_1^{(FC)} &= CR(\mathcal{M}_1^{(FB)}), \quad \mathcal{M}_2^{(FC)} = CR(\mathcal{M}_2^{(FB)}) \\
\mathcal{M}_1^{(TC)} &= CR(\mathcal{M}_1^{(TB)}), \quad \mathcal{M}_2^{(TC)} = CR(\mathcal{M}_2^{(TB)})
\end{aligned} \tag{9}$$

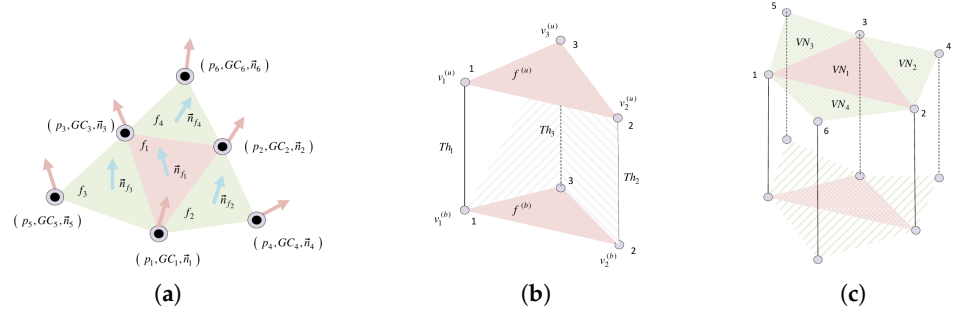


Figure 3. Cartilage mesh face structure. (a) A 1-ring neighborhood of mesh face. (b) A 3D prism formed by an upper and bottom triangular cell (face). (c) Neighborhood structure around a triangular cell.

5.3. Graphs of Volumetric Nodes

In this section, we elaborate on the critical part of the *C_Shape.Net*, namely, the description of volumetric nodes (VN) along with the respective volumetric graphs. Both are especially tailored to represent the structural characteristics of the knee cartilage volume. The generic VN model proposed herein is shown in Figure 3b, and formally, it can be described as a 3D structure of the form:

$$VN = struct(f^{(u)}, f^{(b)}) \tag{10}$$

As can be seen, the VN shape comprises an upper triangular face $f^{(u)}$ that lies on a reference mesh, and a corresponding face $f^{(b)} = CR(f^{(u)})$ at the bottom, belonging to the target mesh. Particularly, for the joint space region $\mathcal{R}_0$, the faces belong to the *FB* and *TB* meshes, i.e., $f^{(u)} \in \mathcal{M}_0^{(FB)}$ and $f^{(b)} \in \mathcal{M}_0^{(TB)}$. There are also three lateral faces of a quadrilateral shape. The three lateral edges connecting the corresponding vertices of $f^{(u)}$ and $f^{(b)}$ represent the cartilage thicknesses between these points. Finally, we pay due attention to the volume *VN*, which is a critical measure in *KOA* assessment analysis. This quantity is readily computed by suitably decomposing the *VN* shape into three triangular pyramids.

Given the meshes formed at the different surfaces, we can create *VNs* across all regions, thus covering the entire *CV*. The *VNs* of a specific region form a hypergraph structure of the form: $\mathcal{G} = \{\mathcal{V}^{(\mathcal{G})}, \mathcal{E}^{(\mathcal{G})}, \mathcal{H}^{(\mathcal{G})}, \mathcal{X}^{(\mathcal{G})}\}$, where $\mathcal{V}^{(\mathcal{G})}$ denotes the set of *VNs*, $\mathcal{E}^{(\mathcal{G})}$ represents the spatial connections among the different *VNs*, $\mathcal{H}^{(\mathcal{G})}$ denotes the incidence matrix of hyperedges and $\mathcal{X}^{(\mathcal{G})}$ represents the node features. Accordingly, we generate the graph $\mathcal{G}_0$ attached to the joint space $\mathcal{R}_0$, as well as the regional sub-graphs $\{\mathcal{G}_1^{(FC)}, \mathcal{G}_2^{(FC)}\}$ and $\{\mathcal{G}_1^{(TC)}, \mathcal{G}_2^{(TC)}\}$, assigned to the respective *FC* and *TC* parts, respectively. The overall graph $\mathcal{G}$ is then formed as a collection of the above regional subgraphs:

$$\mathcal{G} = \{\mathcal{G}_0, \mathcal{G}_1^{(FC)}, \mathcal{G}_2^{(FC)}, \mathcal{G}_1^{(TC)}, \mathcal{G}_2^{(TC)}\} \tag{11}$$

5.4. Features of Volumetric Nodes

The formation of VN features involves two issues, namely, the determination of the face features and the collection of the additional volumetric features characterizing the VN. In regard to the face features, we follow the constructive approach and the feature types suggested in [23]. In this work, the feature vector of a face includes the geometric characteristics of this face itself, along with the corresponding ones of its 1-hop face neighbors. As detailed in Figure 3a, a face f_1 on a mesh is attached with three neighboring faces $\{f_2, f_3, f_4\}$ and a vertex set $\{v_1, v_2, v_3, v_4, v_5, v_6\}$. Then, the feature vector $\mathbf{XF} \in \mathbb{R}^{57}$ is obtained by considering several spatial and structural characteristics, as follows:

$$\mathbf{XF} = [\{\mathbf{P}_v\}, \{\mathbf{N}_v\}, \{\mathbf{GC}_v\}, \{\mathbf{N}_f\}, \{\boldsymbol{\theta}\}] \quad (12)$$

where

$$\mathbf{P}_v = (p_1, p_2, p_3, p_4, p_5, p_6) \quad (13)$$

$$\mathbf{N}_v = (\vec{n}_1, \vec{n}_2, \vec{n}_3, \vec{n}_4, \vec{n}_5, \vec{n}_6) \quad (14)$$

$$\mathbf{GC}_v = (\mathbf{GC}_1, \mathbf{GC}_2, \mathbf{GC}_3, \mathbf{GC}_4, \mathbf{GC}_5, \mathbf{GC}_6) \quad (15)$$

$$\mathbf{N}_f = (\vec{n}_{f_1}, \vec{n}_{f_2}, \vec{n}_{f_3}, \vec{n}_{f_4}) \quad (16)$$

$$\boldsymbol{\theta} = (\theta_{1,1}, \theta_{1,2}, \theta_{1,3}, \theta_{1,4}) \quad (17)$$

$\mathbf{P}_v$ includes the vertex spatial positions, while $\mathbf{N}_v$ contains the vertex normals. $\mathbf{GC}_v$ is the Gaussian Curvature, a significant geometrical indicator in surface description, and $\mathbf{N}_f$ includes the face normals. Finally, $\boldsymbol{\theta}$ includes the angles between the central face and its 1-hop face neighbors. Detailed derivations of these features can be found in [23]. The consideration of the neighboring face information in Equation (16) provides an effective description of a broader local area around a central face.

Most importantly, as can be seen from Figure 3c, the one-to-one CR mapping ensures that volumetric nodes retain a similar neighborhood structure as the ones of the upper and bottom faces, pertaining in the construction of a VN. The 1-hop neighbors of a central $\text{VN}_1 = \text{struct}(f_1^{(u)}, f_1^{(b)})$ are $\{\text{VN}_2, \text{VN}_3, \text{VN}_4\}$, where $\text{VN}_j = \text{struct}(f_j^{(u)}, f_j^{(b)})$, $f_j^{(b)} = \text{CR}(f_j^{(u)})$, $j = 2, 3, 4$. Hence, similarly to the faces, in the design of the VN feature vector, we incorporate the information of the neighboring VNs, to acquire a better perception of the local volume. Concretely, the VN descriptor embraces face and volumetric features, as defined in the following:

$$\mathbf{X} = [\mathbf{XF}^{(u)}, \mathbf{XF}^{(b)}, \mathbf{X}^{(vol)}] \quad (18)$$

$$\mathbf{X}^{(vol)} = [\{\mathbf{Th}_v\}, \{\mathbf{Vol}_v\}] \quad (19)$$

$$\{\mathbf{Th}_v\} = (\text{Th}_1, \text{Th}_2, \text{Th}_3; \text{Th}_4, \text{Th}_5, \text{Th}_6) \quad (20)$$

$$\{\mathbf{Vol}_v\} = (\text{Vol}_1; \text{Vol}_2, \text{Vol}_3, \text{Vol}_4) \quad (21)$$

where $\mathbf{X}^{(vol)}$ contains the volumetric features, $\{\mathbf{Th}_v\}$ subsumes the thickness values amongst corresponding vertices of the central VN and its 1-hop neighbors, while $\{\mathbf{Vol}_v\}$ includes the respective cartilage volumes.

The collection of features in VN should be regarded within the multi-view framework used in this paper. Accordingly, each constituent part in Equation (18) represents a different view (source), which encodes its own aspect of node data. The first two views correspond to the local features, $\mathbf{XF}^{(u)}$ and $\mathbf{XF}^{(b)}$. They are devoted to capturing the bone/cartilage surface variabilities, with the goal to detect possible appearance of osteophytes and denudation.

The third view corresponds to the features $\mathbf{X}^{(vol)}$, which convey valuable volumetric information of local thicknesses and volumes, especially regarding the joint-space area. These features assist in defining the joint space width, and hence the joint space narrowing over time, which are essential factors for the assessment of KOA grading and progression. Overall, the integration of the above views provides a more complete representation of the local features across the entire CV.

6. Cartilage 3D Shape Network

The proposed *C_Shape.Net* adopts a multilayered hierarchical graph architecture, as shown in Figure 4. The network receives as inputs the local features derived in Equation (18), and produces a global 3D shape descriptor, representing the structural characteristics of CV. At each layer, the model involves three main units: the *DHGCN* module that combines local and global *HGCN* convolutions on the VN graphs, the graph coarsening unit aiming to reduce the graph size via a self-attention scoring mechanism on VNs, and the readout units which extract global shape features from the graph across the different layers. These features are then submitted to the *FC* network to generate the final shape representation.

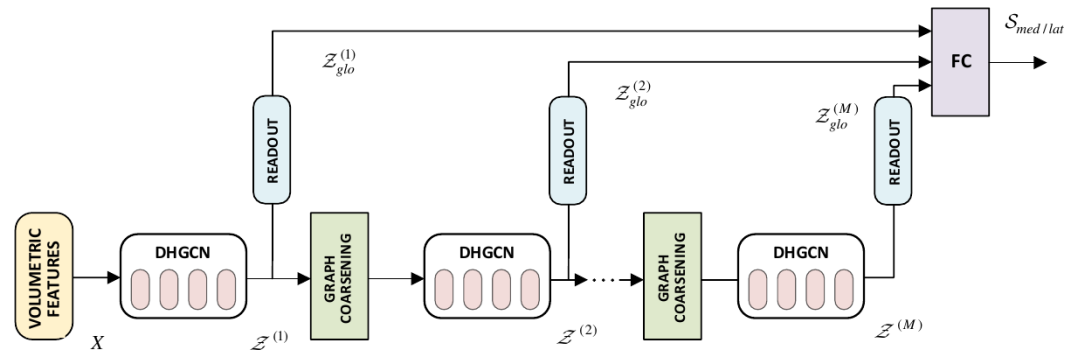


Figure 4. Proposed *C_Shape.Net* Hierarchical Architecture. It includes the *DHGCN* convolutional blocks, the graph coarsening, and readout layers. Global features across the different layers are passed through *FC*, to provide the global descriptor of a cartilage compartment.

6.1. *DHGCN* Blending Block

The *DHGCN* is a densely connected structure conducting *HGCN* convolutions, which integrates both local and global learning on hypergraphs (Figure 5). It comprises the local multi-view convolutional (*LMV*) module and the global multi-view convolutional (*GMV*) module, respectively, that are intertwined into a four-layered model.

The *LMV* undertakes the local learning task, acting upon the hypergraph $\mathcal{G}_s(\mathcal{V}, \mathcal{E}_s, \mathbf{H}_s, \mathbf{U}_s)$, where $\mathcal{V}$ denotes the common set of N nodes ($|\mathcal{V}| = N$). The local hyperedge, set $\mathcal{E}_s$ in conjunction with its respective incidence matrix $\mathbf{H}_s$, reflects the hyperedge structure between the nodes at the local level. The hyperedge associated with a central VN $\mathbf{x}_j$, $j = 1, \dots, N$, is formed by establishing attention weights $\mathbf{H}_s(i, j)$ with respect to its neighboring nodes in a local 3-hop neighborhood around $\mathbf{x}_j$. These weights are adaptively computed at each layer using the *AHL* mechanism discussed in Section 3.3. For each VN, we construct the incidence matrices $\mathbf{H}_s^{(u)} \in \mathbb{R}^{(N \times N)}$, $\mathbf{H}_s^{(b)} \in \mathbb{R}^{(N \times N)}$ and $\mathbf{H}_s^{(vol)} \in \mathbb{R}^{(N \times N)}$, attached to the respective view data, independently. Then, following the multi-view setting, we concatenate these matrices to construct an aggregated matrix $\mathbf{H}_s$, suitable for multi-modal data [43,44]:

$$\mathbf{H}_s = [\mathbf{H}_s^{(u)}, \mathbf{H}_s^{(b)}, \mathbf{H}_s^{(vol)}] \in \mathbb{R}^{(N \times 3N)} \quad (22)$$

where $H_s^{(u)}$ and $H_s^{(b)}$ contain the upper/bottom mesh hyperedges, while $H_s^{(vol)}$ includes the volumetric features' hyperedges. Given H_s , the *LMV* unit generates local *HGCN* convolutions, exploring the spatially local relationships between *VNs*. The local feature representations $\mathcal{X}_s^{(l)} \in \mathbb{R}^{(N \times F^{(l+1)})}$ at layer l are derived as follows:

$$\mathcal{X}_s^{(l)} = \sigma \left(\left(D_{n,s}^{-1/2} \right) H_s W_s \left(D_{e,s}^{-1} \right) (H_s)^T \left(D_{n,s}^{-1/2} \right) \mathcal{U}_s^{(l)} \Theta_s^{(l)} \right) \quad (23)$$

where $D_{n,s} \in \mathbb{R}^{N \times N}$ and $D_{e,s} \in \mathbb{R}^{M \times M}$, ($M = 3N$) denote the node degree and hyperedge degree matrices, as computed by H_s , and $W_s = \text{diag}(\mathbf{W}_s^{(u)}, \mathbf{W}_s^{(b)}, \mathbf{W}_s^{(vol)}) \in \mathbb{R}^{(M \times M)}$ is a diagonal matrix of learnable hyperedge weights. Further, $\mathcal{U}_s^{(l)} \in \mathbb{R}^{N \times F^{(l)}}$ and $\Theta^{(l)} \in \mathbb{R}^{F^{(l)} \times F^{(l+1)}}$ denote the input of the *LMV* unit, and the set of learnable filter parameters, at layer l .

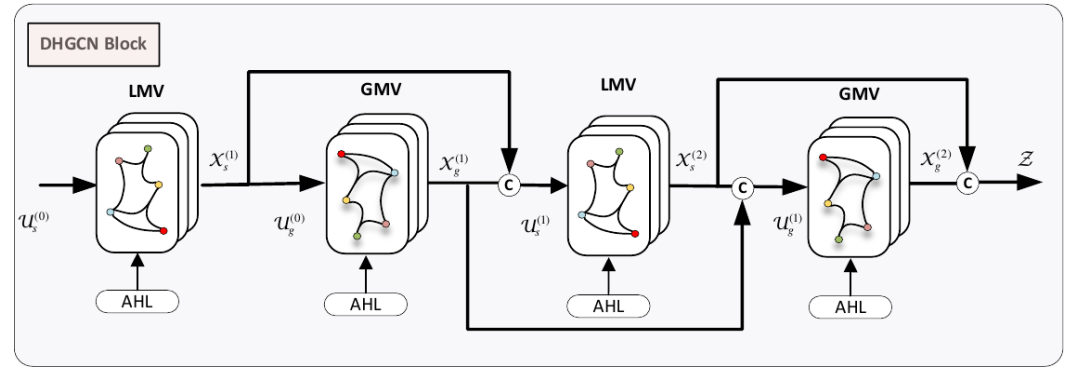


Figure 5. *DHGCN* Convolutional Block. It comprises the local multi-view convolutional (*LMV*) and the global multi-view convolutional (*GMV*) modules, intertwined into a four-layered model.

The *GMV* units, on the other hand, conduct global feature convolutions, operating on the hypergraph $\mathcal{G}_g(\mathcal{V}, \mathcal{E}_g, H_g, \mathcal{U}_g)$. The global hyperedge set $\mathcal{E}_g$ along with its respective incidence matrix H_g now reflect the hyperedge connectivity among the nodes at the global level. In that case, the incidence matrices of the different views $H_g^{(u)} \in \mathbb{R}^{N \times N}$, $H_g^{(b)} \in \mathbb{R}^{N \times N}$, and $H_g^{(vol)} \in \mathbb{R}^{N \times N}$ are determined by computing the attention weights between *VNs* located at spatially distant locations in *CV* using the *AHL* mechanism. Then, after concatenation, we obtain the global multi-modal matrix

$$H_g = [H_g^{(u)}, H_g^{(b)}, H_g^{(vol)}] \in \mathbb{R}^{(N \times M)} \quad (24)$$

The global *HGCN* feature representations acquired by *GMV* at layer l are obtained by

$$\mathcal{X}_g^{(l)} = \sigma \left(\left(D_{n,g}^{-1/2} \right) H_g W_g \left(D_{e,g}^{-1} \right) (H_g)^T \left(D_{n,g}^{-1/2} \right) \mathcal{U}_g^{(l)} \Theta_g^{(l)} \right) \quad (25)$$

where $D_{n,g} \in \mathbb{R}^{N \times N}$ and $D_{e,g} \in \mathbb{R}^{M \times M}$ are the node and hyperedge degree matrices attached to H_g and $W_g = \text{diag}(\mathbf{W}_g^{(u)}, \mathbf{W}_g^{(b)}, \mathbf{W}_g^{(vol)}) \in \mathbb{R}^{M \times M}$ is a diagonal matrix of learnable hyperedge weights. Finally, $\mathcal{U}_g^{(l)} \in \mathbb{R}^{N \times F^{(l)}}$ and $\Theta_g^{(l)} \in \mathbb{R}^{F^{(l)} \times F^{(l+1)}}$ are the *GMV* input and learnable embedding matrix, respectively.

Given the connectivity between the *LMV* and *GMV* units in Figure 5, the signal flow across the *DHGCN* proceeds as follows:

$$\mathcal{U}_s^{(0)} = \mathbf{X}^{(0)}, \quad \mathcal{U}_g^{(0)} = \mathcal{X}_s^{(1)} \quad \mathcal{U}_s^{(1)} = \mathcal{X}_s^{(1)} \oplus \mathcal{X}_s^{(1)}, \quad \mathcal{U}_g^{(1)} = \mathcal{X}_s^{(2)} \oplus \mathcal{X}_g^{(1)} \quad (26)$$

where $\mathbf{X}^{(0)}$ denotes the initial node data.

Concluding, we provide some comments on the convolution process. First, we should notice that the formation of hyperedges in H_s and H_g is performed in a regional manner. Specifically, the neighboring nodes of a central VN are confined to belong to the same regional subgraph, as the one of the VN. By doing so, we decouple the convolutions conducted across the different subgraphs. As a result, the acquired feature representation of VNs in each subgraph is imposed to describe the structural characteristics of the respective part of CV. Secondly, the DHGCN block encompasses both local-global and multi-view learning, simultaneously. The former property allows the aggregation of neighboring information at different spatial scales around VNs, while the latter one combines multiple HGCN convolutions associated with the different shape features of the views. Most importantly, the learnable weight matrices $(\mathbf{W}_{s,g}^{(u)}, \mathbf{W}_{s,g}^{(b)}, \mathbf{W}_{s,g}^{(vol)})$ place proper attention to the views $(\mathbf{X}\mathbf{F}^{(u)}, \mathbf{X}\mathbf{F}^{(b)}, \mathbf{X}\mathbf{F}^{(vol)})$, according to their importance in the convolutions. Lastly, the succession of the four layers in DHGCN implicitly increases the neighborhood depth around central nodes. This suggests that VNs at later layers aggregate neighboring information from broader surface/volume areas, which has a beneficial effect on the obtained feature representations.

6.2. Graph Coarsening and Readout

The coarsening operation performs graph pooling, aiming to reduce the size of the original graph by retaining a portion of representative nodes. To achieve the above task, we harness the principles of the *SagPool* [47] approach, which is adapted to the HGCN convolutions and the cartilage shape needs. *SagPool* is a self-attentional graph pooling method, leveraging both the node features and the graph topology. Particularly, it uses the self-attention scores attained via graph convolution to distinguish between the nodes to be preserved and those to be dropped out.

In this work, the self-attention scores of nodes are computed using HGCN convolutions, as follows:

$$\mathbf{Z}\mathbf{S}^{(l)} = \sigma\left((\mathbf{D}_{n,g})^{-1/2} \mathbf{H}_g \mathbf{W}_g (\mathbf{D}_{e,g})^{-1} (\mathbf{H}_g)^T (\mathbf{D}_{n,g})^{-1/2} \mathbf{Z}^{(l)} \boldsymbol{\Theta}_{pool}^{(l)}\right) \quad (27)$$

where $\mathbf{Z}\mathbf{S}^{(l)} \in \mathbb{R}^{N \times 1}$ contains the self-attention scores of VN nodes, $\mathbf{Z}^{(l)}$ is the output of the DHGCN at layer l of *C_Shape.Net*, and $\boldsymbol{\Theta}_{pool}^{(l)} \in \mathbb{R}^{(N \times 1)}$ is the vector of learnable parameters, used for the embedding of node features.

The scores in $\mathbf{Z}\mathbf{S}^{(l)}$ are sorted in descending order, providing a global ranking list for all VNs in the entire graph. Next, based on $\mathbf{Z}\mathbf{S}^{(l)}$, we create distinct ranking lists $\mathbf{Z}\mathbf{S}_r^{(l)}, r = 0, \dots, 4$, for each regional graph, individually. Then, for each regional graph, we select the top-ranked VNs:

$$idx_r = top\text{-}rank\left(\mathbf{Z}\mathbf{S}_r^{(l)}, [kN_r]\right) \quad (28)$$

where $k \in (0, 1]$ is the pooling ratio, determining the portion of VNs of the input graph to be preserved, and N_r is the number of VNs contained in the r -th regional graph. The rest of VNs are masked, and the respective adjacency matrices are reshaped accordingly. Given that the great majority of VNs lie on the joint space graph $\mathcal{G}_0$, the regional node selection described above is dictated by the fact that we should not risk the case of discarding VNs in the less populated lateral subgraphs, which possibly convey valuable information for the description of their respective volumes.

The readout unit aggregates all the node shape features in the graph, which are updated via graph convolutions by the DHGCN, at the node level. Thus, it provides a

global (pooled) shape feature representation of the hypergraph. In our case, we apply the max, min, and average pooling operators to create the global shape feature, at layer l :

$$\mathcal{Z}_{glo} = \left(\max\{\mathcal{Z}_{(i)}^{(l)}\}, \min\{\mathcal{Z}_{(i)}^{(l)}\}, \text{avg}\{\mathcal{Z}_{(i)}^{(l)}\} \right), \quad l = 1, \dots, M \quad (29)$$

The min and the average pooling operators are introduced with the scope to extract useful information regarding the important indicators of joint-space width and the cartilage volume. The overall global shape feature is obtained by concatenating the pooled features at all layers in the hierarchy of the *C_Shape.Net*:

$$\mathcal{Z}_{glo} = \bigoplus_{l=1}^M \mathcal{Z}_{glo}^{(l)} \quad (30)$$

$\mathcal{Z}_{glo}$ is then submitted to a fully connected network which generates the final 3D shape descriptor $\mathcal{S} = FC(\mathcal{Z}_{glo})$. The final output of the *DHGCN* is obtained by

$$\mathcal{S}_{med/lat} = FC(\mathcal{Z}_{glo}) \quad (31)$$

From a structural point of view, the knee cartilage comprises two distinct compartments, namely, the medial and lateral cartilage volumes. These volumes exhibit their own shape characteristics and convey complementary evidence towards *KOA* grading of the entire knee of a subject. To circumvent this issue, we opt to integrate these two volumes by considering a two-branch architecture, whereby each branch implements the exact pipeline of Figure 4. Concretely, the slides of the knee *MRI* are suitably partitioned into two groups along the medial-lateral direction, which decomposes the overall cartilage into the medial and lateral bodies. These bodies are then processed as described in Section 6, including the formation of the cartilage meshes and graphs, as well as the extraction of face and volumetric features in Section 5.4. After passing through the two-branch *C_Shape.Net*, we obtain the 3D shape descriptors $\mathcal{S}_{med}$ and $\mathcal{S}_{lat}$, corresponding to the medial and lateral shapes, respectively. The total 3D shape descriptor of the entire knee shape is then attained by

$$\mathcal{S} = \lambda \mathcal{S}_{med} + (1 - \lambda) \mathcal{S}_{lat} \quad (32)$$

where the parameter λ is utilized to learn the balance between the two volumes according to their importance.

7. KOA Prediction Network

In this section, we elaborate on the proposed *ST_HGCN* predictor network used for longitudinal *KOA* predictions. The model comprises three main parts. The first part performs spatial *HGCN* convolutions for the different time steps of the historical data, while the second one implements an attention-based temporal fusion of the convolutional feature representations. Finally, a transformer network is designed to transform the knee's feature sequences to future sequences of *KOA* predictions.

7.1. Notations and Problem Definition

The knee record used in the following analysis mainly includes the 3D shape feature obtained by the *C_Shape.Net* and the demographic features {Age, BMI, Gender}. Temporally, the data are distinguished here into two stages, namely, the historical stage and the prediction stage. The historical stage extends to a depth of P years, containing the knee data at time steps $\{t_0, t_1, \dots, t_P\}$, where t_0 corresponds to the baseline visit. Further, the prediction stage refers to the future period of the next T years $\{t_{P+1}, t_{P+2}, \dots, t_{P+T}\}$.

Figure 6 shows the evolution of the different hypergraphs over time, which indicates our spatial-temporal approach of how the patients' knee data are represented. Along the vertical direction, at each time slice $t = t_0, t_1, \dots$, the collection of knees is spatially arranged to form a graph, $\mathcal{G}_t = (\mathcal{V}_t, \mathcal{E}_t, \mathbf{H}_t, \mathbf{XS}_t)$. $\mathcal{V}_t$ is the set of N nodes ($|\mathcal{V}_t| = N$), whereby each node corresponds now to a specific knee, with their data $\mathbf{XS}_t$ considered at time t . $\mathcal{E}_t$ and $\mathbf{H}_t$ represent the set of different hyperedges between the nodes and the incidence matrix, at time t . Moreover, along the horizontal (temporal) direction, each node is associated to a respective sequence of shape data, evolving over time.

At the historical stage, the node data matrix $\mathbf{XS}_t$ at time steps $t = 0, 1, \dots, P$ is defined as

$$\mathbf{XS}_t = (\mathbf{s}_t^{(1)}, \dots, \mathbf{s}_t^{(i)}, \dots, \mathbf{s}_t^{(N)}) \in \mathbb{R}^{(N \times F)} \quad (33)$$

where $\mathbf{s}_t^{(i)} \in \mathbb{R}^F$ denotes the F -dimensional shape feature vector of the i -th node at time t . We also consider the matrix

$$\mathbf{XS}_P = (\mathbf{XS}_0, \dots, \mathbf{XS}_t, \dots, \mathbf{XS}_P) \in \mathbb{R}^{(N \times P \times F)} \quad (34)$$

which subsumes the node data at all time slices of the historical stage. For the time steps of the prediction stage $t' = P + 1, \dots, P + T$, we define the matrix

$$\mathbf{Y}_{t'} = (\mathbf{y}_{t'}^{(1)}, \dots, \mathbf{y}_{t'}^{(i)}, \dots, \mathbf{y}_{t'}^{(N)}) \in \mathbb{R}^{(N \times C)} \quad (35)$$

where $\mathbf{y}_{t'}^{(i)} \in \mathbb{R}^C$ is the label vector of KOA grades in the C cartilage classes. Similarly, the matrix

$$\mathbf{Y}_T = (\mathbf{Y}_{P+1}, \dots, \mathbf{Y}_{t'}, \dots, \mathbf{Y}_{P+T}) \in \mathbb{R}^{(N \times T \times C)} \quad (36)$$

subsumes the label vectors across all nodes and time slices of the prediction stage.

We now consider the node sequence

$$\mathbf{XS}_P(i) = (\mathbf{s}_0^{(i)}, \dots, \mathbf{s}_t^{(i)}, \dots, \mathbf{s}_P^{(i)}) \in \mathbb{R}^{(P \times F)}, \quad i = 1, \dots, N \quad (37)$$

which includes the shapes of a node across all time steps of the historical stage. This sequence represents the shape evolution of a specific knee over time. Similarly, we introduce the label sequence

$$\mathbf{Y}_T(i) = (\mathbf{y}_{P+1}^{(i)}, \dots, \mathbf{y}_{t'}^{(i)}, \dots, \mathbf{y}_{P+T}^{(i)}) \in \mathbb{R}^{(T \times C)}, \quad i = 1, \dots, N \quad (38)$$

which contains the target label vectors $\mathbf{y}_{t'}^{(i)} \in \mathbb{R}^{(C \times 1)}$ of a node, at the prediction stage. We let $\mathbf{y}_{t'}^{(i)}(k) = 1$ if the patient i belongs to the k -th KL class, and $\mathbf{y}_{t'}^{(i)}(k) = 0$ otherwise.

In view of the above definitions, the forecasting problem is posed as follows: Using exclusively the historical data of nodes $\mathbf{XS}_P$, determine the multi-step ahead KOA predictions $\hat{\mathbf{Y}}_T = (\hat{\mathbf{Y}}_{P+1}, \dots, \hat{\mathbf{Y}}_{t'}, \dots, \hat{\mathbf{Y}}_{P+T}) \in \mathbb{R}^{(N \times T \times C)}$. For each node, the above task implements a mapping between the sequences $\mathbf{XS}_P(i) \rightarrow \mathbf{Y}_T^{(i)}, i = 1, \dots, N$. The latter indicates the sequence-to-sequence learning perspective of our approach.

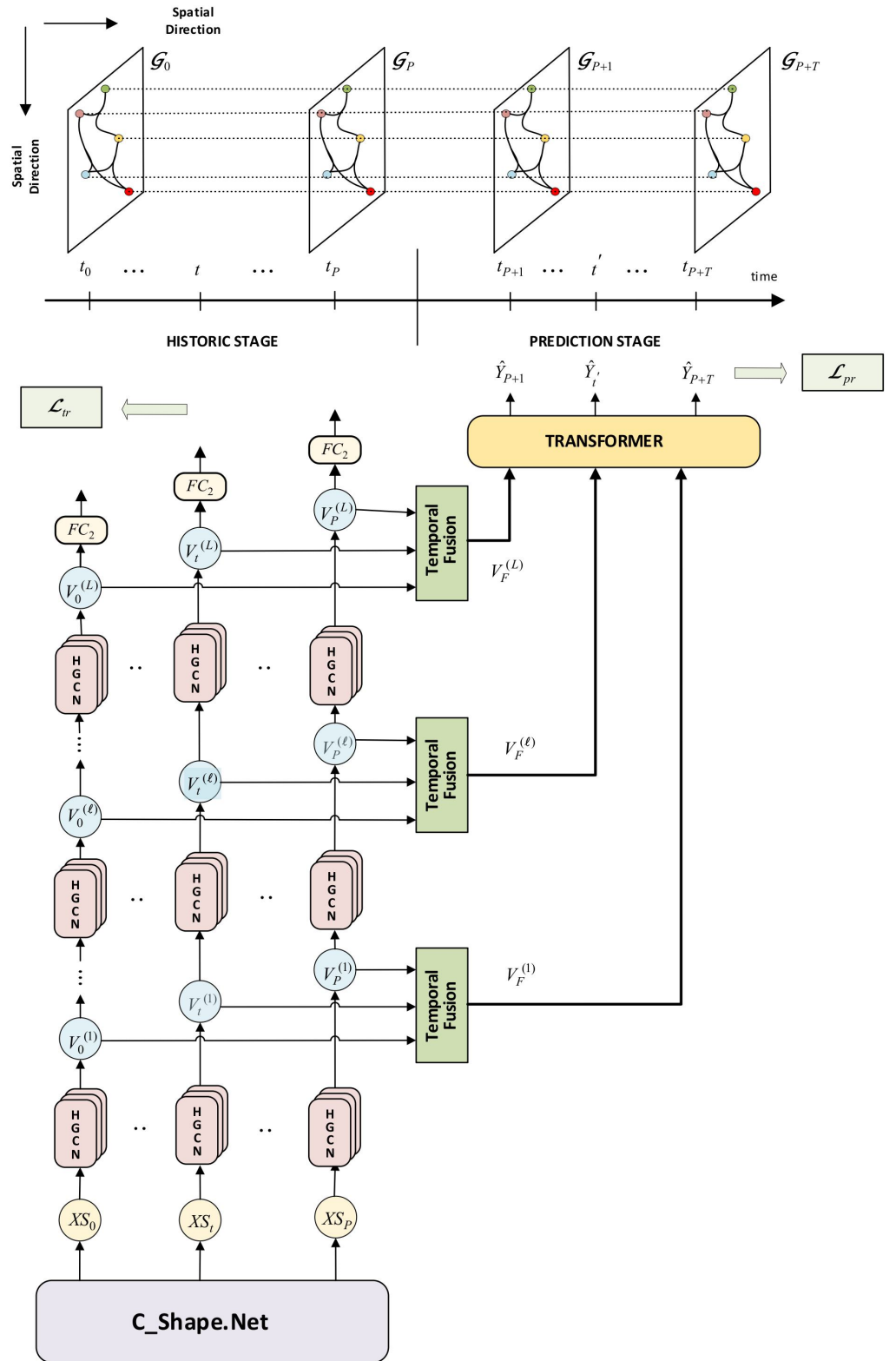


Figure 6. Proposed ST_HGCN Predictor Network for longitudinal KOA predictions. It incorporates the spatial HGCN convolutions, the attention-based temporal fusion, and the transformer module. Data are distinguished into a historical stage of knee shape sequences, and a prediction stage of KOA prediction sequences. We also show the spatio-temporal interconnection of hypergraphs along the historic/prediction follow-up times.

7.2. Spatial Convolutions

The node shape data are processed in this part by establishing P independent convolutional branches of L layers (Figure 6), where each branch performs multi-view HGCN

convolutions on the graphs $\mathcal{G}_t, t = 1, \dots, P$, of the historical stage. The aim of this task is to acquire more descriptive shape representations, taking into consideration the various relationships among the nodes of the knee dataset. For the formation of the hyperedges, we consider four different views, corresponding to the shape feature and the three demographic features, respectively:

$$\mathbf{H}_t = [\mathbf{H}_t^{(sp)}, \mathbf{H}_t^{(age)}, \mathbf{H}_t^{(BMI)}, \mathbf{H}_t^{(gender)}] \in \mathbb{R}^{(N \times M)} \quad (39)$$

$\mathbf{H}_t^{(sp)} \in \mathbb{R}^{(N \times N)}$ includes the spectral affinities between the nodes, i.e., the similarities between their shape features. They are adaptively constructed at each layer using the AHL mechanism in Section 3.3. The latter three parts in Equation (39) encode pairwise node relations in terms of age group, BMI, and gender, and are determined differently. For instance, the age domain is distinguished into four age groups (hyperedges): $\{< 50, [50, 59), [60, 69), \geq 70\}$. Each hyperedge is described by a membership function:

$$\mu_j(\mathbf{x}) = \exp\left[-(x - p_j)^2 / \sigma^2\right], \quad j = 1, \dots, 4 \quad (40)$$

where p_j is a central value of the group and σ is the standard deviation between all nodes in $\mathcal{V}_t$. Then, the degree of membership of the i -th node in the j -th hyperedge is defined as $\mathbf{H}_t^{(age)}(i, j) = \mu_j(\mathbf{x}_i) \in [0, 1]$. Similarly, we create three hyperedges for $\mathbf{H}_t^{(BMI)}$ (Normal, Pre-obesity, Obesity) and two hyperedges for $\mathbf{H}_t^{(gender)}$ (Male, Female), respectively, leading to a total number of $M = N + 9$ hyperedges. Contrary to the spectral $\mathbf{H}_t^{(sp)}$, the three demographic incidence matrices remain fixed across the layers and time steps.

At each time step of the historical stage $t = t_0, \dots, t_P$, the feature embeddings $\mathbf{V}_t^{(l+1)} \in \mathbb{R}^{(N \times F^{(l+1)})}$ acquired by the spatial HGCN at layer $l = 0, 1, \dots, L$ are given by

$$\mathbf{V}_t^{(l+1)} = \sigma\left((\mathbf{D}_{n,t})^{-1/2} \mathbf{H}_t \mathbf{W}_t (\mathbf{D}_{e,t})^{-1} (\mathbf{H}_t)^T (\mathbf{D}_{n,t})^{-1/2} \mathbf{U}_t^{(l)} \boldsymbol{\Theta}_{pr,t}^{(l)}\right) \quad (41)$$

$\mathbf{D}_{n,t}$, $\mathbf{D}_{e,t}$, and $\mathbf{W}_t$ are the respective degree matrices and the learnable hyperedge weights. $\boldsymbol{\Theta}_{pr,t}^{(l)}$ denotes the embedding matrix of tuneable parameters, while $\mathbf{U}_t^{(l)}$ is the input signal at layer l . For $l = 0$, we have $\mathbf{U}_t^{(0)} = \mathbf{X}\mathbf{S}_t$, where $\mathbf{X}\mathbf{S}_t$ contains the original shape data, and $\mathbf{U}_t^{(0)} = \mathbf{V}_t^{(l)}$ for later layers, at branch t . The demographic data $\mathbf{X}\mathbf{D}_t$ are implicitly used at each layer to generate their respective hyperedges.

7.3. Attention-Based Temporal Fusion

At each layer of spatial convolutions, we apply an attention mechanism with the aim to aggregate the feature representations $\{\mathbf{V}_0^{(l)}, \dots, \mathbf{V}_t^{(l)}, \dots, \mathbf{V}_P^{(l)}\}, l = 1, \dots, L$, along the temporal direction of the historical stage. As a first step, we compute the following self-attention scores:

$$g_t^{(l)} = \sigma\left(\tilde{\mathbf{V}}_t^{(l)} \mathbf{W}_F^{(l)} + \mathbf{b}_F^{(l)}\right) \quad (42)$$

where $\tilde{\mathbf{V}}_t^{(l)} \in \mathbb{R}^{(N \cdot F)}$ is a flattened one-dimensional vector of $\mathbf{V}_t^{(l)}$, while $\mathbf{W}_F^{(l)}$ and $\mathbf{b}_F^{(l)}$ are the weight embedding matrix and bias of learnable parameters, shared across all nodes and time steps. The attentional scores are then normalized via the softmax function as

$$\beta_t^{(l)} = \frac{\exp[g_t^{(l)}]}{\sum_{q=1}^P \exp[g_q^{(l)}]} \quad (43)$$

Finally, the multi-temporal aggregation is formed by

$$\mathbf{V}_F^{(l)} = \sum_{t=1}^P \beta_t^{(l)} \mathbf{V}_t^{(l)} \quad (44)$$

The obtained fusion $\mathbf{V}_F^{(l)}$ pays due attention to specific time steps of the node shape sequences of the historical data. In that respect, it can capture significant temporal changes in the knee shapes (cartilage volumes), which provides valuable evidence for assessing KOA incidence and progression.

7.4. Sequence Transformation

In this part of the predictor, we develop an FC-based transformer network to transform the historical sequences of shape feature embeddings to future sequences of KOA grade estimates. Formally, the transformer performs the following sequence mapping:

$$\mathbf{V}_F = (\mathbf{V}_F^{(1)}, \dots, \mathbf{V}_F^{(l)}, \dots, \mathbf{V}_F^{(L)}) \xrightarrow{\text{TRANSF}} \hat{\mathbf{Y}}_T = (\hat{\mathbf{Y}}_{P+1}, \dots, \hat{\mathbf{Y}}_{P'}, \dots, \hat{\mathbf{Y}}_{P+T}) \quad (45)$$

where $\mathbf{V}_F \in \mathbb{R}^{(N \times L \times F)}$ is the input shape feature data, formed as the concatenation of the fusion results $\mathbf{V}_F^{(l)}$, at all layers of the spatial convolutions of the historical stage. The output $\hat{\mathbf{Y}}_T \in \mathbb{R}^{(N \times T \times C)}$ comprises the KOA predictions of nodes $\hat{\mathbf{Y}}_{P'} \in \mathbb{R}^{(N \times C)}$, at all times of the prediction stage. As a first step, $\mathbf{V}_F$ is reshaped as a two-dimensional matrix. Next, we use T two-layered fully connected networks to produce the multi-step ahead predictions [48]:

$$\hat{\mathbf{Y}}_{P'} = \sigma(\tilde{\mathbf{V}}_F \mathbf{W}_1^{(tr)} + \mathbf{b}_1^{(tr)}) \cdot \mathbf{W}_2^{(tr)} + \mathbf{b}_2^{(tr)} \quad (46)$$

where $\mathbf{W}_1^{(tr)} \in \mathbb{R}^{(L \times F \times \tilde{C})}$ and $\mathbf{b}_1 \in \mathbb{R}^{\tilde{C}}$ denote the embedding matrix and bias of the first layer used to embed the inputs to an intermediate $\tilde{C}$ -dimensional space. $\mathbf{W}_2^{(tr)} \in \mathbb{R}^{(\tilde{C} \times C)}$ and $\mathbf{b}_2 \in \mathbb{R}^C$ are the respective weights and bias of the second layer, and $\sigma(\cdot)$ denotes the *ReLU* function. The above learnable parameters are shared across all nodes (patients) of the graph. The outcomes from Equation (46) are passed through a softmax layer to obtain the final KOA grades, which are submitted to the prediction loss.

The predictor's training is accomplished by applying semi-supervised learning (SSL). Specifically, the shape data $\mathcal{X}\mathcal{S}_P$ is composed of a labeled part $\mathcal{X}\mathcal{S}_{P,l} \in \mathbb{R}^{(N_l \times P \times F)}$ and an unlabeled part $\mathcal{X}\mathcal{S}_{P,u} \in \mathbb{R}^{(N_u \times P \times F)}$ with $N = N_l + N_u$. The former part contains the nodes with historical shape sequences having labeled sequences of KOA grades in the prediction stage. The nodes in the latter part form the testing dataset, and hence, their output sequences are unknown. Following the SSL approach, we take leverage of the shape contents of both the training and the testing data, which can lead to more enhanced forecasting results. Finally, we employ the cross-entropy measure over the labeled samples to penalize the errors between the ground true labels and the predictor's estimates:

$$\mathcal{L}_{pr} = \sum_{t=P+1}^T \sum_{c=1}^C \sum_{p=1}^{N_l} \mathcal{Y}_T(t, c, p) \ln(\hat{\mathbf{Y}}_T(t, c, p)) \quad (47)$$

where $\mathcal{Y}_T(t, c, p)$ indicates a labeled node at time t and KL class c . Using Equation (47), we perform an end-to-end training of the architecture in Figure 6, including the different parts of the predictor and the $C_Shape.Net$.

8. Experimental Results

Our proposed approach is comprehensively evaluated in this section on the widely used *OAI* dataset. In the experiments, we validate the produced longitudinal *KOA* predictions for varying depths of the historical stage as defined by P . Concretely, given the maximum length of 8 follow-up years for the patient sequences, we consider four different historical data cases $\{t_0, \dots, t_P\}$, assuming that P takes values in the range $P \in \{0, 1, 2, 3\}$. The case of $P = 0$ corresponds to the graph data $\mathcal{G}_0$, i.e., it contains the population of $N = 6228$ knees at the baseline visit. The case $P = 1$ subsumes the graphs $\{\mathcal{G}_0, \mathcal{G}_1\}$ (Figure 6), i.e., it encompasses the $2 \cdot N$ knees at the baseline visit and the first follow-up step, and so on. Finally, the case $P = 3$ defines the maximum depth of the historical data, incorporating the $4 \cdot N$ knees at the first four consecutive visits. Then, given a specific historical case $\{t_0, \dots, t_P\}$, its respective prediction stage is extended to the next four time steps $\{t_{P+1}, \dots, t_{P+4}\}$, i.e., $T = 4$, where longitudinal *KOA* predictions are produced. MRI and shape data are used solely in the historical stage. The above experimental design is primarily dictated by space limitations. Nevertheless, our approach is general, allowing the consideration of alternative prediction settings, according to the user's preference and the database contents.

The quality of the acquired predictions ahead is assessed using a variety of accuracy measures. The dataset is partitioned into training, validation, and testing subsets to a portion of 60%, 20%, and 20%, respectively. The reported accuracy results are obtained by averaging over five different data partitions. The various hyperparameters pertaining to our networks are selected through 5-fold cross-validation.

8.1. Implementation Details

All models presented in this study were developed using the PyTorch 2.7 deep learning framework (<https://pytorch.org/> (accessed on 1 September 2024)). The initial feature engineering step, involving the creation of 3D surface meshes for extraction of volumetric and surface features, which are then processed by the *C_Shape.Net*, is carried out utilizing the trimesh (<https://trimesh.org/> (accessed on 1 September 2024)) and CGAL (<https://www.cgal.org/> (accessed on 1 September 2024)) libraries. Table 2 holds the list of parameters with regards to model architecture, and their respective values. To better streamline the hyperparameter optimization process, we made use of the Optuna (<https://optuna.org/> (accessed on 1 September 2024)) open source optimization framework. After an initial experimentation stage, the layer dimensionalities for the *LMV* and *GMV* units comprising the *C_Shape.Net*, and those of the 2-layer *FC-Transformer* network were set to 256, 64, and [256, 128], respectively. The models were trained for a maximum number of 500 epochs, whereby training was halted by an early stopping criterion, after no reduction had been recorded in the validation set for more than 20 consecutive epochs. The initial learning rate was set to 0.1, with a multi-step scheduler that decays that number by a factor of 10 at 3 evenly spaced milestones. The code for all models, feature extraction pipeline, and the training process utilized in this study, can be found at https://gitlab.com/mri_cartilage_segmentation/longitudinal_KOA_prediction (accessed on 15 April 2025).

In the following, we present a series of experiments to examine different aspects of our method.

Table 2. Summary of the proposed model's hyperparameter ranges and optimal values determined via Grid-Search + 5-fold Cross-Validation. Bold indicates the optimal parameter value.

	#Attention-Heads	Pooling Ratio	Convex Comb. Parameter (λ)	Batch Size	Dropout Prob.
Range & Opt. value	[2, 4, 6 , 8]	[.20, .40 , 0.60, .80]	[.20, .40, .60 , .80]	[64, 128, 256]	[.25, 0.50, .75]

8.2. Node KOA Classifications

In this experiment, we elaborate on the node classification task, i.e., generate *KL*-grade predictions for each node (knee) of the historical stage ($P = 3$), individually. The accuracy of these predictions indicates primarily the efficacy of the shape descriptors given by *C_Shape.Net*, and secondarily that of the spatial convolutions.

All knees of the historical data matrix $\mathcal{XS}_P = \mathbb{R}^{(N \times P \times F)}$ are considered herein as independent nodes in an overall graph, regardless of the time they are located. Similarly to Equation (36), let $\mathcal{Y}_P = (Y_0, \dots, Y_t, \dots, Y_P) \in \mathbb{R}^{(N \times P \times C)}$ denote the matrix of *KOA* labels of nodes in the historical stage. In that case, the network conducts spatial *HGCN* convolutions on all nodes of $\mathcal{XS}_P$. The finally acquired feature representations are passed through a two-layered network (*FC2*) to obtain *KOA* predictions, which are subsequently submitted to the training loss $\mathcal{L}_{tr}$. The goal is then to establish a classification mapping $\mathcal{XS}_P \rightarrow \mathcal{Y}_T$ after semi-supervised learning. The knee temporal sequences are dissolved, and no future predictions are sought; the attention-based temporal fusion mechanism and the transformer unit are herein disregarded.

Table 3 shows the node classification accuracies for various depths of the historical data. The results are evaluated in terms of the Balanced overall Accuracy (*BA*) on the five *KL* classes, the *F1-score*, and Cohen's Kappa coefficient (*K*). The results indicate a comparable performance for the medial and lateral knee compartments, with a slight edge observed for the latter part. Regarding the *Kappa* coefficient, it can be seen that the combined treatment yields a meaningful boost. Noticeably, the suggested predictor provides a sufficiently high performance, with the accuracies varying between 84.07 and 95.98%, 85.94 and 97.47%, and 86.29 and 97.92%, in regard to the *BA*, *F1-score*, and *Kappa* measures, respectively. The lower results are obtained when learning is performed on the baseline data ($P = 0$). Moreover, it can be noticed that the accuracies are proportionally increased for higher depths of the historical data, with the best scores achieved for $P = 3$. This trend is attributed to the fact that for higher P -values, the datasets are progressively supplemented with more severe *KOA* cases, i.e., knees exhibiting greater *KL* grades (Table 3).

Table 3. Performance of node *KOA* classifications under varying-size (P) of historical windows. Results are reported for the Medial/Lateral compartments and their combination. **Metrics:** [Balanced Accuracy (*BA*), F1-weighted measure (*F1*), Cohen's Kappa (*K*)].

P	Medial			Lateral			Medial + Lateral		
	BA	F1	K	BA	F1	K	BA	F1	K
0	84.21%	86.05%	85.61%	83.92%	85.88%	87.03%	84.07%	85.94%	86.29%
	(±0.054)	(±0.081)	(±0.049)	(±0.087)	(±0.101)	(±0.024)	(±0.092)	(±0.037)	(±0.062)
1	88.46%	89.13%	88.24%	87.69%	88.71%	90.54%	88.12%	89.02%	90.22%
	(±0.032)	(±0.114)	(±0.078)	(±0.094)	(±0.045)	(±0.079)	(±0.117)	(±0.065)	(±0.089)
2	92.18%	93.01%	94.17%	93.09%	95.08%	95.12%	92.51%	93.98%	94.79%
	(±0.028)	(±0.076)	(±0.088)	(±0.039)	(±0.105)	(±0.066)	(±0.082)	(±0.057)	(±0.083)
3	96.82%	97.19%	97.83%	95.47%	97.85%	98.61%	95.98%	97.49%	97.92%
	(±0.067)	(±0.092)	(±0.11)	(±0.039)	(±0.075)	(±0.041)	(±0.056)	(±0.083)	(±0.077)

8.3. Longitudinal KL Predictions

Table 4 shows the longitudinal *KOA* grading accuracies for varying depths of the historical data. The results are evaluated in terms of *BA*, *F1-score*, and *K* at the different yearly steps ahead, along with the overall accuracy *OA* on the entire prediction stage. In this experiment, we deploy the complete setup of our *ST_HGCN* predictor, including the spatial convolutions, the temporal fusion, and the (historic) sequence-to-(*KL*) sequence transformer.

Table 4. Performance of KL longitudinal predictions under varying-size (P) of historical windows. **Metrics:** [Balanced Accuracy (BA), F1-weighted measure (F1), Cohen’s Kappa (K), Overall Accuracy averaged across future predictions ($\tilde{t}_H$)].

P	t_{P+1}			t_{P+2}			t_{P+3}			t_{P+4}			$\tilde{t}_P$
	BA	F1	K	BA	F1	K	BA	F1	K	BA	F1	K	OA
0	67.49%	65.19%	65.02%	60.24%	62.19%	61.75%	58.02%	60.13%	59.88%	54.19%	56.01%	55.41%	60.02%
	(± 0.089)	(± 0.075)	(± 0.093)	(± 0.105)	(± 0.088)	(± 0.164)	(± 0.079)	(± 0.112)	(± 0.069)	(± 0.122)	(± 0.096)	(± 0.080)	(± 0.101)
1	75.84%	77.01%	76.73%	72.16%	74.28%	73.64%	68.07%	69.54%	67.91%	62.08%	64.10%	63.59%	69.53%
	(± 0.068)	(± 0.082)	(± 0.075)	(± 0.094)	(± 0.088)	(± 0.059)	(± 0.076)	(± 0.081)	(± 0.069)	(± 0.081)	(± 0.104)	(± 0.095)	(± 0.111)
2	84.95%	86.61%	85.39%	79.28%	80.37%	77.87%	75.17%	77.23%	76.49%	71.58%	73.14%	72.01%	77.74%
	(± 0.048)	(± 0.066)	(± 0.081)	(± 0.057)	(± 0.062)	(± 0.074)	(± 0.088)	(± 0.053)	(± 0.071)	(± 0.099)	(± 0.071)	(± 0.057)	(± 0.094)
3	91.89%	93.13%	92.81%	88.11%	89.69%	88.81%	84.35%	87.06%	85.92%	79.41%	82.02%	80.19%	85.94%
	(± 0.052)	(± 0.067)	(± 0.058)	(± 0.078)	(± 0.095)	(± 0.072)	(± 0.088)	(± 0.064)	(± 0.083)	(± 0.093)	(± 0.069)	(± 0.101)	(± 0.091)

The first observation is that the historical depth plays a crucial role in the KL grading performance. Concretely, the larger the length of the patients’ sequences, the higher the accuracies attained at follow-up times. The worst performance is obtained for $P = 0$ (first row), exhibiting BA in the range (54.19–67.49%) and $OA = 60.02\%$. In that case, learning relies solely on the baseline data, with the temporal fusion mechanism essentially deactivated. Increasing the historic sequence length elevates the accuracy values considerably. The best results are acquired for $P = 3$ (maximum depth), which provides a sufficiently high BA between 79.41 and 91.89% and $OA = 85.94\%$. The enhanced performance is primarily due to the temporal fusion, which is now able to detect significant shape trends across the knees’ historic sequence. To highlight the influence of historic sequence length, let us consider the KOA grading of a patient at the fourth follow-up visit ahead of the baseline (+48 months), as derived from four different values of P . As can be seen from Table 4, we obtain the accuracies of 54.19% ($t_{P+4}, P = 0$), 68.07% ($t_{P+3}, P = 1$), 79.28% ($t_{P+2}, P = 2$), and 91.89% ($t_{P+1}, P = 3$), respectively, which suggests that higher historic depths lead to sufficiently better predictions.

The second observation is that, reasonably, the KOA grading accuracies at later follow-up times are reduced, compared to the one-step ahead predictions. This trend applies across all historic depths, where we can notice a reduction of about 13% for the accuracies at t_{P+4} in contrast to the ones at t_{P+1} . Specifically, for $P = 3$, the one-step ahead prediction achieves an accuracy of 91.89% at t_{P+1} (fourth follow-up visit) which is gradually diminished to 79.40% at t_{P+4} (seventh follow-up visit). The drop in accuracies can be justified by the fact that, apart from the historical data, no shape, radiological, or other information is utilized during the prediction stage. Similar findings in the experiment can be also drawn in terms of the other quality metrics.

In order to statistically verify the above observations, we perform Nemenyi’s post hoc test [49] after rejecting the null hypothesis of statistical equivalence between the examined cases via the use of the Friedman statistical test [50]. The results presented in Table 5 allow us to confirm the fact that expanding the depth of the historical record considered by *ST_HGCN*, we are able to attain statistically significant improvements in performance.

Table 5. Pairwise comparisons using the Nemenyi statistical test. We evaluate the performance of *ST_HGCN* under various historical depths for multiple step-ahead predictions. Reported are the corresponding p -values.

	t_{P+1}			t_{P+2}			t_{P+3}			t_{P+4}		
	$P = 0$	$P = 1$	$P = 2$	$P = 0$	$P = 1$	$P = 2$	$P = 0$	$P = 1$	$P = 2$	$P = 0$	$P = 1$	$P = 2$
$P = 1$	1.91×10^{-4}	-	-	4.97×10^{-5}	-	-	7.19×10^{-4}	-	-	2.07×10^{-3}	-	-
$P = 2$	5.44×10^{-6}	0.22×10^{-6}	-	3.18×10^{-5}	7.54×10^{-4}	-	1.33×10^{-5}	0.87×10^{-3}	-	5.08×10^{-4}	9.17×10^{-3}	-
$P = 3$	3.62×10^{-12}	7.41×10^{-08}	9.37×10^{-6}	8.49×10^{-10}	1.65×10^{-7}	6.44×10^{-6}	3.27×10^{-9}	1.98×10^{-7}	8.27×10^{-4}	4.37×10^{-6}	0.59×10^{-5}	1.28×10^{-5}

8.4. Demographic-Based KL Accuracies

Table 6 hosts the KOA accuracies at future times for different demographic groups, categorized by Gender, Age, and BMI. Regarding the gender groups, the accuracies for the female group are consistently superior compared to the male group, across all time steps ahead. For the Age groups, the three lower age groups are more accurately identified in the KL classes, as contrasted to the elderly group of patients, which shows a relatively degraded performance. We can also notice a general trend of reducing performance as age advances from younger to more aged patients. Finally, in terms of BMI groups, there is also a slight tendency of diminishing accuracies as BMI moves from normal to the obese group. Normal and pre-obese patients are more accurately classified, while the obese ones exhibit reduced scores compared to the average performance in the population.

Table 6. Balanced accuracy (BA) of KL longitudinal predictions for a historical depth of $P = 3$ w.r.t. demographic groups of {Gender, Age, and BMI}.

	Gender		Age				BMI		
	Male	Female	<50	[50, 59]	[60, 69]	≥ 70	Normal	Pre-Obesity	Obesity
t_{P+1}	90.21% (± 0.116)	92.37% (± 0.099)	93.04% (± 0.078)	92.12% (± 0.091)	89.65% (± 0.107)	88.34% (± 0.077)	91.95% (± 0.077)	93.02% (± 0.079)	87.57% (± 0.072)
t_{P+2}	86.98% (± 0.084)	90.19% (± 0.092)	90.24% (± 0.049)	90.01% (± 0.061)	88.53% (± 0.066)	86.67% (± 0.087)	91.05% (± 0.109)	89.92% (± 0.081)	86.52% (± 0.064)
t_{P+3}	83.47% (± 0.097)	86.02% (± 0.086)	84.49% (± 0.053)	86.72% (± 0.071)	82.48% (± 0.047)	80.91% (± 0.073)	85.82% (± 0.101)	83.91% (± 0.083)	83.11% (± 0.055)
t_{P+4}	77.97% (± 0.072)	80.18% (± 0.098)	81.58% (± 0.119)	81.73% (± 0.091)	79.18% (± 0.066)	78.14% (± 0.081)	82.37% (± 0.059)	79.85% (± 0.082)	76.63% (± 0.090)

8.5. Accuracies According to the Stage of KOA

In this experiment, patients are classified with respect to the degree of severity of their OA. Following the setting in [51], we consider an alternative population partition into three new classes $C = \{C_1, C_2, C_3\}$, which are formed by grouping the KL grades as follows: $C_1 = NO_OA = \{0, 1\}$, $C_2 = EARLY_OA = \{2\}$, and $C_3 = SEVERE_OA = \{3, 4\}$. Table 7 presents the longitudinal KOA predictions in the different classes in C , considering a historic depth $P = 3$. For each of the follow-up times independently, we create three binary classification problems: NO_OA vs. rest, $EARLY_OA$ vs. rest, and $SEVERE_OA$ vs. rest. Based on these formulations, the accuracy is evaluated in terms of each class-specific measures BA , $F1$ -score, and K . We also include the average accuracy over all time steps of the prediction stage.

The results show that the suggested predictor identifies the classes C_1 (mild KOA grades) and C_2 (moderate KOA grades) most accurately, achieving an average BA of 89.72% and 85.95%, respectively. The class C_3 (severe KOA grades) exhibits a lower level of prediction 82.22%. The above average picture is also reflected in the sequences of follow-up predictions. Specifically, the longitudinal predictions for the classes C_1 and C_2 attain enhanced accuracies, between 94.87 and 84.37% and 92.13 and 79.39%, respectively. On the other hand, the prediction scores for C_3 take values in the range 88.65–74.68%. Contrasting the results in the first three rows with those of the fourth row, it can be concluded that patients with mild or moderate KOA severity (C_1, C_2) are identified more precisely, while those with severe KOA grades (C_3) are classified to a relatively lower level of accuracy. We can also notice the previously observed trend for the temporal accuracies. Concretely, for every class, the best performance is obtained for the one-step predictions at t_{P+1} , while predictions at later times are gradually diminishing.

Table 7. Performance of KOA longitudinal predictions for historic depth $P = 3$. Subjects are partitioned according to the stage of KL grades at baseline: *NON-OA*: $\{0, 1\}$, *EARLY-OA*: $\{2\}$, *SEVERE-OA*: $\{3, 4\}$. **Metrics:** [Balanced Accuracy (*BA*), F1-weighted measure (*F1*), Cohen’s Kappa (*K*), Overall Accuracy averaged across future predictions ($\bar{t}_P$)].

	t_{P+1}			t_{P+2}			t_{P+3}			t_{P+4}			$\bar{t}_P$
	<i>BA</i>	<i>F1</i>	<i>K</i>	<i>BA</i>	<i>F1</i>	<i>K</i>	<i>BA</i>	<i>F1</i>	<i>K</i>	<i>BA</i>	<i>F1</i>	<i>K</i>	<i>OA</i>
C_1	94.87% (± 0.041)	95.74% (± 0.067)	95.02% (± 0.77)	91.06% (± 0.089)	92.73% (± 0.103)	91.33% (± 0.097)	88.61% (± 0.082)	89.09% (± 0.099)	88.96% (± 0.105)	84.37% (± 0.055)	86.22% (± 0.079)	84.81% (± 0.101)	89.73% (± 0.083)
C_2	92.13% (± 0.106)	94.05% (± 0.099)	93.49% (± 0.113)	88.14% (± 0.124)	89.43% (± 0.085)	89.07% (± 0.092)	84.19% (± 0.103)	85.67% (± 0.077)	84.55% (± 0.081)	79.39% (± 0.108)	81.27% (± 0.090)	80.02% (± 0.102)	85.94% (± 0.092)
C_3	88.65% (± 0.077)	90.16% (± 0.092)	89.21% (± 0.064)	85.17% (± 0.081)	86.67% (± 0.102)	85.93% (± 0.079)	80.43% (± 0.076)	81.12% (± 0.091)	80.78% (± 0.107)	74.68% (± 0.099)	76.33% (± 0.111)	75.47% (± 0.088)	82.23% (± 0.097)

8.6. Longitudinal Predictions of KOA Progression

This experiment aims at investigating the capabilities of our predictor network to identify the KOA progression of patients, across the follow-up period. That is, we seek to accurately detect the changes in KOA grades occurring over time, as the patient’s disease aggravates from lower to greater grades.

In our setting, the progression $Progr(t, t_{ref})$ is measured as the KOA grade change observed at a future time step $t \in (t_{P+1}, t_{P+2}, t_{P+3}, t_{P+4})$, in regard to a previous reference time point $t_{ref} = t_P$. This choice suggests that all future KOA changes are evaluated in reference to the last time step of the historical stage, with available KOA labels. To quantify the results, we define three progression variables which are derived according to the temporal transitions among the classes $C_1 \rightarrow C_2 \rightarrow C_3$. The variable $No_Progr(t, t_{ref})$ indicates the absence of any KOA change in a patient in the interval (t, t_{ref}) . The variable $Progr_{+1}(t, t_{ref})$ signifies the existence of KOA grade changes causing the following class transitions: $C_1(NO_OA) \rightarrow C_2(EARLY_OA)$ or $C_2(EARLY_OA) \rightarrow C_3(SEVERE_OA)$. Finally, the variable $Progr_{+2}(t, t_{ref})$ indicates the occurrence of significant KL-grade changes, leading to a two-class shift $C_1(NO_OA) \rightarrow C_3(SEVERE_OA)$.

Table 8 hosts the longitudinal progression results at the different follow-up times, considering a historical length of $P = 3$. In that case, the KL classifications provided by the predictor are re-organized according to the three progression classes ($No_Progr, Progr_{+1}, Progr_{+2}$). Similarly to the previous section, the accuracies of progressions are evaluated in terms of the *BA*, *F1*, and the kappa coefficient *K*. The results show that as expected, the No_Progr class of subjects that retain the same grading throughout the prediction stage, achieve high *BA* accuracies (94.49%, 93.16%, 88.47%, 84.72%). Notice that these accuracies are in close accordance with the ones in Table 7 (first row). The $Progr_{+1}$ is a more important class in two aspects. On one hand, it encompasses the critical task of early detection of KOA ($C_1 \rightarrow C_2$) for patients with or without minimal symptoms, which can assist the clinicians to schedule effective treatment strategies. On the other hand, we have the early-to-severe KOA progression ($C_2 \rightarrow C_3$), which is also a critical task for apparent reasons. The $Progr_{+1}$ is also precisely identified, with the accuracy scores varying between (89.75%, 84.25%, 79.39%, 74.37%). Noticeably, the $Progr_{+2}$ class attains a sufficiently high performance of accuracies, which indicates that the predictor can identify the cases of strong grade changes from normal-to-severe KOA ($C_1 \rightarrow C_3$). In Figure 7, we use the Area Under ROC curve (*AUC*) values for various time steps ahead, along with the average *AUC* over the prediction horizon. In average terms, all progressions, and especially the early KOA appearance, are adequately captured with *AUC* higher than 0.8.

Table 8. Prediction accuracies of KOA progression for historical depth $P = 3$. KOA progression of patients over time is categorized according to the following progress classes: $Prog_{+1}$ indicates the progressions $NON-OA \rightarrow EARLY-OA$ and $EARLY-OA \rightarrow SEVERE-OA$, while $Prog_{+2}$ indicates the progression $NON-OA \rightarrow SEVERE-OA$. *No-Prog* refers to subjects that remain static in their respective OA grading. **Metrics:** [Balanced Accuracy (BA), F1-weighted measure (F1), Cohen’s Kappa (K)].

	$Ref \rightarrow t_{P+1}$			$Ref \rightarrow t_{P+2}$			$Ref \rightarrow t_{P+3}$			$Ref \rightarrow t_{P+4}$		
	BA	F1	K	BA	F1	K	BA	F1	K	BA	F1	K
<i>No-Prog</i>	94.49% (± 0.059)	95.02% (± 0.082)	95.15% (± 0.077)	93.16% (± 0.064)	93.52% (± 0.092)	93.74% (± 0.061)	88.47% (± 0.049)	89.23% (± 0.056)	89.02% (± 0.074)	84.72% (± 0.092)	85.01% (± 0.071)	85.61% (± 0.062)
$Prog_{+1}$	89.75% (± 0.087)	90.14% (± 0.092)	89.01% (± 0.076)	84.25% (± 0.107)	84.72% (± 0.114)	83.88% (± 0.091)	79.39% (± 0.123)	80.27% (± 0.099)	77.91% (± 0.104)	74.37% (± 0.086)	76.19% (± 0.091)	75.59% (± 0.098)
$Prog_{+2}$	91.41% (± 0.066)	92.31% (± 0.082)	92.03% (± 0.059)	86.92% (± 0.087)	87.49% (± 0.092)	85.90% (± 0.105)	83.05% (± 0.092)	83.56% (± 0.088)	82.37% (± 0.103)	79.14% (± 0.116)	79.88% (± 0.108)	78.24% (± 0.099)

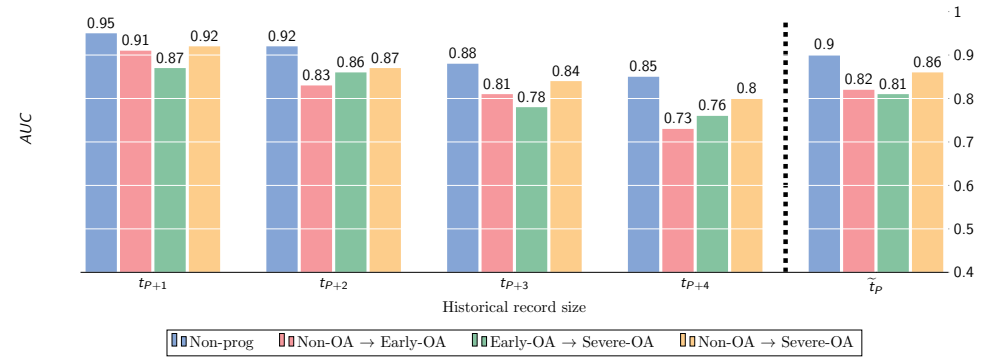


Figure 7. Performance (AUC) w.r.t. prediction of KOA severity progression (maximum historical depth $P = 3$), under different transition scenarios. $\bar{t}_H$ denotes the averaged performance along future time steps.

Figure 8 shows a more detailed picture of the results for varying values of historical lengths. In that case, we examine the KL-grade progression incidence task ($\Delta KL \geq 1$), namely, we predict the KL changes greater than 1 over the prediction stage, compared to the KL grades at the reference time (last time step) of the respective historic periods. For each P , we present the area under ROC curves (AUC) values for the future time steps, along with the average AUC over all times of the prediction period. These results are computed after casting the appropriate binary classification problems (KL progression vs. No Progression). In terms of average AUCs, we achieved progression accuracies in the range (0.81, 0.77, 0.73, 0.68) for increasing values of P . This also indicates that the performance is sufficiently enhanced when considering longer historic periods. The AUC scores receive their highest values at t_{P+1} (fast progression) and gradually lower ones at further time steps (middle- and longer-term progression). Viewing the above issue in a different perspective, suppose we want to predict the progress incidence of a patient 48 months (four time steps) after their first visit (baseline). Figure 8 shows that this can be obtained with an AUC of 0.68 ($P = 0$), 0.76 ($P = 1$), 0.85 ($P = 2$), and 0.94 ($P = 3$), respectively.

8.7. Ablation Studies

In this section, we conduct experiments to demonstrate the impact of several components of the predictor network, including the temporal fusion component, the *C_Shape.Net*, and the incorporation of the demographic risk factors in the convolutions.

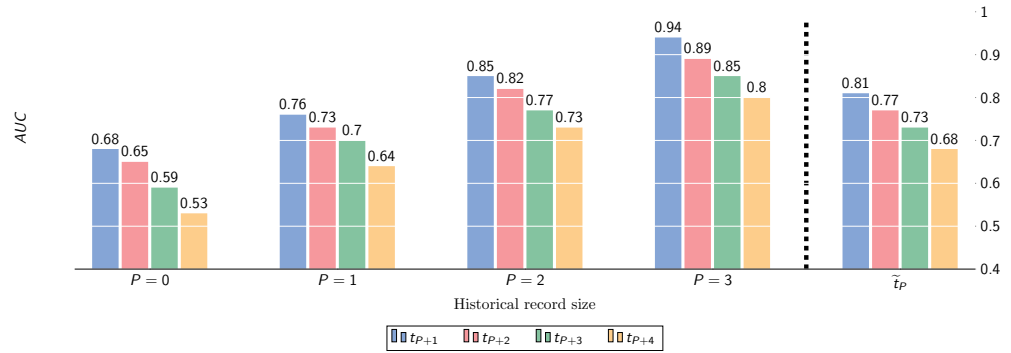


Figure 8. Performance (AUC) w.r.t. prediction of OA progression incidence (KL -grade change ≥ 1) at multiple time points ahead, for varying historic depths P .

8.7.1. Temporal Fusion Effect

Table 9 shows the KOA classifications along the follow-up times, for varying sizes of the historical windows. For each window, we contrast the efficacy of the proposed attention-based temporal fusion (*ATTN*) against the temporal fusion via mean aggregation (*MEAN*). The results demonstrate that the trainable temporal fusion component consistently outperforms the simple mean aggregator by a percentage of 2–6% across all historic depths and future time steps. Noticeably, the accuracy excess increases for larger historic lengths and later future times. Given a sufficiently large historic length, the above finding suggests that the attention-based aggregation on the patients' data sequences is a more effective approach. Particularly, it allows our temporal fusion to discern shape changes, i.e., indicators of joint space narrowing, thus leading to better KOA assessments.

Table 9. Summary of KOA classification results of the proposed method, under varying-size historical windows. We examine the effect of temporal fusion via adaptive attention mechanism *vs.* temporal fusion via mean aggregation. *ATTN* indicates temporal fusion via trainable attention mechanism, while *MEAN* indicates temporal fusion via simple averaging. **Metrics:** [Balanced Accuracy (*BA*), F1-weighted measure (*F1*), Cohen's Kappa (*K*)].

P		t_{P+1}			t_{P+2}			t_{P+3}			t_{P+4}		
		BA	F1	K	BA	F1	K	BA	F1	K	BA	F1	K
P_1	<i>ATTN</i>	75.84% (± 0.068)	77.01% (± 0.082)	76.73% (± 0.075)	72.16% (± 0.094)	74.28% (± 0.088)	73.64% (± 0.059)	68.07% (± 0.076)	69.54% (± 0.081)	67.91% (± 0.069)	62.08% (± 0.081)	64.10% (± 0.104)	63.59% (± 0.095)
	<i>MEAN</i>	73.49% (± 0.092)	75.31% (± 0.104)	74.98% (± 0.112)	70.24% (± 0.141)	71.87% (± 0.098)	71.12% (± 0.132)	64.31% (± 0.107)	65.68% (± 0.113)	64.93% (± 0.127)	58.53% (± 0.106)	59.67% (± 0.089)	59.02% (± 0.117)
P_2	<i>ATTN</i>	84.95% (± 0.048)	86.61% (± 0.066)	85.39% (± 0.081)	79.28% (± 0.057)	80.37% (± 0.062)	77.87% (± 0.074)	75.17% (± 0.088)	77.23% (± 0.053)	76.49% (± 0.071)	71.58% (± 0.099)	73.14% (± 0.071)	72.01% (± 0.057)
	<i>MEAN</i>	80.26% (± 0.095)	81.17% (± 0.108)	80.79% (± 0.102)	75.07% (± 0.088)	76.91% (± 0.090)	75.67% (± 0.112)	71.31% (± 0.131)	73.08% (± 0.108)	72.56% (± 0.143)	65.22% (± 0.109)	67.08% (± 0.135)	66.14% (± 0.116)
P_3	<i>ATTN</i>	91.89% (± 0.052)	93.13% (± 0.067)	92.81% (± 0.058)	88.11% (± 0.078)	89.69% (± 0.095)	88.81% (± 0.072)	84.35% (± 0.088)	87.06% (± 0.064)	85.92% (± 0.083)	79.41% (± 0.093)	82.02% (± 0.069)	80.19% (± 0.101)
	<i>MEAN</i>	86.91% (± 0.088)	88.63% (± 0.092)	87.38% (± 0.084)	82.92% (± 0.091)	83.16% (± 0.079)	81.33% (± 0.101)	79.58% (± 0.099)	82.42% (± 0.073)	81.36% (± 0.082)	75.14% (± 0.095)	75.67% (± 0.101)	74.94% (± 0.097)

8.7.2. Demographic Risk Factors Effect

The longitudinal KOA grade predictions in Table 10 aim to justify the inclusion of demographic features. In this experiment, we examine two different cases of features used in the spatial *HGCN* convolutions. In the former case, we deploy solely the global shape descriptors obtained by the *C_Shape.Net* as the main feature representation, while in the latter one, we combine the shape features with demographic risk factors of gender, age, and BMI. Considering different views, these factors are incorporated in the convolutions via the respective hyperedge neighborhoods (Section 7.2), which allows the consideration of both shape and demographic relationships between the patients, simultaneously. The re-

sults show that the combination of shape and demographic factors provides a superior performance compared to the one attained by shape features alone. Higher values of accuracy can be noticed for the various historical lengths and future time steps, and across all quality measures. Moreover, the accuracy improvements tend to grow for more distant follow-up times. The above outcomes underscore the significance of demographic risk factors in acquiring more precise assessments for KOA grading and progression, an issue also corroborated by previous studies [32].

Table 10. Summary of KOA classification results of the proposed method, under varying-size (P) of historical windows. We examine the effect of using solely features learned by *ShapeNet* vs. incorporating additional demographic data of {Age, BMI, Gender}. **Metrics:** [Balanced Accuracy (BA), F1-weighted measure ($F1$), Cohen’s Kappa (K)].

P		t_{P+1}			t_{P+2}			t_{P+3}			t_{P+4}		
		BA	F1	K	BA	F1	K	BA	F1	K	BA	F1	K
0	<i>ShapeNet + Dem</i>	67.49% (± 0.089)	65.19% (± 0.075)	65.02% (± 0.093)	60.24% (± 0.105)	62.19% (± 0.088)	61.75% (± 0.164)	58.02% (± 0.079)	60.13% (± 0.112)	59.88% (± 0.069)	54.19% (± 0.122)	56.01% (± 0.096)	55.41% (± 0.080)
	<i>ShapeNet</i>	65.38% (± 0.112)	63.72% (± 0.108)	62.83% (± 0.137)	57.98% (± 0.165)	59.11% (± 0.142)	58.64% (± 0.119)	55.47% (± 0.137)	56.06% (± 0.121)	55.24% (± 0.106)	50.63% (± 0.113)	51.12% (± 0.136)	50.96% (± 0.120)
1	<i>ShapeNet + Dem</i>	75.84% (± 0.068)	77.01% (± 0.082)	76.73% (± 0.075)	72.16% (± 0.094)	74.28% (± 0.088)	73.64% (± 0.059)	68.07% (± 0.076)	69.54% (± 0.081)	67.91% (± 0.069)	62.08% (± 0.081)	64.10% (± 0.104)	63.59% (± 0.095)
	<i>ShapeNet</i>	73.71% (± 0.095)	74.98% (± 0.108)	74.19% (± 0.111)	69.91% (± 0.089)	71.57% (± 0.113)	70.45% (± 0.099)	65.19% (± 0.127)	66.02% (± 0.099)	65.76% (± 0.101)	58.36% (± 0.118)	59.17% (± 0.131)	59.03% (± 0.109)
2	<i>ShapeNet + Dem</i>	84.95% (± 0.048)	86.61% (± 0.066)	85.39% (± 0.081)	79.28% (± 0.057)	80.37% (± 0.062)	77.87% (± 0.074)	75.17% (± 0.088)	77.23% (± 0.053)	76.49% (± 0.071)	71.58% (± 0.099)	73.14% (± 0.071)	72.01% (± 0.057)
	<i>ShapeNet</i>	82.97% (± 0.089)	84.39% (± 0.101)	83.82% (± 0.092)	75.88% (± 0.078)	76.46% (± 0.097)	76.11% (± 0.112)	70.32% (± 0.088)	73.06% (± 0.090)	71.19% (± 0.102)	67.02% (± 0.076)	68.13% (± 0.089)	67.74% (± 0.081)
3	<i>ShapeNet + Dem</i>	91.89% (± 0.052)	93.13% (± 0.067)	92.81% (± 0.058)	88.11% (± 0.078)	89.69% (± 0.095)	88.81% (± 0.072)	84.35% (± 0.088)	87.06% (± 0.064)	85.92% (± 0.083)	79.41% (± 0.093)	82.02% (± 0.069)	80.19% (± 0.101)
	<i>ShapeNet</i>	88.34% (± 0.076)	90.17% (± 0.092)	89.61% (± 0.079)	85.87% (± 0.102)	87.14% (± 0.133)	86.42% (± 0.089)	80.38% (± 0.114)	82.19% (± 0.105)	81.33% (± 0.096)	74.85% (± 0.118)	76.17% (± 0.131)	75.49% (± 0.119)

8.7.3. Shape Network Contribution

This experiment aims to demonstrate the efficiency of the automatically acquired shape descriptors in representing the cartilage volume. For this purpose, we contrast the prediction performance using two different inputs to the predictor network. Concretely, in addition to the shape features provided by the *C_ShapeNet*, we also consider an alternative shape input obtained by a *CDI*-based approach [34]. Briefly, the method proceeds as follows. From the collection of MRI slices in the medial compartment, we select 20 tibiofemoral slices, across the medial-to-lateral axis. For each of the slices, we apply sparse sampling along the anterior-posterior direction, to collect 4, 8, and 12 locations at the $\mathcal{S}_0^{(FB)}$, $\mathcal{S}_1^{(FB)}$, and $\mathcal{S}_2^{(FB)}$ of the femoral bone surfaces, and a similar number of locations on the tibial bone surface, located at the $\mathcal{S}_0^{(TB)}$, $\mathcal{S}_1^{(TB)}$, and $\mathcal{S}_2^{(TB)}$, respectively. Then, for each of the femoral/tibial cartilage locations, we compute the respective femoral/tibial cartilage volumes, as a product of the thickness, the cartilage length (anterior-posterior), and the voxel size. In this way, we obtain 800 cartilage volumes overall for the medial compartment. A similar process is applied to compute another 800 cartilage volumes from the lateral compartment, yielding a total of 1600 volumes for the entire knee cartilage. The concatenation of these values forms the new feature vector. After *PCA* analysis, we are led to a reduced feature dimensionality by retaining the first 60 *PCA* components. Experiments show that the consideration of additional components does not improve performance. The large number of locations selected intends to ensure that all regions of articular surfaces are sampled where denudation frequently occurs.

Table 11 presents the longitudinal KOA predictions for the two inputs. As can be seen, the shape descriptors achieve considerably higher accuracies in terms of historical depths,

future times, and evaluation metrics, compared to the more simplistic stacking of sparsely located volumes. The improvements lie in the range of 3–8% and tend to increase for later follow-up times. The better performance can be explained as follows: (a) the graph-based representation of VNs offers a more detailed description of the structural properties of the entire cartilage volume. (b) The VNs are equipped with a set of local features, including geometric characteristics of the surfaces, cartilage thicknesses, and volumes. (c) As a result of deep HGCN convolutions and hierarchical aggregations conducted by the *C_ShapeNet*, the shape descriptors comprehensively encapsulate the VN information, acquiring a global representation of cartilage.

Table 11. Summary of KOA classification results under varying-size (P) of historical windows. We examine the effect of using the automatically learned volumetric features extracted by *C_ShapeNet* vs. computing the Cartilage Damage Index (CDI), at specified points in the MRI scans. Both cases incorporate additional demographic features of {Age, BMI, Gender}. **Metrics:** [Balanced Accuracy (BA), F1-weighted measure (F1), Cohen’s Kappa (K)].

P		t_{P+1}			t_{P+2}			t_{P+3}			t_{P+4}		
		BA	F1	K	BA	F1	K	BA	F1	K	BA	F1	K
0	<i>ShapeNet + Dem</i>	67.49% (± 0.089)	65.19% (± 0.075)	65.02% (± 0.093)	60.24% (± 0.105)	62.19% (± 0.088)	61.75% (± 0.164)	58.02% (± 0.079)	60.13% (± 0.112)	59.88% (± 0.069)	54.19% (± 0.122)	56.01% (± 0.096)	55.41% (± 0.080)
	<i>CDI + Dem</i>	64.28% (± 0.199)	62.27% (± 0.223)	62.85% (± 0.176)	55.73% (± 0.208)	56.08% (± 0.264)	55.47% (± 0.188)	53.19% (± 0.217)	53.94% (± 0.232)	52.71% (± 0.186)	45.33% (± 0.235)	46.12% (± 0.250)	46.09% (± 0.193)
1	<i>ShapeNet + Dem</i>	75.84% (± 0.068)	77.01% (± 0.082)	76.73% (± 0.075)	72.16% (± 0.094)	74.28% (± 0.088)	73.64% (± 0.059)	68.07% (± 0.076)	69.54% (± 0.081)	67.91% (± 0.069)	62.08% (± 0.081)	64.10% (± 0.104)	63.59% (± 0.095)
	<i>CDI + Dem</i>	73.47% (± 0.151)	74.02% (± 0.186)	73.38% (± 0.154)	67.92% (± 0.192)	69.14% (± 0.218)	68.53% (± 0.165)	60.64% (± 0.202)	61.79% (± 0.229)	61.27% (± 0.171)	54.18% (± 0.194)	55.07% (± 0.231)	54.81% (± 0.187)
2	<i>ShapeNet + Dem</i>	84.95% (± 0.048)	86.61% (± 0.066)	85.39% (± 0.081)	79.28% (± 0.057)	80.37% (± 0.062)	77.87% (± 0.074)	75.17% (± 0.088)	77.23% (± 0.053)	76.49% (± 0.071)	71.58% (± 0.099)	73.14% (± 0.071)	72.01% (± 0.057)
	<i>CDI + Dem</i>	81.37% (± 0.139)	82.15% (± 0.166)	82.03% (± 0.153)	74.21% (± 0.178)	75.22% (± 0.199)	73.96% (± 0.133)	70.18% (± 0.184)	71.42% (± 0.205)	71.29% (± 0.149)	64.55% (± 0.180)	66.37% (± 0.193)	64.93% (± 0.142)
3	<i>ShapeNet + Dem</i>	91.89% (± 0.052)	93.13% (± 0.067)	92.81% (± 0.058)	88.11% (± 0.078)	89.69% (± 0.095)	88.81% (± 0.072)	84.35% (± 0.088)	87.06% (± 0.064)	85.92% (± 0.083)	79.41% (± 0.093)	82.02% (± 0.069)	80.19% (± 0.101)
	<i>CDI + Dem</i>	88.36% (± 0.098)	89.34% (± 0.104)	88.92% (± 0.117)	84.41% (± 0.091)	85.03% (± 0.133)	84.67% (± 0.111)	80.29% (± 0.128)	82.53% (± 0.106)	81.16% (± 0.133)	73.86% (± 0.147)	74.45% (± 0.116)	74.07% (± 0.154)

8.8. Visual Demonstration

In this brief section, we showcase an example of KL-grade progression for a specific knee, across all the time points in the historical and future stages. Figure 9 illustrates the cartilage degradation that is indicative of KOA incidence and progression.

Table 12 illustrates the performance of our proposed approach with regards to the size of the historical record considered, for the progression of the specific knee depicted in Figure 9. As can be seen, expanding the historic depth has a definite positive effect on the predictions of the model, across all the future time points:

1. Starting from the trivial case of using only the baseline data to perform future estimations, we can observe that *ST_HGCN* struggles to produce accurate predictions, with the model’s performance progressively deteriorating. While the immediate step-ahead prediction is correct, all multi-step ahead predictions are erroneous, with the furthest one also being the largest in discrepancy.
2. The second and third rows of Table 12 showcase the model’s performance for a historic size of $P = 2$ and $P = 3$, respectively. Progressing from a single past observation to two, *ST_HGCN* is able to slightly boost its performance. Supplying the additional data points to the historical sequence further enhances the model’s predictive capabilities. In particular, *ST_HGCN* 1) performs more accurate predictions with increasing P and 2) the magnitude of the errors is decreased (2/4 in the $P = 0$ case vs. 3/4 in the $P = 2, 3$ cases.)

3. Finally, for the last step, we can confirm that expanding the historical depth to the maximum size (4), enables *ST_HGCN* to accurately predict all future *KL* grades.

Table 12. Table showcasing future *KL* predictions under varying historical depths, against the ground truth values (Predicted Value/Ground Truth).

	+0	+1	+2	+3	+4	+5	+6	+7
$P = 0$	✗	2/2	2/3	2/3	2/4	✗	✗	✗
$P = 1$	✗	✗	2/2	2/3	3/3	3/4	✗	✗
$P = 2$	✗	✗	✗	2/2	3/3	3/3	3/4	✗
$P = 3$	✗	✗	✗	✗	2/2	3/3	3/3	4/4

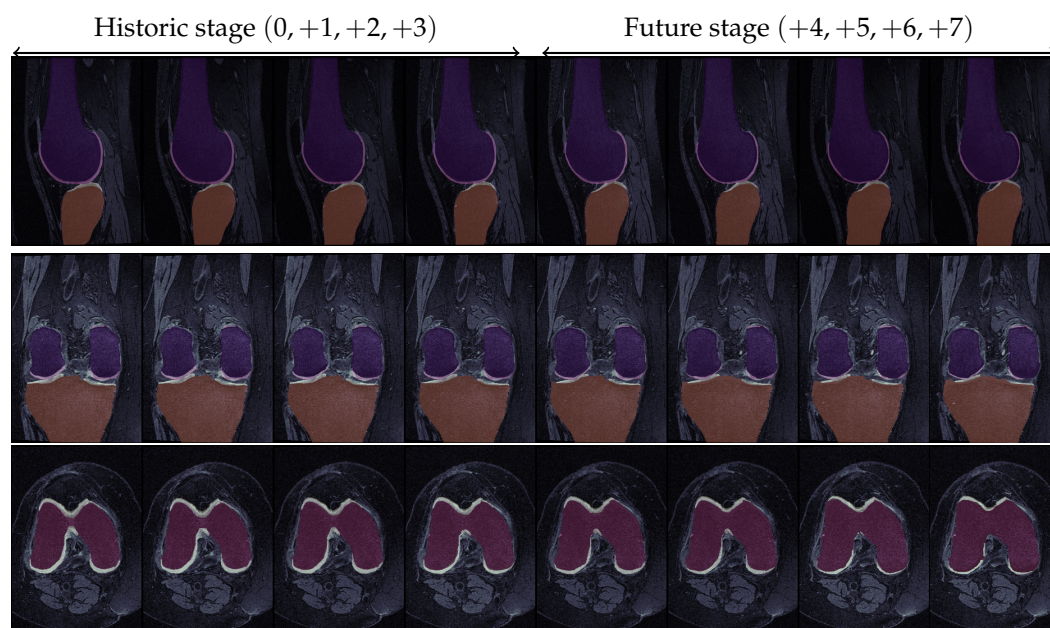


Figure 9. KOA progression for a specific knee in the *OAI* dataset. The first four images in each row correspond to the historic stage, while the latter four refer to the prediction stage. For the first two rows, dark orange and dark purple represent the tibial and femoral bones, respectively, while light yellow and light pink represent the corresponding cartilage structures. In the last row, the red part constitutes the femoral bone while the light yellow represents the femoral cartilage. **Ground Truth** *KL*-grade progression: 0 → 1 → 1 → 2 → 2 → 3 → 3 → 4 / **Predicted** *KL*-grade progression: 0 → 0 → 1 → 2 → 2 → 2 → 3 → 4 (green indicates correct prediction, red indicates erroneous prediction).

8.9. Comparative Results

In this section, we present comparative results in Table 13, considering a collection of existing representative works in the field. The different methods are categorized into three groups, namely, the *KL* grading, the longitudinal *KL* grading, and the *KOA* progression. The methods in the *KL* grading group conduct static node (patient) classifications into the five *KL* classes. The second group contains a recent method, which, to the best of our knowledge, matches the multi-step ahead *KL* prediction setting of our scheme. Finally, the third group contains the methods which provide predictions of *KOA* progression. Usually, they consider the baseline visit as the reference time and generate predictions at a single follow-up time. For each category, we report the best results obtained using the max depth of historic sequences ($P = 3$).

Table 13. Comparative performance of KOA prediction methods dealing w.r.t. (1) static KL-grade classification and (2) longitudinal KL-grade prediction and (3) KL-grade progression.

Method	Prediction Type	Data	Source	Add. Inputs	Model	Results
Kishore et.al [29]	KL Grading	OAI	X-ray	-	12 DL models	ACC = 98.36%
Yong et.al [27]	KL Grading	OAI	X-ray	-	{VGG, ResNet, DenseNet}	ACC _{macro} = 88.09%
Gorriz et.al [52]	KL Grading	OAI, MOST	X-ray	-	VGG + Attn. Module	ACC = 64.3%
Chen et.al [28]	KL Grading	OAI-baseline	X-ray	-	{ResNet, VGG, DenseNet} + Novel Ord. loss	ACC = 69.7%
Zhang et.al [25]	KL Grading	OAI-baseline	X-ray	-	ResNet+ CBAM	ACC = 74.81%
Tiulpin et.al [24]	KL Grading	OAI _{tm.} → MOST _{lst.}	X-ray	-	Deep Siamese CNN	K = 82%
Ashamari et.al [33]	KL Grading	Kaggle	X-ray	-	{Seq.CNN, VGG, ResNet-50}	ACC = 92.17%
Proposed	KL Grading	OAI	MRI	-	C_Shape.Net + SptHGNC	BA = 95.98%, K = 97.92%
Hu et.al [36]	Long. KL Prediction	OAI	X-ray	-	Adversarial NN	Sens = 0.639 (+1), 0.632 (+2), 0.618 (+3), 0.602 (+4)
Proposed	Long. KL Prediction	OAI	MRI	-	C_Shape.Net + SptHGNC + Transf.	BA = 95.98%, K = 97.92%
Panfilov et.al. [37]	Long. Progr. Prediction	OAI	X-ray, MRI (multi-modal)	+Demographics	CNN + Transf.	AUC = 0.76 (+1), 0.72 (+2), 0.70 (+3), 0.74 (+4)
Halilaj et.al. [35]	Long. Progr. Prediction (Progr. vs No-Progr.)	OAI	X-ray	Pain scores (WOMAC)	LASSO + Clustering	AUC = 0.86
Guan et.al. [32]	JSL Progr. (baseline → +48)	OAI	MRI	+{Demographics, Injury hist., Tibiofemoral angle, etc.}	Deep Learning	AUC = 0.85
Du et.al. [34]	KL Progr. Prediction (baseline → +2)	OAI	MRI	CDI at 36 locations	Comparison of ML methods	AUC = 0.76, F = 0.714
Ahmad et.al. [33]	Long. Progr. Prediction	OAI	X-ray, MRI (multi-modal)	+Demographics	CNN + Transf.	AUC = 0.76 (+1), 0.72 (+2), 0.70 (+3), 0.74 (+4)
Tiulpin et.al. [51]	No-Progr. vs Progr. (0 → +5)	OAI _{tm.} → MOST _{lst.}	X-ray	+{Demographics, WOMAC, Injury hist., KL}	DenseNet	AUC = 0.834, Sens. = 0.78, Spec. = 0.77
Pedioia et.al. [53]	No-OA(≤ 1) vs OA(> 1))	OAI	MRI	{Demographics, WOMAC}	DenseNet	AUC = 0.83, Sens. = 0.77, Spec. = 0.78
Alexopoulos et.al. [54]	No_OA vs EARLY_OA	OAI	MRI	+Demographics	{ResNet, DenseNet, CVAE}	AUC = 0.65
Schiratti et.al. [55]	JSN Progr. (+1) ahead	OAI	MRI	-	EfficientNet-B0, DenseNet, CVAE	AUC = 0.67
Proposed	KL Progression Incidence	OAI	MRI	+Demographics	C_Shape.Net + ST_HGNC + Transf.	AUC = [0.94 (+1), 0.89 (+2), 0.85 (+3), 0.80 (+4)], OA = 85.94%

In the *KL* grading, our predictor achieved high accuracy results comparable to the best ones reported in [29]. For the second category, we obtained considerably better longitudinal *KL* predictions with max historic depth, compared to the *A-EEN* approach in [36]. Superior results are also produced for smaller historic horizons (Table 13). Finally, regarding the longitudinal progression incidence of the third category, we also obtained good results, favorably compared to existing methods in the group. It should be noticed that straightforward comparisons in this category are not feasible, because the various methods define the *KOA* progression differently. For instance, direct contrast with our setting is valid for the works [37], where *KOA* progression is defined as the change in *KL* grade over a follow-up period. However, other approaches define progression in terms of joint space narrowing [35,55]. Concluding, the above outcomes signify that the usage of previous historical data is indispensable for acquiring more precise *KL* and progression predictions ahead.

9. Conclusions

An integrated approach is presented in this paper for longitudinal *KL* predictions and progression. To accomplish the above goal, we propose two novel and distinct networks, namely, the *C_Shape.Net* and the predictor network, operating in cascade. *C_Shape.Net* acts upon a graph of volumetric nodes, especially designed to represent the surface and volumetric characteristics of the knee articular cartilage. It conducts *HGCN* convolutions, graph pooling, and readout operations across a hierarchy of layers, providing comprehensive 3D shape descriptors of the cartilage volume. Contrary to most previous works in *KOA*, grading the predictor network exhibits spatio-temporal capabilities, following the sequence-to-sequence learning approach. Concretely, the model encompasses spatial *HGCN* convolutions, attention-based temporal fusion of feature representations at multiple layers, and a transformer module that generates the longitudinal predictions at follow-up visits.

Complying with the sequence-to-sequence scenario, the *OAI* data are distinguished in the experiments into a historical stage of varying depth and a prediction stage of future times ahead. The predictor is then invoked to process the patient's historical sequences and transform it into a prediction sequence of *KOA* grades and progression. We have conducted a thorough investigation to evaluate the performance of our method in different cases, including the node classification, longitudinal *KOA* predictions, *KOA* assessments in terms of demographic features, and longitudinal predictions of *KOA* progression. In these cases, our method acquired adequate accuracies, favorably compared to existing approaches.

The fundamental finding of this work is that the longer the patient's historic sequences, the more accurate prediction sequences are obtained, in both *KL* grading and progression. Accordingly, provided that deeper data sequences are collected in the future, we can explore the opportunity of developing personalized prediction schemes via transfer learning, whereby effective longitudinal *KOA* assessments can be recursively drawn by exploiting temporal trends and knowledge embedded in previous historic visits. Along this direction, in our future work, we intend to further enhance the spatio-temporal capabilities of our scheme by resorting to more effective temporal processing approaches, such as the Gated Recurrent Unit (*GRU*) under the encoder-decoder paradigm. Additionally, drawing inspiration from recent work proposed in the field of knee segmentation [56], another possible route of experimentation is available, involving the joint treatment of the knee cartilage shape description and the *KOA* progression, formulated as a multi-task learning problem.

Author Contributions: Conceptualization, J.B.T.; data curation, C.G.C.; formal analysis, J.B.T.; methodology, J.B.T. and C.G.C.; project administration, J.B.T.; software, C.G.C.; supervision, J.B.T.; validation, J.B.T., C.G.C., and A.L.S.; visualization, J.B.T. and C.G.C.; writing—original draft, J.B.T.;

writing—review and editing, J.B.T. and C.G.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The original data used in this study are available in the Osteoarthritis Initiative repository at <https://nda.nih.gov/oai> (accessed on 1 March 2018).

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A

Let us consider two surfaces $\mathcal{S}_1$ and $\mathcal{S}_2$, and a mesh vertex $P \in \mathcal{S}_1$. The corresponding point $Q = CR(P) \in \mathcal{S}_2$ is described as shown in Figure 3b. First, we draw the lines $\vec{l}_1 = PR \perp \mathcal{S}_2$ and $\vec{l}_2 = PT \perp \mathcal{S}_1$, perpendicular to the surfaces $\mathcal{S}_2$ and $\mathcal{S}_1$, respectively. Then, the intersection of the line $\vec{l} = (\vec{l}_1 + \vec{l}_2)/2$ with the surface $\mathcal{S}_2$ defines the location of point $Q \in \mathcal{S}_2$. Further, the length of the segment PQ defines the cartilage thickness between these two CR vertices, $Th = \|PQ\|$.

References

1. Wang, Z.; Xiao, Z.; Sun, C.; Xu, G.; He, J. Global, regional and national burden of osteoarthritis in 1990–2021: A systematic analysis of the global burden of disease study 2021. *BMC Musculoskelet. Disord.* **2024**, *25*, 1021. [CrossRef] [PubMed]
2. Bastick, A.N.; Runhaar, J.; Belo, J.N.; Bierma-Zeinstra, S.M. Prognostic factors for progression of clinical osteoarthritis of the knee: A systematic review of observational studies. *Arthritis Res. Ther.* **2015**, *17*, 152. [CrossRef] [PubMed]
3. Kellgren, J.H.; Lawrence, J.S. Radiological Assessment of Osteo-Arthrosis. *Ann. Rheum. Dis.* **1957**, *16*, 494–502. [CrossRef]
4. Zhao, H.; Ou, L.; Zhang, Z.; Zhang, L.; Liu, K.; Kuang, J. The value of deep learning-based X-ray techniques in detecting and classifying K-L grades of knee osteoarthritis: A systematic review and meta-analysis. *Eur. Radiol.* **2024**, *35*, 327–340. [CrossRef] [PubMed]
5. Joo, P.; Borjali, A.; Chen, A.; Muratoglu, O.; Varadarajan, K. Defining and predicting radiographic knee osteoarthritis progression: A systematic review of findings from the osteoarthritis initiative. *Knee Surgery Sport. Traumatol. Arthrosc.* **2022**, *30*, 4015–4028. [CrossRef]
6. Lee, L.S.; Chan, P.K.; Wen, C.; Fung, W.C.; Cheung, A.; Chan, V.W.K.; Cheung, M.H.; Fu, H.; Yan, C.H.; Chiu, K.Y. Artificial intelligence in diagnosis of knee osteoarthritis and prediction of arthroplasty outcomes: A review. *Arthroplasty* **2022**, *4*, 16. [CrossRef]
7. Chadoulos, C.; Tsaopoulos, D.; Symeonidis, A.; Moustakidis, S.; Theocharis, J. Dense Multi-Scale Graph Convolutional Network for Knee Joint Cartilage Segmentation. *Bioengineering* **2024**, *11*, 278. [CrossRef]
8. Peterfy, C.; Schneider, E.; Nevitt, M. The osteoarthritis initiative: Report on the design rationale for the magnetic resonance imaging protocol for the knee. *Osteoarthr. Cartil.* **2008**, *16*, 1433–1441. [CrossRef]
9. Bai, J.; Gong, B.; Zhao, Y.; Lei, F.; Yan, C.; Gao, Y. Multi-Scale Representation Learning on Hypergraph for 3D Shape Retrieval and Recognition. *IEEE Trans. Image Process.* **2021**, *30*, 5327–5338. [CrossRef]
10. Su, H.; Maji, S.; Kalogerakis, E.; Learned-Miller, E. Multi-view Convolutional Neural Networks for 3D Shape Recognition. *arXiv* **2015**, arXiv:1505.00880. [CrossRef]
11. Feng, Y.; Zhang, Z.; Zhao, X.; Ji, R.; Gao, Y. GVCNN: Group-View Convolutional Neural Networks for 3D Shape Recognition. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 264–272. [CrossRef]
12. He, X.; Huang, T.; Bai, S.; Bai, X. View N-Gram Network for 3D Object Retrieval. In Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Republic of Korea, 27 October–2 November 2019; pp. 7514–7523. [CrossRef]
13. Yu, T.; Meng, J.; Yuan, J. Multi-view Harmonized Bilinear Network for 3D Object Recognition. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 186–194. [CrossRef]
14. Charles, R.Q.; Su, H.; Kaichun, M.; Guibas, L.J. PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 77–85. [CrossRef]

15. Qi, C.R.; Yi, L.; Su, H.; Guibas, L.J. PointNet++: Deep hierarchical feature learning on point sets in a metric space. In Proceedings of the 31st International Conference on Neural Information Processing Systems, Red Hook, NY, USA, 4–9 December 2017; NIPS'17, pp. 5105–5114.
16. Zhao, H.; Jiang, L.; Fu, C.W.; Jia, J. PointWeb: Enhancing Local Neighborhood Features for Point Cloud Processing. In Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 15–20 June 2019; pp. 5560–5568. [\[CrossRef\]](#)
17. Wu, Z.; Song, S.; Khosla, A.; Yu, F.; Zhang, L.; Tang, X.; Xiao, J. 3D ShapeNets: A deep representation for volumetric shapes. In Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015; pp. 1912–1920. [\[CrossRef\]](#)
18. Maturana, D.; Scherer, S. VoxNet: A 3D Convolutional Neural Network for real-time object recognition. In Proceedings of the 2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Hamburg, Germany, 28 September–3 October 2015; pp. 922–928. [\[CrossRef\]](#)
19. Qi, C.R.; Su, H.; NieBner, M.; Dai, A.; Yan, M.; Guibas, L.J. Volumetric and Multi-view CNNs for Object Classification on 3D Data. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 5648–5656. [\[CrossRef\]](#)
20. Feng, Y.; Feng, Y.; You, H.; Zhao, X.; Gao, Y. MeshNet: Mesh Neural Network for 3D Shape Representation. *AAAI* **2019**, *33*, 8279–8286. [\[CrossRef\]](#)
21. Hanocka, R.; Hertz, A.; Fish, N.; Giryas, R.; Fleishman, S.; Cohen-Or, D. MeshCNN: A Network with an Edge. *ACM Trans. Graph.* **2019**, *38*, 1–12. [\[CrossRef\]](#)
22. Wei, X.; Yu, R.; Sun, J. Learning View-Based Graph Convolutional Network for Multi-View 3D Shape Analysis. *IEEE Trans. Pattern Anal. Mach. Intell.* **2023**, *45*, 7525–7541. [\[CrossRef\]](#)
23. Tang, W.; Qiu, G. Dense graph convolutional neural networks on 3D meshes for 3D object segmentation and classification. *Image Vis. Comput.* **2021**, *114*, 104265. [\[CrossRef\]](#)
24. Tiulpin, A.; Thevenot, J.; Rahtu, E.; Lehenkari, P.; Saarakkala, S. Automatic Knee Osteoarthritis Diagnosis from Plain Radiographs: A Deep Learning-Based Approach. *Sci. Rep.* **2018**, *8*, 1727. [\[CrossRef\]](#)
25. Zhang, B.; Tan, J.; Cho, K.; Chang, G.; Deniz, C.M. Attention-based CNN for KL Grade Classification: Data from the Osteoarthritis Initiative. In Proceedings of the 2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI), Iowa City, IA, USA, 3–7 April 2020; pp. 731–735. [\[CrossRef\]](#)
26. Woo, S.; Park, J.; Lee, J.; Kweon, I.S. CBAM: Convolutional Block Attention Module. *arXiv* **2018**, arXiv:1807.06521.
27. Yong, C.W.; Teo, K.; Murphy, B.P.; Hum, Y.C.; Tee, Y.K.; Xia, K.; Lai, K.W. Knee osteoarthritis severity classification with ordinal regression module. *Multimed. Tools Appl.* **2022**, *81*, 41497–41509. [\[CrossRef\]](#)
28. Chen, P.; Gao, L.; Shi, X.; Allen, K.; Yang, L. Fully automatic knee osteoarthritis severity grading using deep neural networks with a novel ordinal loss. *Comput. Med. Imaging Graph.* **2019**, *75*, 84–92. [\[CrossRef\]](#)
29. V, V.K.; Kalpana, V.; Kumar, G.H. Evaluating the efficacy of deep learning models for knee osteoarthritis prediction based on Kellgren-Lawrence grading system. *E-Prime Adv. Electr. Eng. Electron. Energy* **2023**, *5*, 100266. [\[CrossRef\]](#)
30. Leung, K.; Zhang, B.; Tan, J.; Shen, Y.; Geras, K.J.; Babb, J.S.; Cho, K.; Chang, G.; Deniz, C.M. Prediction of Total Knee Replacement and Diagnosis of Osteoarthritis by Using Deep Learning on Knee Radiographs: Data from the Osteoarthritis Initiative. *Radiology* **2020**, *296*, 584–593. [\[CrossRef\]](#)
31. Alexopoulos, A.; Hirvasniemi, J.; Klein, S.; Donkervoort, C.; Oei, E.; Tümer, N. Early detection of knee osteoarthritis using deep learning on knee MRI. *Osteoarthr. Imaging* **2023**, *3*, 100112. [\[CrossRef\]](#)
32. Guan, B.; Liu, F.; Haj-Mirzaian, A.; Demehri, S.; Samsonov, A.; Neogi, T.; Guermazi, A.; Kijowski, R. Deep learning risk assessment models for predicting progression of radiographic medial joint space loss over a 48-MONTH follow-up period. *Osteoarthr. Cartil.* **2020**, *28*, 428–437. [\[CrossRef\]](#)
33. Almhdi-Imjabbar, A.; Nguyen, K.L.; Toumi, H.; Jennane, R.; Lespessailles, E. Prediction of knee osteoarthritis progression using radiological descriptors obtained from bone texture analysis and Siamese neural networks: Data from OAI and MOST cohorts. *Arthritis Res. Ther.* **2022**, *24*, 66. [\[CrossRef\]](#)
34. Du, Y.; Almajalid, R.; Shan, J.; Zhang, M. A Novel Method to Predict Knee Osteoarthritis Progression on MRI Using Machine Learning Methods. *IEEE Trans. Nanobiosci.* **2018**, *17*, 228–236. [\[CrossRef\]](#)
35. Halilaj, E.; Le, Y.; Hicks, J.; Hastie, T.; Delp, S. Modeling and predicting osteoarthritis progression: Data from the osteoarthritis initiative. *Osteoarthr. Cartil.* **2018**, *26*, 1643–1650. [\[CrossRef\]](#)
36. Hu, K.; Wu, W.; Li, W.; Simic, M.; Zomaya, A.; Wang, Z. Adversarial Evolving Neural Network for Longitudinal Knee Osteoarthritis Prediction. *IEEE Trans. Med. Imaging* **2022**, *41*, 3207–3217. [\[CrossRef\]](#) [\[PubMed\]](#)
37. Panfilov, E.; Saarakkala, S.; Nieminen, M.T.; Tiulpin, A. End-To-End Prediction of Knee Osteoarthritis Progression with Multi-Modal Transformers. *arXiv* **2023**, arXiv:2307.00873. [\[CrossRef\]](#)

38. Yu, B.; Yin, H.; Zhu, Z. Spatio-Temporal Graph Convolutional Networks: A Deep Learning Framework for Traffic Forecasting. In Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, Stockholm, Sweden, 13–19 July 2018; pp. 3634–3640. [\[CrossRef\]](#)
39. Guo, S.; Lin, Y.; Feng, N.; Song, C.; Wan, H. Attention Based Spatial-Temporal Graph Convolutional Networks for Traffic Flow Forecasting. *AAAI* **2019**, *33*, 922–929. [\[CrossRef\]](#)
40. Zhao, L.; Song, Y.; Zhang, C.; Liu, Y.; Wang, P.; Lin, T.; Deng, M.; Li, H. T-GCN: A Temporal Graph Convolutional Network for Traffic Prediction. *IEEE Trans. Intell. Transport. Syst.* **2020**, *21*, 3848–3858. [\[CrossRef\]](#)
41. Defferrard, M.; Bresson, X.; Vandergheynst, P. Convolutional Neural Networks on Graphs with Fast Localized Spectral Filtering. *arXiv* **2016**, arXiv:1606.09375.
42. Kipf, T.N.; Welling, M. Semi-Supervised Classification with Graph Convolutional Networks. *arXiv* **2017**, arXiv:1609.02907. [\[CrossRef\]](#)
43. Ma, Z.; Jiang, Z.; Zhang, H. Hyperspectral Image Classification Using Feature Fusion Hypergraph Convolution Neural Network. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5517314. [\[CrossRef\]](#)
44. Bai, S.; Zhang, F.; Torr, P.H. Hypergraph convolution and hypergraph attention. *Pattern Recognit.* **2021**, *110*, 107637. [\[CrossRef\]](#)
45. Ambellan, F.; Tack, A.; Ehlke, M.; Zachow, S. Automated segmentation of knee bone and cartilage combining statistical shape knowledge and convolutional neural networks: Data from the Osteoarthritis Initiative. *Med. Image Anal.* **2019**, *52*, 109–118. [\[CrossRef\]](#) [\[PubMed\]](#)
46. Elshakhs, Y.; Deliparaschos, K.; Charalambous, T.; Oliva, G.; Zolotas, A. A Comprehensive Survey on Delaunay Triangulation: Applications, Algorithms, and Implementations Over CPUs, GPUs, and FPGAs. *IEEE Access* **2024**, *12*, 12562–12585. [\[CrossRef\]](#)
47. Lee, J.; Lee, I.; Kang, J. Self-Attention Graph Pooling. *arXiv* **2019**, arXiv:1904.08082. [\[CrossRef\]](#)
48. Song, C.; Lin, Y.; Guo, S.; Wan, H. Spatial-Temporal Synchronous Graph Convolutional Networks: A New Framework for Spatial-Temporal Network Data Forecasting. *AAAI* **2020**, *34*, 914–921. [\[CrossRef\]](#)
49. Friedman, M. The Use of Ranks to Avoid the Assumption of Normality Implicit in the Analysis of Variance. *J. Am. Stat. Assoc.* **1937**, *32*, 675–701. [\[CrossRef\]](#)
50. Nemenyi, P. Distribution-Free Multiple Comparisons. Ph.D. Thesis, Princeton University, Princeton, NJ, USA, 1963.
51. Tiulpin, A.; Saarakkala, S. Automatic Grading of Individual Knee Osteoarthritis Features in Plain Radiographs Using Deep Convolutional Neural Networks. *Diagnostics* **2020**, *10*, 932. [\[CrossRef\]](#)
52. Gorriz, M.; Antony, J.; Antony, J.; Ie, D.; McGuinness, K.; McGuinness, K.; Ie, D.; Giro-i Nieto, X.; O'Connor, N.E.; O'Connor, N.; et al. Assessing Knee OA Severity with CNN attention-based end-to-end architectures. *Proc. Mach. Learn. Res.* **2019**, *102*, 197–204.
53. Pedoia, V.; Lee, J.; Norman, B.; Link, T.; Majumdar, S. Diagnosing osteoarthritis from T2 maps using deep learning: An analysis of the entire Osteoarthritis Initiative baseline cohort. *Osteoarthr. Cartil.* **2019**, *27*, 1002–1010. [\[CrossRef\]](#) [\[PubMed\]](#)
54. Alexopoulos, A.; Hirvasniemi, J.; Tumer, N. Early detection of knee osteoarthritis using deep learning on knee magnetic resonance images. *arXiv* **2022**, arXiv:2209.01192.
55. Schiratti, J.B.; Dubois, R.; Herent, P.; Cahané, D.; Dachary, J.; Clozel, T.; Wainrib, G.; Keime-Guibert, F.; Lalande, A.; Pueyo, M.; et al. A deep learning method for predicting knee osteoarthritis radiographic progression from MRI. *Arthritis Res. Ther.* **2021**, *23*, 262. [\[CrossRef\]](#) [\[PubMed\]](#)
56. Li, X.; Lv, S.; Li, M.; Zhang, J.; Jiang, Y.; Qin, Y.; Luo, H.; Yin, S. SDMT: Spatial Dependence Multi-Task Transformer Network for 3D Knee MRI Segmentation and Landmark Localization. *IEEE Trans. Med Imaging* **2023**, *42*, 2274–2285. [\[CrossRef\]](#)

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.