Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Craniofacial Growth Transformer: A Growth-Aware Framework for Longitudinal Craniofacial Analysis and Forecasting

Mostafa A. Salama

mostafa.salama@bue.edu.eg

BUE: The British University in Egypt

Article

Keywords: Craniofacial growth modeling, Longitudinal facial analysis, LSTM, Transformer, Growth forecasting, Facial landmarks, Deep learning

Posted Date: November 16th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11010034/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Craniofacial Growth Transformer: A Growth-Aware Framework for Longitudinal Craniofacial Analysis and Forecasting

Mostafa A. Salama[1*]

[1*]Faculty of Infromatics and Computer Science, British University in Egypt, Sherouk, Cairo, 11111, Egypt.

Corresponding author(s). E-mail(s): mostafa.salama@bue.edu.eg;

Abstract

Accurate characterization of craniofacial growth is important for orthodontic planning, craniofacial assessment, and long-term monitoring of facial development. Most existing methods rely on sparse or cross-sectional data. As a result, they do not fully capture how facial structures change over time or how different facial regions develop together.

This study presents a growth-aware framework for longitudinal craniofacial analysis and forecasting. The framework is based on a dense longitudinal dataset containing nearly ten years of near-daily facial observations from three individuals. Using this dataset, we analyze temporal dependencies, age-related differences, developmental consistency across individuals, and relationships among craniofacial measurements.

The results show that several craniofacial measurements follow similar developmental trajectories across individuals. This finding suggests the presence of shared growth patterns despite individual differences. The analysis also identifies stable relationships between anatomically related regions, including bilateral ocular structures and mandibular-facial dimensions. These findings indicate that craniofacial growth follows a coordinated process rather than independent changes in individual features.

Based on these findings, we propose a Long Short-Term Memory–Temporal–Spatial Transformer (LSTM-TS-TF) model. The model captures temporal continuity, long-range developmental dependencies, and spatial interactions between facial measurements. Experimental results show that the proposed model improves forecasting accuracy and reduces prediction error by approximately 63–72% compared with conventional LSTM models.

These findings show that dense longitudinal modeling can identify interpretable growth descriptors and coordinated developmental patterns. The results also support the use of growth-aware deep learning models for data-driven craniofacial analysis and predictive modeling in biomedical applications.

Keywords: Craniofacial growth modeling, Longitudinal facial analysis, LSTM, Transformer, Growth forecasting, Facial landmarks, Deep learning

1 Introduction

Craniofacial growth is a complex biological process that shapes the human face during childhood, adolescence, and early adulthood. Understanding facial growth is important for orthodontic treatment planning, craniofacial surgery, forensic science, anthropometry, age progression, and biometric analysis. Knowledge of facial development helps clinicians and researchers anticipate future changes, improve diagnosis, and support personalized treatment decisions. As a result, facial growth modeling has received attention in biomedical research and computer vision, particularly in facial aging, age estimation [1], and clinical decision-support systems.

Traditional studies of craniofacial growth have relied on cephalometric measurements, statistical growth curves, geometric analysis, and handcrafted descriptors. These methods have improved our understanding of facial development. However, they often assume simplified growth patterns. They also have limited ability to capture the nonlinear and long-term relationships involved in craniofacial growth. In addition, most studies examine individual facial measurements rather than the joint development of multiple facial structures.

Recent advances in machine learning and deep learning have improved facial analysis and growth prediction [2, 3]. Deep neural networks, recurrent models, convolutional networks, attention mechanisms, and transformers have achieved strong results in age estimation, age progression, cephalometric analysis, landmark detection, and craniofacial prediction. Recent reviews also show increasing use of artificial intelligence in orthodontics and clinical decision support, where deep learning supports craniofacial assessment, treatment planning, and growth analysis [4, 5]. These methods can learn complex patterns from facial data and often outperform traditional statistical approaches in prediction tasks. Recent work has also improved automatic facial landmark detection and craniofacial measurement extraction, providing reliable geometric representations for growth analysis [6, 7]. Despite these advances, important limitations remain in both data and methodology.

First, most craniofacial growth studies use cross-sectional data or sparsely sampled longitudinal observations. Such data provide limited information about continuous facial development. Dense longitudinal datasets that contain facial observations across many years are still rare. Second, many studies work directly with images or cephalometric radiographs without examining the developmental behavior of individual craniofacial descriptors. As a result, the growth characteristics of specific facial

measurements are often overlooked. Third, most forecasting methods focus on prediction performance while giving less attention to developmental consistency across individuals or growth relationships between facial regions. Finally, transformer-based models have shown strong performance in facial analysis, but they are mostly used for age estimation, facial synthesis, and landmark localization rather than longitudinal craniofacial growth forecasting.

To address these limitations, this study presents a framework for longitudinal craniofacial growth analysis and prediction. The framework is built on a dense longitudinal facial-growth dataset containing thousands of temporally ordered observations collected over approximately ten years from three individuals. The dataset captures the transition from adolescence to early adulthood. To the best of our knowledge, datasets with this level of temporal coverage remain rare. Instead of relying only on image representations, facial development is described using geometric craniofacial descriptors extracted from facial landmarks. This allows detailed analysis of structural growth patterns.

A key contribution of this study is a growth-oriented analysis framework for longitudinal facial development. The framework examines facial growth from several perspectives, including temporal dependency, age discriminability, neural predictive importance, developmental consistency across individuals, and coordinated growth relationships among facial structures. These analyses help identify descriptors that capture meaningful growth processes and reveal developmental patterns shared across individuals.

The findings from these analyses are then used to develop a new forecasting model called the **LSTM–Temporal–Spatial Transformer (LSTM-TS-TF)**. The model combines recurrent temporal modeling, temporal self-attention, and spatial self-attention in a single framework. The LSTM component captures short-term developmental continuity. The Temporal Transformer learns long-range growth dependencies across years. The Spatial Transformer learns relationships between anatomically related facial descriptors. This design allows the model to use both temporal and structural information during forecasting.

The main contributions of this study are as follows:

- Construction and analysis of a dense longitudinal facial-growth dataset covering approximately ten years of craniofacial development and containing thousands of temporally ordered observations from adolescence to early adulthood.
- Development of a growth-oriented analysis framework that combines temporal dependency analysis, age discriminability evaluation, neural saliency assessment, growth-pattern characterization, and coordinated growth analysis.
- Identification of facial descriptors that show developmental relevance, cross-individual consistency, and coordinated growth behavior, helping reveal shared craniofacial growth patterns.
- Development of a novel LSTM–Temporal–Spatial Transformer architecture that jointly models temporal continuity, long-range developmental dependencies, and coordinated anatomical interactions.
- Comprehensive experimental evaluation of the proposed framework for longitudinal facial-growth forecasting using both complete and reduced feature sets.

A preliminary version of this work was presented in [8]. That study introduced an ensemble feature-ranking framework for identifying growth-sensitive craniofacial descriptors. The present work extends that research in several ways. It introduces a dense longitudinal facial-growth dataset. It adds growth-pattern characterization and coordinated growth analyses. It also develops a new LSTM–Temporal–Spatial Transformer model for longitudinal growth forecasting. As a result, this study moves beyond feature ranking and provides a complete framework for craniofacial growth analysis and prediction.

The rest of this paper is organized as follows. Section 2 reviews related work in craniofacial growth prediction, facial aging, and transformer-based facial analysis. Section 3 describes the dataset, feature extraction process, and preprocessing steps. Section 4 presents the proposed growth-oriented analysis framework. Section 5 introduces the LSTM–Temporal–Spatial Transformer model. Section 6 reports the experimental results and comparative evaluations. Section 7 concludes the paper and discusses future research directions.

2 Related Work

Predicting craniofacial growth and facial aging has been an important research topic in orthodontics, forensic science, computer vision, and biomedical research. Early studies mainly relied on cephalometric measurements, statistical growth curves, and geometric analysis to estimate future facial shape. These methods provided useful clinical information. However, they had limited ability to model nonlinear growth patterns and long-term developmental changes. Recent advances in machine learning and deep learning have improved craniofacial growth analysis. Zhang *et al.* [9] proposed a deep convolutional neural network for predicting mandibular growth from cephalometric radiographs. Their ResNet-based model achieved high prediction accuracy and identified facial regions associated with mandibular development. Kim *et al.* [10] evaluated several machine-learning models for longitudinal craniofacial growth prediction. Their results showed that nonlinear learning methods outperform traditional statistical approaches for skeletal growth forecasting. Several recent studies have also explored machine-learning methods for long-term craniofacial prediction. Myers *et al.* [11] investigated multiple machine-learning algorithms for predicting skeletal and dental growth changes. Their study showed the potential of data-driven models for orthodontic applications. Roseth *et al.* [12] compared artificial-intelligence-based models with traditional partial least-squares methods. Their results showed improved prediction performance with deep learning models. At the same time, substantial research has focused on facial aging and age progression. Modern facial-aging systems use deep neural networks, generative adversarial networks (GANs), diffusion models, and transformer architectures to estimate age or generate future facial appearance. Reviews by Kale *et al.* [13] and other recent surveys show a shift from handcrafted aging models to deep-learning approaches. Transformer-based age-estimation methods have also demonstrated the ability of self-attention mechanisms to capture global facial structure and long-range facial relationships [14, 15]. Artificial intelligence has also become increasingly common in orthodontics and craniofacial analysis. Recent studies

have applied deep learning to landmark detection, cephalometric analysis, treatment planning, and growth assessment. Reviews show improvements in automated craniofacial measurements and growth prediction [4]. Attention-based models have also been used for three-dimensional craniofacial landmark localization and soft-tissue analysis. These studies demonstrate the value of transformer mechanisms for modeling complex anatomical structures [16]. Despite these advances, important limitations remain. First, most studies use cross-sectional data or sparsely sampled longitudinal observations. Dense longitudinal datasets collected over many years remain rare. Second, many approaches work directly with images or cephalometric radiographs without examining the developmental behavior of individual craniofacial descriptors. Third, most forecasting models focus on prediction accuracy and do not examine whether facial descriptors follow consistent growth patterns across individuals or coordinated growth relationships across facial regions. Finally, transformer-based methods are usually developed for age estimation, landmark detection, or image synthesis rather than longitudinal craniofacial growth forecasting. This study addresses these limitations. Unlike previous work, it uses a dense longitudinal facial-growth dataset covering approximately ten years. The analysis is performed at the descriptor level rather than only at the image level. A growth-oriented analysis framework is introduced to identify descriptors with temporal dependency, age discriminability, predictive importance, developmental consistency, and coordinated growth behavior. The proposed LSTM–Temporal–Spatial Transformer model is then designed based on these findings. The model captures temporal continuity, long-range developmental dependencies, and coordinated relationships between facial structures. As a result, the proposed framework provides a growth-aware approach for longitudinal craniofacial forecasting that extends beyond conventional facial-aging, age-estimation, and craniofacial prediction methods.

3 Dataset and Feature Extraction

3.1 Dataset Description

This study uses a longitudinal facial-growth dataset collected over approximately ten years. The dataset covers the developmental period from early adolescence to early adulthood. Unlike conventional facial datasets that contain independent samples from many individuals, this dataset contains dense temporal observations from three male individuals.

Each individual is represented by a separate dataset of frontal facial images collected on a near-daily basis. After data cleaning and feature extraction, the datasets contained 2944, 2097, and 1713 observations for Individuals I, II, and III, respectively. The corresponding age ranges were $[12, 22]$, $[13, 23]$, and $[14, 24]$ years. Using only male subjects reduces sex-related growth variation and allows the analysis to focus on longitudinal developmental patterns. The three datasets were later aligned to a common developmental interval, as described in Section 3.3.

To protect privacy, the original facial images are not publicly available. Instead, the extracted geometric feature datasets are publicly available through Kaggle [17].

The dataset was designed to capture long-term craniofacial development rather than identity-specific appearance. Therefore, each observation is represented by a set of geometric facial descriptors derived from facial landmarks.

3.2 Facial Landmark Detection and Feature Extraction

Facial landmarks were detected using the pretrained `shape_predictor_68_face_landmarks` model. The model provides 68 anatomical landmarks located on the eyebrows, eyes, nose, mouth, and facial contour.

Using these landmarks, we extracted 29 geometric descriptors from each image. These descriptors capture facial dimensions, anatomical proportions, angular relationships, and symmetry properties. They are also less sensitive to image scale and small acquisition variations. The extracted features were grouped into the following categories:

- **Ocular and eyebrow descriptors:** eye distance, left and right eye width, left and right eye height, left and right eye aspect ratio, left and right eyebrow height, and eyebrow distance.
- **Nasal and oral descriptors:** nose width, nose height, nose ratio, mouth width, mouth height, and mouth ratio.
- **Symmetry descriptors:** eye symmetry, eyebrow symmetry, and jaw symmetry.
- **Angular descriptors:** jaw angle, nose angle, and eye angle.
- **Normalized proportional descriptors:** eye-to-face width ratio, mouth-to-face width ratio, and nose-to-face height ratio.
- **Global craniofacial dimensions:** face width, face height, jaw width, and forehead-to-nose length.

Together, these descriptors provide a compact representation of craniofacial morphology. They form the basis of the growth analyses and forecasting models presented in this study.

3.3 Temporal Preprocessing and Age Alignment

To support longitudinal growth analysis, a temporal preprocessing pipeline was applied to all feature sequences. First, outlier correction was performed for each descriptor separately. Extreme values were identified and replaced with the mean value of the corresponding feature. This step reduced the influence of anomalous measurements while preserving the overall growth pattern. Second, all descriptors were normalized to the range $[0, 1]$. This allowed direct comparison between features with different numerical scales. A moving-average smoothing filter was then applied to reduce measurement noise and short-term fluctuations. This produced smoother growth trajectories over time.

Third, long-term growth trends were extracted using the Fast Fourier Transform (FFT), a common signal-processing method for separating long-term trends from high-frequency variations in temporal data [18, 19]. Each feature sequence was transformed into the frequency domain. High-frequency components associated with short-term variation were removed. The filtered signals were then reconstructed using the inverse

FFT. This process produced smoother developmental trajectories that emphasized long-term craniofacial growth patterns. It also reduced the influence of short-term fluctuations and acquisition noise.

Finally, age alignment was performed because the three individuals covered different age ranges. Only the common developmental interval between ages 14 and 22 years was retained. The original age ranges were $[12, 22]$, $[13, 23]$, and $[14, 24]$ years for Individuals I, II, and III, respectively. After alignment, the final datasets contained 2355, 1678, and 1370 observations for Individuals I, II, and III, respectively. This step enabled direct comparison of growth trajectories and developmental relationships across individuals.

The aligned and preprocessed datasets form the basis of the growth analyses presented in Section 4 and the forecasting framework described in Section 5.

4 Growth-Oriented Feature Analysis

The proposed dataset differs from conventional facial datasets used for recognition or classification. Instead of independent images from many individuals, it contains dense longitudinal observations collected daily over an eight-year period. The data cover the transition from adolescence to early adulthood. As a result, each facial descriptor forms a temporal signal that reflects the growth of a specific craniofacial structure.

This temporal structure introduces two main challenges. First, facial descriptors do not contribute equally to growth analysis. Some descriptors change with age. Others remain stable or show short-term variation. Second, a descriptor that appears important for one individual may not represent a common growth pattern across different individuals. Therefore, meaningful growth descriptors must be evaluated based on both developmental relevance and cross-individual consistency.

To address these challenges, a two-stage growth-oriented analysis framework was developed. The first stage, referred to as *Feature Importance Analysis*, identifies descriptors with high temporal dependency, age discriminability, and predictive importance within facial growth trajectories. The second stage, referred to as *Growth Pattern Characterization*, examines the temporal behavior of each descriptor after noise reduction and trend extraction. This stage measures the extent to which growth trajectories are shared across individuals.

Together, these analyses identify facial descriptors that are both informative for age progression and representative of common craniofacial growth patterns. These descriptors provide suitable inputs for subsequent growth-aware facial forecasting models.

4.1 Feature Importance Analysis

The goal of this analysis is to identify the facial descriptors that provide the most information about craniofacial growth and age progression. Feature importance is evaluated from three perspectives. The Temporal Dependency (TD) score measures how much past values of a feature relate to future facial development. The Age Discriminability (AD) score measures the relationship between a feature and chronological age. The Neural Saliency (NS) score measures how much a feature contributes to a

temporal prediction model. Features with higher NS values have a greater impact on prediction. Let

$$X = \left\{ x_t^{(f)} \right\}_{t=1}^{T}$$

be the temporal sequence of facial feature f, where T is the number of observations. Before analysis, all features are standardized using z-score normalization,

$$\hat{x}_t^{(f)} = \frac{x_t^{(f)} - \mu_f}{\sigma_f}$$

where μ_f and σ_f are the mean and standard deviation of feature f.

4.1.1 Temporal Dependency Score

The Temporal Dependency score measures the relationship between past feature values and later facial development. For a temporal lag L, the score is defined as

$$S_{\mathrm{TD}}^{(f)} = \left| \mathrm{corr} \left(x_{1:T-L}^{(f)}, \bar{X}_{L+1:T} \right) \right|,$$

where $\bar{X}$ is the average representation across all facial descriptors. Features with stronger temporal relationships receive higher scores.

4.1.2 Age Discriminability Score

The Age Discriminability score measures how much age-related information is contained in a feature. It is calculated using mutual information between feature values and age labels,

$$S_{\mathrm{MI}}^{(f)} = I(X^{(f)}; A),$$

where A is the age variable and $I(\cdot)$ denotes mutual information. Mutual information is widely used for feature evaluation because it captures both linear and nonlinear relationships between variables [20, 21].

4.1.3 Neural Saliency Score

A lightweight Long Short-Term Memory (LSTM) network is used to estimate the predictive contribution of each feature. The saliency score is computed from the average absolute gradient magnitude,

$$S_{\mathrm{NS}}^{(f)} = \frac{1}{T} \sum_{t=1}^{T} \left| \frac{\partial y}{\partial x_t^{(f)}} \right|.$$

Features with larger gradients have a greater influence on the learned temporal representation. Gradient-based saliency analysis is widely used to estimate feature importance in deep learning models because it measures how sensitive model outputs are to input changes [22].

4.1.4 Consensus Feature Ranking

The three scores are normalized to the range $[0, 1]$,

$$\tilde{S}_k^{(f)} = \frac{S_k^{(f)} - \min(S_k)}{\max(S_k) - \min(S_k)},$$

and combined using equal weights,

$$S_{\text{Final}}^{(f)} = \frac{\tilde{S}_{\text{TD}}^{(f)} + \tilde{S}_{\text{MI}}^{(f)} + \tilde{S}_{\text{NS}}^{(f)}}{3}.$$

Because the dataset contains three individuals, feature rankings are calculated separately for each one. The rankings are then combined using the average rank across all datasets. The final ranking identifies descriptors that consistently show temporal relevance, age dependency, and predictive importance across individuals.

4.1.5 Results and Analysis

Table 1 presents the feature rankings obtained from the TD, AD, and NS scores, together with the final consensus ranking.

Table 1: Comparison of feature rankings obtained using Temporal Dependency (TD), Age Discriminability (AD), Neural Saliency (NS), and Consensus Ranking.

Rank	TD	AD (MI)	NS	Consensus
1	Face Height	Face Width	Jaw Width	Left Eye Width
2	Left Eye Width	Jaw Angle	Mouth Ratio	Jaw Width
3	Face Width	Mouth Width	Eye Aspect Ratio (Right)	Right Eye Width
4	Right Eye Width	Left Eye Width	Left Eye Width	Face Width
5	Nose Width	Face Height	Eye Distance	Nose Height
6	Eye Distance	Nose Height	Left Eyebrow Height	Forehead-to-Nose
7	Jaw Width	Forehead-to-Nose	Left Eye Height	Face Height
8	Mouth Width	Jaw Width	Eye Symmetry	Eye Distance
9	Nose Height	Right Eye Width	Right Eyebrow Height	Jaw Angle
10	Eyebrow Distance	Nose-to-Face Height Ratio	Nose Ratio	Nose Width

The TD rankings show that global facial dimensions, eye measurements, and nasal features have the strongest temporal growth patterns. The AD rankings show that facial width, jaw shape, eye measurements, and nasal development contain the most age-related information. The NS rankings provide a prediction-based view of feature importance. Jaw width, eye measurements, and several facial proportions received high saliency scores. This indicates that these features contribute strongly to temporal prediction. Although the NS rankings differ from the TD and AD rankings in some cases, they support the importance of several growth-related descriptors.

The consensus ranking identifies a common group of informative features. These include left eye width, jaw width, right eye width, face width, nose height, face height, and eye distance. These descriptors consistently show high temporal dependency, strong association with age, and high predictive contribution. Together, they provide a concise representation of longitudinal craniofacial growth.

4.2 Growth Pattern Characterization

The previous stage identifies informative descriptors. However, it does not determine whether their growth trajectories follow shared developmental patterns. Therefore, a second analysis stage is introduced to examine the temporal behavior of each descriptor and assess its consistency across individuals.

For each feature sequence, extreme observations are first suppressed to reduce the effect of outliers. The signal is then smoothed using a moving-average filter,

$$s_t = \frac{1}{W} \sum_{i=0}^{W-1} x_{t-i},$$

where W is the smoothing window size. The smoothed signal is then normalized using min-max normalization,

$$\hat{s}_t = \frac{s_t - \min(s)}{\max(s) - \min(s)}.$$

To extract long-term growth trends, frequency-domain filtering is applied using the Fast Fourier Transform (FFT). Let

$$F(k) = \mathcal{F}\{\hat{s}_t\}$$

represent the frequency-domain signal. High-frequency components associated with short-term variation are removed. The filtered signal is then reconstructed using the inverse transform,

$$s_t^* = \mathcal{F}^{-1}\{F(k)\}.$$

After preprocessing, the trajectories of all individuals are aligned to a common length. Pairwise Pearson correlation is then computed. The mean cross-individual correlation is defined as

$$C_f = \frac{1}{M} \sum_{i<j} \operatorname{corr}\left(s_i^{*(f)}, s_j^{*(f)}\right),$$

where M is the total number of pairwise comparisons. This measure quantifies the degree to which a descriptor follows a shared growth trajectory across individuals.

4.2.1 Growth Signal Characterization

Correlation measures the similarity between growth trajectories. However, it does not describe the statistical properties of the signals. Therefore, several signal measures are

extracted from the average developmental trajectory. Signal energy is calculated to measure the cumulative magnitude of growth,

$$\text{Energy} = \sum_{t=1}^{T} (s_t^*)^2,$$

Signal variance is calculated to measure variability, where $\bar{s}$ is the mean signal value,

$$\text{Variance} = \frac{1}{T} \sum_{t=1}^{T} (s_t^* - \bar{s})^2,$$

The root mean square (RMS) value measures the overall growth magnitude,

$$\text{RMS} = \sqrt{\frac{1}{T} \sum_{t=1}^{T} (s_t^*)^2},$$

Signal entropy is calculated as

$$\text{Entropy} = -\sum_{i} p_i \log p_i,$$

where p_i is the probability of the i-th histogram bin in the signal distribution. Skewness and kurtosis are also calculated. Skewness measures the asymmetry of the growth trajectory. Kurtosis measures the concentration of values around the central region of the distribution. These measures describe the magnitude, variability, complexity, and overall structure of craniofacial growth signals.

4.2.2 Results and Analysis

Table 2 presents the ten facial descriptors with the highest cross-individual consistency scores. These descriptors show the most similar growth trajectories across the three individuals after trend extraction and temporal alignment.

The highest consistency scores were observed for eyebrow distance (0.839), jaw width (0.836), eye distance (0.836), left eye width (0.822), nose height (0.810), and right eye width (0.795). These descriptors show the most similar developmental patterns across individuals. They are therefore strong candidates for representing common craniofacial growth processes.

Table 3 reports the signal measures for the ten highest-ranked descriptors. Jaw width, left eye width, eye distance, face height, and right eye width show high energy, RMS, and variance values. This pattern indicates substantial growth and clear temporal change during adolescence. Most descriptors also have negative kurtosis values, suggesting wider signal distributions. The entropy values indicate that many growth trajectories follow organized and stable patterns over time.

The correlation and signal analyses show that several craniofacial descriptors follow similar developmental trajectories across individuals. Measurements related to

Table 2: Top-10 facial descriptors ranked by mean cross-individual correlation.

Feature	Mean Correlation
eyebrow_distance	0.839
jaw_width	0.836
eye_distance	0.836
left_eye_width	0.822
nose_height	0.810
right_eye_width	0.795
face_width	0.759
face_height	0.637
eye_symmetry	0.586
left_eye_height	0.574

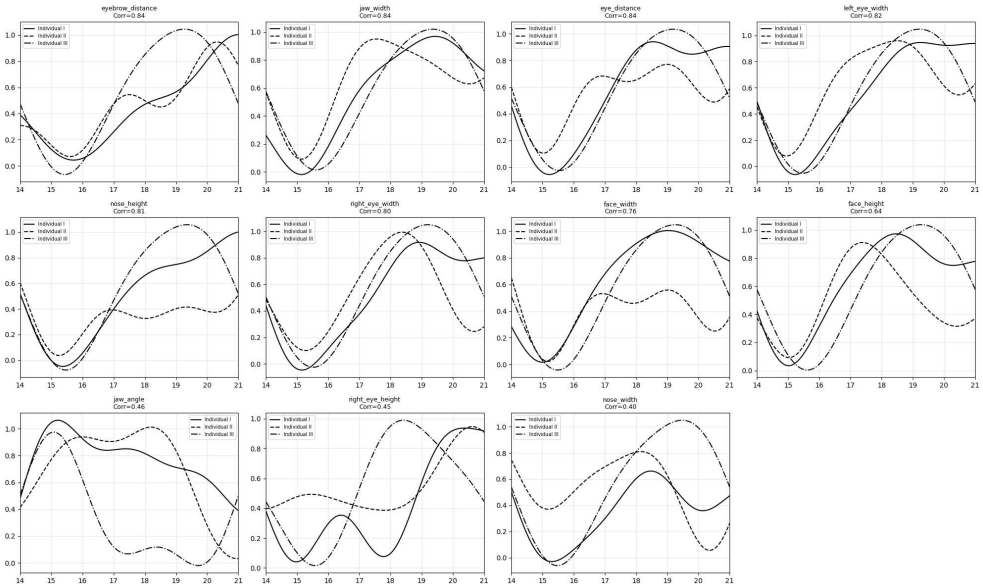


Fig. 1: Normalized developmental trajectories of the highest-ranked growth-consistent facial descriptors across the three individuals.

the eyes, jaw, nose, and overall facial dimensions show both high consistency and substantial growth. These findings suggest that parts of facial development follow shared growth patterns. They also indicate that facial growth can be represented as an organized temporal process.

Table 3: Signal characterization metrics for the ten facial descriptors exhibiting the highest cross-individual growth consistency.

Feature	Energy	Variance	Skewness	Kurtosis	RMS	Entropy
eyebrow_distance	338.666	0.088	-0.139	-1.415	0.562	-5.919
jaw_width	463.854	0.097	-0.501	-1.316	0.658	-9.353
eye_distance	412.525	0.090	-0.449	-1.288	0.621	-7.438
left_eye_width	458.249	0.109	-0.421	-1.319	0.654	-4.947
nose_height	299.881	0.075	-0.503	-1.270	0.529	-14.868
right_eye_width	397.037	0.101	-0.187	-1.351	0.609	-3.803
face_width	371.553	0.081	-0.472	-1.174	0.589	-7.332
face_height	408.895	0.076	-0.503	-1.204	0.618	-10.523
eye_symmetry	319.146	0.046	-0.857	-0.937	0.546	-32.023
left_eye_height	282.044	0.061	0.051	-1.269	0.513	-11.141

4.3 Cross-Group Coordinated Growth Analysis

The previous analysis identified descriptors with consistent growth trajectories across individuals. However, it did not examine how different facial regions develop together. Craniofacial growth is a coordinated process. Changes in one facial structure are often associated with changes in related structures. Therefore, understanding the relationships between facial descriptors is important for studying facial development.

To examine these relationships, a cross-group coordinated growth analysis is performed. The goal is to identify facial descriptors that change together during adolescence and early adulthood. The analysis also evaluates whether these coordination patterns are shared across individuals.

4.3.1 Inter-Feature Correlation Analysis

For each individual dataset, pairwise correlations were calculated between all facial descriptors. Let $x^{(f_i)}$ and $x^{(f_j)}$ denote the temporal sequences of features f_i and f_j, respectively. Coordinated growth between the two features was measured using Spearman's rank correlation coefficient,

$$\rho_{ij} = \mathrm{corr}_{\mathrm{Spearman}}\left(x^{(f_i)}, x^{(f_j)}\right).$$

Spearman correlation measures monotonic relationships and is less sensitive to nonlinear growth patterns and outliers than Pearson correlation [23]. The resulting correlation matrix describes how facial descriptors change together over time. For each individual, all feature pairs were ranked according to the absolute value of their correlation coefficients. The highest-ranked pairs were considered the most coordinated growth relationships.

4.3.2 Common Coordination Patterns

To identify growth relationships shared across individuals, the highest-ranked feature pairs from each dataset were compared. Let

$$P_k$$

denote the set of top-ranked feature pairs for individual k. Common coordination patterns were identified using set intersection,

$$P_{\text{common}} = \bigcap_{k=1}^{N} P_k,$$

where N is the number of individuals. Feature pairs that appeared in all datasets were considered stable growth relationships that persisted across individuals.

4.3.3 Results and Analysis

Table 4 summarizes the strongest coordinated growth relationships identified for each individual. The analysis examined all feature pairs. To keep the presentation concise, only the three highest-ranked pairs are reported for each subject. Although the rankings vary across individuals, several common patterns are observed.

High correlations are found between anatomically related features. These include bilateral eye measurements, eyebrow features, and global facial dimensions. In Individual I, the strongest relationships occur mainly among eye and eyebrow descriptors. This pattern suggests coordinated development between eye dimensions and nearby facial structures. Individual II shows similar coordination between bilateral features. It also shows close relationships between facial dimensions and jaw morphology. In Individual III, several global facial measurements form a connected growth group. Face width appears in multiple high-ranking relationships with jaw width, face height, and eye measurements. The high correlations in this subject indicate coordinated growth among major craniofacial structures during adolescence and early adulthood.

To identify growth relationships shared across individuals, the highest-ranked feature pairs were compared across all datasets. The common relationships are presented in Table 5. Two feature pairs appeared among the strongest coordinated growth patterns in all individuals. The first pair is *left eye width* and *right eye width*. The second pair is *face width* and *jaw width*.

The relationship between left and right eye width reflects coordinated bilateral eye development. Both features show proportional growth throughout the observation period. The relationship between face width and jaw width indicates a stable connection between overall facial width and mandibular growth. This pattern suggests that jaw development contributes to changes in overall facial shape. The presence of these relationships in all datasets indicates that some aspects of craniofacial growth follow common structural patterns across individuals.

Table 4: Top coordinated feature pairs identified for each individual using Spearman correlation analysis.

Individual	Feature 1	Feature 2	Correlation
Individual I	right_eye_height	eye_aspect_ratio_right	0.961
	left_eye_height	eye_aspect_ratio_left	0.951
	left_eyebrow_height	right_eyebrow_height	0.922
Individual II	left_eyebrow_height	right_eyebrow_height	0.929
	left_eye_height	right_eye_height	0.894
	face_width	jaw_width	0.889
Individual III	left_eye_width	right_eye_width	0.972
	face_width	jaw_width	0.971
	face_width	face_height	0.966

Overall, these findings show that facial growth involves both temporal change and structural coordination between facial regions. The presence of shared growth relationships across individuals suggests that craniofacial development follows an organized spatial pattern rather than isolated changes in individual features. These observations motivate the inclusion of spatial modeling components in the proposed forecasting framework, allowing the model to learn interactions among related facial descriptors during growth prediction.

To identify growth relationships shared across all individuals, the top-ranked correlation pairs were compared between datasets. Two feature pairs appeared consistently in all three datasets:

Table 5: Common coordinated growth relationships observed across all individuals.

Feature 1	Feature 2	Correlation Range
left_eye_width	right_eye_width	0.807 – 0.972
face_width	jaw_width	0.770 – 0.971

The relationship between left eye width and right eye width reflects coordinated bilateral eye development. The correlation appears in all three individuals. This finding suggests that both eyes grow in a proportional manner throughout adolescence and early adulthood.

The relationship between face width and jaw width reflects coordinated development between facial width and mandibular growth. Because the jaw contributes to overall facial shape, changes in these two measurements occur together across the developmental period.

Overall, the coordinated growth analysis shows that facial development does not occur through isolated changes in individual descriptors. Several facial regions develop together and maintain stable relationships over time. These findings suggest that craniofacial growth follows an organized spatial pattern. This observation supports

Table 6: Summary of the principal findings obtained from the three growth analyses.

Feature Importance Analysis
Left eye width, jaw width, right eye width, face width, nose height, face height, eye distance
Growth Pattern Characterization
Eyebrow distance, jaw width, eye distance, left eye width, nose height, right eye width, face width, face height
Cross-Group Coordinated Growth Analysis
Left eye width–right eye width Face width–jaw width

the inclusion of spatial attention mechanisms in the proposed forecasting framework so that interactions between related facial descriptors can be learned during growth prediction.

4.4 Growth Characteristics and Model Design Motivation

The previous analyses provide different perspectives on longitudinal facial growth. Feature importance analysis identified descriptors that contain the most growth-related information. Growth pattern characterization identified descriptors with similar developmental trajectories across individuals. Cross-group coordinated growth analysis identified stable relationships between facial regions. Together, these findings suggest that facial growth is both a temporal and coordinated process.

Table 6 summarizes the main descriptors identified by each analysis stage. Several descriptors appear in more than one analysis, including *jaw width*, *eye distance*, *left eye width*, *right eye width*, *face width*, *face height*, and *nose height*. These descriptors show age-related variation, developmental consistency, and predictive importance.

The overlap between the first two analyses is important. Jaw width, eye distance, left eye width, right eye width, face width, face height, and nose height were identified as both informative and consistent across individuals. These descriptors capture key aspects of craniofacial development and provide a compact representation of facial growth.

The coordinated growth analysis also showed that facial descriptors do not evolve independently. Several facial regions exhibit stable relationships throughout development. The clearest examples are the relationship between left and right eye width and the relationship between face width and jaw width. These results indicate that interactions between facial descriptors are part of the growth process.

These findings motivated the design of the proposed LSTM–Temporal–Spatial Transformer (LSTM-TS-TF). The temporal consistency observed in many descriptors suggests that facial growth can be modeled as a sequential process. This motivated the use of Long Short-Term Memory (LSTM) networks to capture local temporal continuity. Long-term developmental trends observed across multiple years motivated the

use of a Temporal Transformer to model long-range temporal dependencies. The coordinated relationships identified between facial regions motivated the use of a Spatial Transformer to model interactions among related facial descriptors.

The proposed forecasting framework was therefore designed to capture three characteristics of facial growth: temporal continuity, long-range developmental dependencies, and coordinated growth between facial regions. By combining these components within a single architecture, the LSTM–Temporal–Spatial Transformer provides a representation of facial growth that incorporates both temporal and structural information.

5 Growth-Aware Facial Forecasting Framework

5.1 Architecture Overview

The analyses presented in the previous section revealed several important characteristics of longitudinal facial growth.

First, the feature importance analysis showed that only a subset of facial descriptors contributes substantially to age-related growth. Features related to eye dimensions, jaw morphology, nasal structures, and overall facial dimensions consistently ranked among the most informative across all individuals. This finding suggests that some facial regions contain more growth-related information than others.

Second, the growth pattern analysis showed that several facial descriptors follow similar temporal trajectories across individuals. Features such as jaw width, eye distance, eye width, face dimensions, and nose height exhibited high agreement between subjects. This observation suggests the presence of shared developmental patterns during craniofacial growth. It also indicates that facial growth can be modeled as a structured temporal process.

Third, the coordinated growth analysis revealed stable relationships between multiple facial descriptors. These relationships were particularly evident between bilateral eye measurements and between facial and jaw dimensions. This finding suggests that facial growth involves interactions between anatomical regions rather than isolated changes in individual features.

Traditional forecasting models cannot capture all of these characteristics. Long Short-Term Memory (LSTM) networks can model local temporal changes and sequential continuity. However, they often have difficulty learning long-range dependencies across extended developmental periods. Recent studies have shown that Transformer models can improve long-horizon forecasting by learning long-range temporal relationships through self-attention mechanisms [24, 25]. However, temporal Transformer models focus mainly on temporal relationships and may not fully exploit structural dependencies between facial regions. Models that focus only on temporal evolution may also overlook coordinated anatomical growth patterns. These observations suggest that a facial growth forecasting model should satisfy three requirements:

1. Capture the sequential nature of facial growth and preserve long-term developmental information.
2. Learn temporal dependencies that span multiple years.

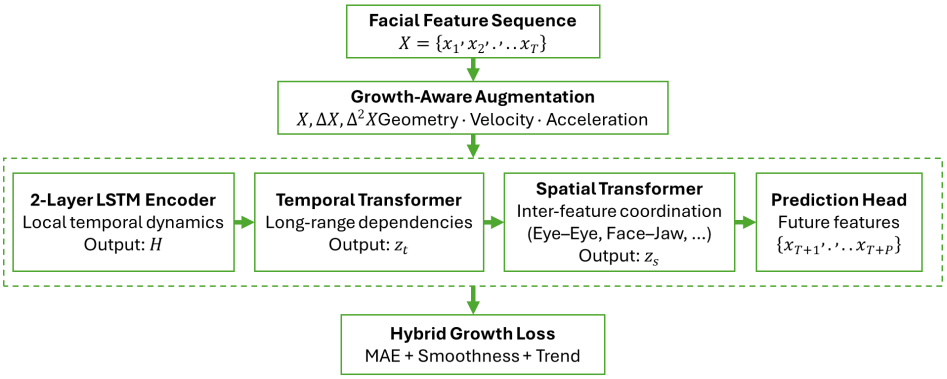


Fig. 2: Overview of the proposed LSTM–Temporal–Spatial Transformer (LSTM-TS-TF). Facial descriptors are augmented with velocity and acceleration information to form a growth-aware representation. A two-layer LSTM encoder captures local growth dynamics. A Temporal Transformer models long-range temporal dependencies. A Spatial Transformer learns relationships among facial descriptors. The final representation is used to predict future facial development and is optimized using a hybrid loss that combines prediction accuracy, smoothness, and trend consistency.

3. Model interactions between anatomically related facial descriptors.

To address these requirements, we propose the **LSTM–Temporal–Spatial Transformer (LSTM-TS-TF)** architecture. The model combines recurrent modeling, temporal self-attention, and spatial self-attention within a single framework. Figure 2 presents an overview of the proposed architecture.

Given an input sequence of facial descriptors,

$$X = x_1, x_2, \ldots, x_T, \qquad x_t \in \mathbb{R}^F,$$

the model first builds a growth-aware representation by augmenting the original descriptors with first-order and second-order temporal derivatives. The resulting representation is then processed through three learning stages. A multi-layer LSTM encoder captures local growth dynamics and short-term temporal continuity. A Temporal Transformer encoder then models relationships between distant developmental stages. Next, a Spatial Transformer encoder learns interactions among facial descriptors and captures structural growth relationships identified during the coordination analysis. The final latent representation is projected into the future feature space to generate growth forecasts. Training uses a hybrid loss function that combines prediction accuracy, temporal smoothness, and trend consistency. This objective encourages realistic growth trajectories.

The proposed framework combines three complementary learning components. The LSTM encoder captures temporal continuity. The Temporal Transformer learns long-range developmental dependencies. The Spatial Transformer models coordinated relationships between facial descriptors. Together, these components address the

Table 7: Configuration of the proposed LSTM–Temporal–Spatial Transformer.

Component	Setting
Input Window	180 time steps
Prediction Horizon	10 time steps
LSTM Encoder	2 layers, 128 units
Temporal Transformer	1 layer, 4 heads
Spatial Transformer	1 layer, 4 heads
Embedding Dimension	128
Optimizer	AdamW (10^{-3})
Batch Size	32

three main characteristics identified in Section 4: temporal continuity, long-range developmental dependencies, and coordinated craniofacial growth.

5.1.1 Implementation Configuration

Table 7 summarizes the main architectural and training parameters used in the experiments.

The same configuration was used for all individuals and experiments. This setup ensures a consistent comparison between forecasting models.

5.1.2 Growth-Aware Input Representation

Let

$$X = \{x_1, x_2, \ldots, x_T\}, \qquad x_t \in \mathbb{R}^F$$

represent a sequence of facial feature vectors over time, where F is the number of geometric facial descriptors. To incorporate growth dynamics, first-order and second-order derivatives are calculated for each feature:

$$\Delta X_t = X_t - X_{t-1},$$

$$\Delta^2 X_t = \Delta X_t - \Delta X_{t-1}.$$

The final input representation is formed by concatenating the original features, velocity features, and acceleration features:

$$X_{aug} = [X, \Delta X, \Delta^2 X].$$

This representation captures facial geometry, growth rate, and growth acceleration.

5.1.3 Temporal Feature Encoding

LSTM Encoder

Long Short-Term Memory (LSTM) networks were introduced to address the vanishing-gradient problem in recurrent neural networks. They are widely used for sequential

and time-series modeling [26]. The first stage uses a multi-layer LSTM network to encode short-term and medium-term growth dynamics:

$$H = \text{LSTM}(X_{aug}),$$

where

$$H \in \mathbb{R}^{T \times d_h}$$

contains the hidden temporal representations. The LSTM captures sequential continuity and smooth growth transitions that are common in craniofacial development.

Temporal Transformer

LSTM networks capture local temporal patterns. However, their ability to model distant dependencies decreases as sequence length increases. To address this limitation, the LSTM representations are processed by a Transformer encoder. The Transformer uses self-attention to model relationships between distant time points [27]. Given the LSTM output

$$H = h_1, h_2, \ldots, h_T, \qquad H \in \mathbb{R}^{T \times d_h},$$

the Temporal Transformer constructs query, key, and value representations:

$$Q = HW_Q, \qquad K = HW_K, \qquad V = HW_V,$$

where W_Q, W_K, and W_V are learnable projection matrices. Self-attention is computed as

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V,$$

where d_k is the dimension of the key vectors. This operation allows each time point to attend to all other time points in the sequence. Multi-head attention is used to learn different temporal dependency patterns. Each attention head focuses on a different representation of the sequence [27]. The final temporal representations are obtained through the Transformer encoder:

$$Z_t = \text{Transformer}_{temp}(W_t H),$$

where W_t is a learnable projection matrix. The Temporal Transformer complements the LSTM encoder by modeling long-range developmental dependencies that

5.1.4 Spatial Dependency Modeling

The coordinated growth analysis in Section 4 identified stable relationships between several facial descriptors. Strong relationships were observed between left and right eye measurements and between face width and jaw width. These findings suggest that craniofacial growth involves interactions between related facial structures rather than independent changes in each descriptor. To model these relationships, a Spatial Transformer encoder is introduced:

$$Z_s = \text{Transformer}_{sp}(W_s Z_t),$$

where W_s projects the temporal representations into a spatial attention space. Unlike temporal attention, which operates across time, spatial attention operates across facial descriptors. This allows the model to learn relationships between facial regions and use coordinated growth patterns during forecasting.

5.1.5 Prediction Layer

The output of the Spatial Transformer is projected back to the original feature space:

$$Y_f = W_o Z_s,$$

where W_o is the output projection matrix. The forecast sequence is then defined as

$$\hat{Y} = \{\hat{x}_{T+1}, \hat{x}_{T+2}, \ldots, \hat{x}_{T+P}\},$$

where $\hat{Y}$ is the predicted future facial-feature sequence and P is the prediction horizon.

5.1.6 Hybrid Growth Loss Function

To improve prediction accuracy and preserve realistic growth behavior, the proposed framework is trained using a hybrid loss function:

$$\mathcal{L} = \mathcal{L}_{MAE} + \alpha \mathcal{L}_{smooth} + \gamma \mathcal{L}_{trend},$$

where $\alpha = 0.05$ and $\gamma = 0.01$ control the contributions of the smoothness and trend-consistency terms. The main forecasting error is measured using Mean Absolute Error (MAE):

$$L_{MAE} = \frac{1}{P} \sum_{t=1}^{P} |\hat{y}_t - y_t| \,,$$

A smoothness term is added to reduce unrealistic fluctuations in the predicted trajectories:

$$\mathcal{L}_{smooth} = \frac{1}{P-1} \sum_{t=2}^{P} \|\hat{y}_t - \hat{y}_{t-1}\|_2^2 \,.$$

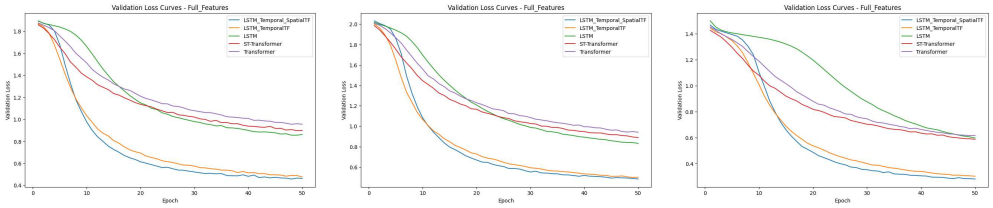
A trend-consistency term is also included to preserve the direction of growth:

$$\mathcal{L}_{trend} = \frac{1}{P-1} \sum_{t=2}^{P} \|(\hat{y}_t - \hat{y}_{t-1}) - (y_t - y_{t-1})\|_2^2 \,.$$

The final objective minimizes prediction error, encourages smooth trajectories, and preserves growth trends.

6 Experimental Evaluation and Model Comparison

The proposed forecasting framework was evaluated using the three longitudinal facial-growth datasets. Experiments were conducted using both the complete feature set



(a) Individual I (b) Individual II (c) Individual III

Fig. 3: Validation loss curves obtained using the full feature set for the five forecasting models across the three individuals. The proposed LSTM–Temporal–Spatial Transformer achieves the lowest validation loss and converges faster than the other models. The results show the benefit of combining temporal continuity, long-range dependencies, and inter-feature relationships within a single framework.

and the reduced feature subset identified by the feature-importance analysis. Five forecasting models were compared:

- LSTM
- Transformer
- ST-Transformer
- LSTM-Temporal Transformer (LSTM-TTF)
- Proposed LSTM-Temporal-Spatial Transformer (LSTM-TS-TF)

Model performance was evaluated using Mean Absolute Error (MAE). Lower MAE values indicate better forecasting accuracy.

6.1 Results on Full Feature Set

Figure 3 shows the validation loss curves for all forecasting models using the complete set of facial descriptors. Across all three individuals, the proposed LSTM–Temporal–Spatial Transformer converges faster and achieves the lowest validation loss. These results indicate that combining recurrent modeling with temporal and spatial attention helps capture both long-term growth patterns and relationships between facial descriptors.

Table 8 summarizes the forecasting results obtained using all extracted facial descriptors.

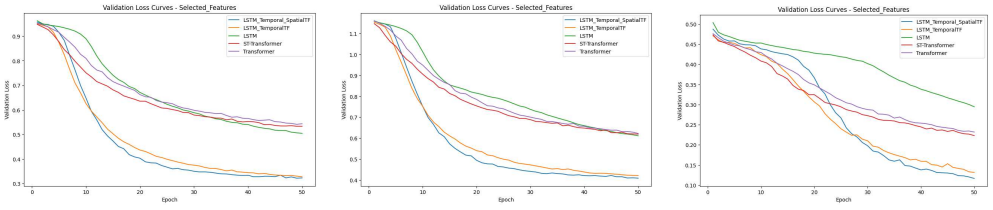
Across all datasets and evaluation settings, the LSTM-TS-TF model achieved the lowest forecasting error. Compared with the conventional LSTM model, the proposed architecture reduced prediction error by approximately 63–72%. This result highlights the value of combining recurrent modeling with attention mechanisms.

6.2 Results on Selected Features

Figure 4 shows the validation loss curves obtained using only the growth-sensitive features identified in Section 3. Across all three individuals, the LSTM–Temporal–Spatial Transformer again achieves the lowest validation loss and converges faster

Table 8: Forecasting results using the full feature set (MAE values). The standalone Transformer produced higher MAE than LSTM in all settings and is omitted for brevity. Complete results are provided in the supplementary material.

Dataset	Split	LSTM	ST-Transformer	LSTM-TTF	LSTM-TS-TF
Individual I	Standard	0.4448	0.4298	0.1813	**0.1652**
Individual I	Random	0.4375	0.4122	0.1758	**0.1632**
Individual II	Standard	0.4593	0.4150	0.1778	**0.1703**
Individual II	Random	0.4502	0.4009	0.1722	**0.1678**
Individual III	Standard	0.3265	0.2510	0.1015	**0.0929**
Individual III	Random	0.3187	0.2403	0.0971	**0.0901**



(a) Individual I (b) Individual II (c) Individual III

Fig. 4: Validation loss curves obtained using the selected feature subset for the five forecasting models across the three individuals. The proposed LSTM–Temporal–Spatial Transformer achieves the lowest validation loss and converges faster than the competing models. The results show that the selected growth-sensitive descriptors retain the main developmental information while reducing the number of input features.

than the other models. These results indicate that the selected descriptors retain the information needed for facial growth forecasting. The similar performance obtained with fewer features also suggests that the feature-selection process removes redundant descriptors while preserving the main developmental information.

Table 9 reports the forecasting performance obtained

The results show that forecasting accuracy improves when the analysis is restricted to growth-related and temporally stable descriptors. The proposed model achieved its best performance on Individual III, with MAE values of 0.0676 and 0.0640 under the standard and random settings, respectively. The performance gains were observed across all individuals and under both evaluation protocols. This consistency suggests that the proposed model generalizes across different data splits and individuals.

6.3 Result Analysis and Discussion

Several observations can be drawn from these results:

Table 9: Forecasting results using selected features (MAE values). The standalone Transformer produced higher MAE values than LSTM in all settings and is omitted for brevity.

Dataset	Split	LSTM	ST-Transformer	LSTM-TTF	LSTM-TS-TF
Individual I	Standard	0.3394	0.3761	0.1290	**0.1276**
Individual I	Random	0.3251	0.3535	0.1222	**0.1225**
Individual II	Standard	0.4657	0.4344	0.1890	**0.1724**
Individual II	Random	0.4593	0.4176	0.1830	**0.1699**
Individual III	Standard	0.2233	0.1718	0.0840	**0.0676**
Individual III	Random	0.2193	0.1630	0.0804	**0.0640**

1. The large performance gap between the LSTM model and the transformer-based models indicates that local temporal information alone is not sufficient for facial-growth forecasting. Long-range developmental dependencies contribute important information for prediction.
2. The improved performance of LSTM-TTF compared with both LSTM and ST-Transformer shows the benefit of combining recurrent temporal encoding with temporal self-attention.
3. The LSTM-TS-TF model achieved the best results across all experiments. This result highlights the importance of modeling relationships between facial descriptors. It is also consistent with the coordinated growth analysis, which revealed stable relationships among features such as eye measurements, nose height, jaw width, and face width.
4. The reduced-feature experiments show that selected growth descriptors provide more predictive information than the complete feature set. This finding suggests that the feature-analysis framework identifies descriptors that capture important aspects of facial growth while reducing the number of input features.

Overall, the results show that combining recurrent modeling, temporal attention, and spatial attention improves long-term craniofacial growth forecasting. The proposed framework captures both temporal growth patterns and relationships among facial structures.

7 Conclusion and Future Work

This study presented a framework for longitudinal craniofacial growth analysis and forecasting using a dense facial-growth dataset collected over approximately ten years. The dataset contains temporally ordered facial records from three individuals and covers the developmental period from adolescence to early adulthood. This dataset enabled the investigation of long-term craniofacial growth patterns from a longitudinal perspective. A multi-stage growth analysis framework was developed to study facial development at both the descriptor and structural levels. The analysis identified a group of growth-sensitive descriptors with high temporal dependency, strong age association, and high predictive value. These descriptors were mainly related to facial dimensions, eye measurements, jaw morphology, and nasal structures. The results also

showed that several facial descriptors follow similar developmental trajectories across individuals. In addition, coordinated growth analysis revealed stable relationships between anatomically related regions, including bilateral eye structures and the relationship between face width and jaw width. The findings from these analyses guided the development of the proposed LSTM–Temporal–Spatial Transformer (LSTM-TS-TF). The model combines recurrent modeling, temporal self-attention, and spatial self-attention within a single forecasting framework. This design allows the model to capture temporal continuity, long-range developmental dependencies, and relationships between facial descriptors. Experimental results showed that the proposed model achieved the lowest forecasting error across all datasets and evaluation settings. The results also provide insight into craniofacial development. Several growth trajectories were consistent across individuals, suggesting the presence of shared developmental patterns. The stable relationships observed between facial regions further indicate that craniofacial growth follows coordinated structural changes rather than isolated feature evolution. Future work will expand the dataset to include more individuals, wider age ranges, and greater demographic diversity. Further improvements may include three-dimensional facial data, multimodal craniofacial measurements, and graph-based anatomical representations. The integration of biological growth knowledge and clinical information may also support personalized craniofacial growth prediction. In summary, this work introduces a dense longitudinal facial-growth dataset, a growth-analysis framework, and an LSTM–Temporal–Spatial Transformer forecasting model. Together, these contributions support longitudinal craniofacial analysis and long-term facial growth prediction.

Editorial Policies for:

Springer journals and proceedings: https://www.springer.com/gp/editorial-policies

Nature Portfolio journals: https://www.nature.com/nature-research/editorial-policies

Scientific Reports: https://www.nature.com/srep/journal-policies/editorial-policies

BMC journals: https://www.biomedcentral.com/getpublished/editorial-policies

Declarations

Funding

The authors received no external funding for this research, authorship, or publication of this article.

Competing interests

The authors declare that they have no competing interests.

Ethics declaration

Not applicable. This study did not involve human participants, human subjects, animals, clinical trials, or the collection of identifiable personal information. All experiments were conducted using an existing publicly available longitudinal craniofacial image dataset, and no new data were collected from participants.

Consent for publication

Not applicable.

Data availability

The adulthood-transition dataset used in this study is publicly available through Kaggle at the following URL: https://www.kaggle.com/datasets/mostafaasalama/adulthood. Additional processed data and source code are available from the corresponding author upon reasonable request.

References

[1] Geng, X., Zhou, Z.-H., Smith-Miles, K.: Automatic age estimation based on facial aging patterns. IEEE Transactions on Pattern Analysis and Machine Intelligence **29**(12), 2234–2240 (2007) https://doi.org/10.1109/TPAMI.2007.70733

[2] Shen, W., Guo, Y., Wang, Y., Zhao, K., Wang, B., Yuille, A.L.: Deep regression forests for age estimation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2304–2313 (2018). https://doi.org/10.1109/CVPR.2018.00244

[3] Niu, Z., Zhou, M., Wang, L., Gao, X., Hua, G.: Ordinal regression with multiple output cnn for age estimation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 4920–4928 (2016). https://doi.org/10.1109/CVPR.2016.532

[4] Lyros, I., Vastardis, H., Tsolakis, I.A., Kotantoula, G., Lykogeorgos, T., Tsolakis, A.I.: Growth prediction in orthodontics: A systematic review of past methods up to artificial intelligence. Children **12**(8), 1023 (2025) https://doi.org/10.3390/children12081023

[5] Heribanová, B., Janáková, K., Tomášik, J., Tichá, D., Harsányi, Thurzo, A.: Current applications and future perspectives of artificial intelligence in face-driven orthodontics: A scoping review. Biomimetics **11**(2), 146 (2026) https://doi.org/10.3390/biomimetics11020146

[6] Berends, B., Bielevelt, F., Schreurs, R., Vinayahalingam, S., Maal, T., Jong, G.: Fully automated landmarking and facial segmentation on 3d photographs. arXiv preprint arXiv:2309.10472 (2023) https://doi.org/10.48550/arXiv.2309.10472

[7] Zhang, H., Wang, C.-W., Muzakky, H., Dai, J., Li, X., Ma, C., Wu, Q., Cui, X., Xu, K., He, P., Guo, D., Wang, X., Lee, H., Zhong, Z., Zhu, Z., Huang, B.: Deep learning techniques for automatic lateral x-ray cephalometric landmark detection: Is the problem solved? arXiv preprint arXiv:2409.15834 (2024) https://doi.org/10.48550/arXiv.2409.15834

[8] Salama, M.A., Hathout, B.: An ensemble feature ranking technique for the analysis of the craniofacial growth. In: Proceedings of the 2020 IEEE 5th International Conference on Signal and Image Processing (ICSIP). IEEE, ??? (2020). https://doi.org/10.1109/ICSIP49896.2020.9339432

[9] Zhang, J.-N., Lu, H.-P., Hou, J., Wang, Q., Yu, F.-Y., Zhong, C., Huang, C.-Y., Chen, S.: Deep learning-based prediction of mandibular growth trend in children with anterior crossbite using cephalometric radiographs. BMC Oral Health **23**(28), 1–12 (2023) https://doi.org/10.1186/s12903-023-02734-4

[10] Kim, E., Kuroda, Y., Soeda, Y., Koizumi, S., Yamaguchi, T.: Validation of machine learning models for craniofacial growth prediction. Diagnostics **13**(21), 3369 (2023) https://doi.org/10.3390/diagnostics13213369

[11] Myers, M., Brown, M.D., Badirli, S., Eckert, G.J., Johnson, D.H.-M., Turkkahraman, H.: Long-term predictive modelling of the craniofacial complex using machine learning on 2d cephalometric radiographs. International Dental Journal **75**(1), 236–247 (2025) https://doi.org/10.1016/j.identj.2024.12.023

[12] Roseth, J., Kim, J.-H., Moon, J.-H., Ko, D.-Y., Oh, H., Lee, S.-J., Suh, H.: Comparison of individualized facial growth prediction models using artificial intelligence and partial least squares based on the mathews growth collection. Angle Orthodontist **95**(3), 249–258 (2025) https://doi.org/10.2319/082124-687.1

[13] Kale, A., Altun, O.: Face age synthesis: A review on datasets, methods, and open research areas. Pattern Recognition **143**, 109791 (2023) https://doi.org/10.1016/j.patcog.2023.109791

[14] Renoult, J.P., Karpinski, R., Sauvadet, L., Kreyer, M., Mbadoumou, R., Roura-Torres, B., Baniel, A., Charpentier, M.J.E.: Vision transformers for age prediction from facial images in a wild primate. Methods in Ecology and Evolution **16**(12), 2858–2872 (2025) https://doi.org/10.1111/2041-210X.70187

[15] Maroun, G., Bekhouche, S.E., Dornaika, F.: Convnext meets vision transformers: A powerful hybrid framework for facial age estimation. Applied Sciences **16**(7), 3281 (2026) https://doi.org/10.3390/app16073281

[16] Qiu, T., Hu, C., Zhang, J., Wu, F., Wang, H., Liu, X., Sun, M.: Attention-based deep learning for 3d craniofacial soft tissue landmark detection and diagnosis in orthodontics. Scientific Reports **16**, 729 (2025) https://doi.org/10.1038/s41598-025-30383-w

[17] Abdelhady, M.S.: TS Facial Features: Longitudinal Facial Growth Dataset. Kaggle. Accessed: 2026-06-02 (2026)

[18] Cooley, J.W., Tukey, J.W.: An algorithm for the machine calculation of complex fourier series. Mathematics of Computation **19**(90), 297–301 (1965) https://doi.org/10.2307/2003354

[19] Oppenheim, A.V., Schafer, R.W.: Discrete-Time Signal Processing, 2nd edn. Prentice Hall, ??? (1999)

[20] Peng, H., Long, F., Ding, C.: Feature selection based on mutual information criteria. IEEE Transactions on Pattern Analysis and Machine Intelligence **27**(8), 1226–1238 (2005) https://doi.org/10.1109/TPAMI.2005.159

[21] Cover, T.M., Thomas, J.A.: Elements of Information Theory, 2nd edn. Wiley-Interscience, ??? (2006)

[22] Simonyan, K., Vedaldi, A., Zisserman, A.: Deep inside convolutional networks: Visualising image classification models and saliency maps. arXiv preprint arXiv:1312.6034 (2014) https://doi.org/10.48550/arXiv.1312.6034

[23] Spearman, C.: The proof and measurement of association between two things. The American Journal of Psychology **15**(1), 72–101 (1904) https://doi.org/10.2307/1412159

[24] Zhou, H., Zhang, S., Peng, J., Zhang, S., Li, J., Xiong, H., Zhang, W.: Informer: Beyond efficient transformer for long sequence time-series forecasting. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 35, pp. 11106–11115 (2021). https://doi.org/10.1609/aaai.v35i12.17325

[25] Wu, H., Xu, J., Wang, J., Long, M.: Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting. In: Advances in Neural Information Processing Systems (NeurIPS), vol. 34, pp. 22419–22430 (2021)

[26] Hochreiter, S., Schmidhuber, J.: Long short-term memory. Neural Computation **9**(8), 1735–1780 (1997) https://doi.org/10.1162/neco.1997.9.8.1735

[27] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I.: Attention is all you need. In: Advances in Neural Information Processing Systems (NeurIPS), vol. 30 (2017)