SESA-KAN: Simultaneous estimation of sex and age from dental panoramic radiographs

Julian Liang, Igor Borja, David Lima, Márcio Júnior, Patrícia Cury, Luciano Oliveira
Intelligent Vision Research Lab, Federal University of Bahia
Salvador, Bahia, Brazil
{julianliang, igorborja, davidlima, santosmarcio, patricia.cury, lrebouca}@ufba.br

***Abstract*—This paper proposes a novel multi-task learning framework for joint sex classification and age estimation from panoramic dental radiographs. Our proposed method combines masked autoencoders for self-supervised pretraining on large-scale unlabeled data, a Kolmogorov-Arnold network (KAN) to model nonlinear relationships between dental features and demographic labels, and a dynamic logarithmic loss function to balance sex classification and age regression tasks within a Vision Transformer (ViT) architecture. The proposed framework, so called Simultaneous Estimation of Sex and Age via KAN (SESA-KAN), achieved a mean absolute error of 3.39 years for age estimation and an F1-score of 94.2% for sex classification on real-world datasets. Comparative evaluations showed average improvements over existing methods, with a 0.5-year reduction in age estimation error and a 7.5 percentage point increase in sex classification accuracy. The results highlight the effectiveness of integrating self-supervised pretraining, KAN-based feature decomposition, and adaptive task balancing for multi-task medical image analysis. This work advances automated demographic analysis in dental radiography, with potential applications in forensic dentistry.**

***Index Terms*—Multi-task learning, panoramic radiographs, forensic dentistry, sex classification, age estimation.**

I. Introduction

Due to the durability of teeth and jaws under extreme conditions, these human structures serve as reliable sources of information about an individual's age, ancestry, geographic origin, sex, past habits, and pathologies [1]. In situations where documentation is unavailable or unreliable, such as the identification of unaccompanied refugees, individuals undergoing adoption, or bodies in advanced decomposition—facial, panoramic radiographs are reference tools for determining biological sex and age with accuracy [2].

Dental panoramic radiographs are widely used as non-destructive diagnostic tools, enabling the extraction of comprehensive information for constructing biological profiles [3]. However, traditional approaches, such as those based on dental maturity stages [4], face limitations. These include subjectivity in the assessment of dental development stages, exclusion of cases involving missing teeth or dental conditions, and reduced applicability in highly complex forensic scenarios [5]–[7]. This highlights the need for more objective and robust methods capable of addressing variability while ensuring reliable performance across diverse contexts.

In the literature of computational dentistry, some methods have demonstrated extensive applicability in medical image analysis [8]–[13]. Deep learning-based methods offer several advantages, including the simultaneous analysis of a broad range of dental features and their interrelationships, remaining effective in cases of suboptimal dental conditions [14]. In [15], for instance, convolutional neural networks (CNNs) have achieved a 96% reduction in processing time when analyzing radiographic databases.

Moving beyond single-task approaches, multi-task learning (MTL) reduces redundancy, optimizes computational resources, and often improves accuracy and generalization across tasks. These advances have facilitated the development of more adaptable and reliable methodologies in forensic dentistry, particularly through the use of panoramic radiographs [7], [16]. Despite the advances in MTL, its application in forensic dentistry still faces challenges, such as balancing tasks of varying complexity and the need for efficient sharing of representations without compromising the performance of each individual task. The scarcity of annotated and diverse datasets limits the robustness of the solutions developed. Therefore it is essential to investigate strategies that improve the adaptability of models, ensuring the reliability required for clinical and forensic applications.

A. Related work

Table I provides an overview of key state-of-the-art studies that address sex classification and/or age estimation using panoramic radiographs. The table outlines the proposed architectures, dataset characteristics, and the results achieved, offering a comprehensive summary of advancements in this field. Each work listed in the table is subsequently examined and discussed in detail.

***Sex classification**:* Milošević et al. (2019) [7] used the VGG-16 architecture to achieve an accuracy of 96.87%, considering a set of 4,000 dental images of individuals aged 19–85 years. The authors observed a slight decline in performance as patient age increased, attributed in part to the scarcity of samples. Ciconelle et al. (2023) [18] worked with a dataset containing 207,946 radiographs and observed that the algorithm's effectiveness was affected by variables such as resolution, age, and sex, presenting the best results for individuals over 16 years of age, mainly women. Hougaz et al. (2023) [19] used a set of 16,824 panoramic radiographs to train and evaluate models aimed at sex classification,

TABLE I
WORKS ON SEX CLASSIFICATION AND AGE ESTIMATION USING PANORAMIC RADIOGRAPHS IN COMPARISON WITH OURS.

Reference	Architecture	Age range	Data set	Performance	
				Sex	Age (mean absolute error)
[17]	VGG-16	19 - 85	4,000 (images)	ACC = 96.87% ± 0.96%	-
[18]	ResNet	-	207,946 (images)	ACC = 95.02%	-
[19]	EfficientNet V2-L	-	16,824 (images)	F1 = 91.43% ± 0.67	-
[20]	VGG-16	19 - 90	4,035 (images)	-	3.96
[15]	CNN architecture	2 - 89	50,000 (images)	-	3,26 ± 3,06
[21]	EfficientNet-B7	1 - 90	7,666 (images)	-	4.46
[22]	VGG-16	19 - 86	86,495 (teeth)	F1=74.90% ACC = 76.41%	4.94
[23]	EfficientNet-B0	3 - 18	5.132 (images)	ACC = 87,38%	1.96
[24]	VGG-16	7 - 25	4.557 (images)	F1= 86,3%	0,864 ± 1,6
[16]	EfficientNet V2-L	1 - 90	7,666 (images)	F1 = 90.37% ± 0.54	5.66 ± 0.22
Ours	SESA-KAN	1 - 90	7,397 (images)	F1 = 94.21% ± 0.004	3.39 ± 3.21

with a sample distribution of 37.7% male patients and 61.3% female patients. Among the pre-trained models of the EfficientNet family tested, EfficientNet V2-L presented the best performance, achieving an F1-score of 91.43%.

***Age estimation**:* Heinrich et al. (2024) [15] trained a custom CNN on 50,000 radiographs spanning ages from 2 to 89 years, with a focus on its integration into human identification systems. In contrast, Milošević et al. (2022a) [20] evaluated several pre-trained CNNs, such as DenseNet201, ResNet50, and VGG16, using 4,035 panoramic radiographs and 76,416 segmented images of individual teeth (ages ranging from 19 to 90 years). Despite the introduction of an attention mechanism, the VGG16 model achieved the best performance. Liang et al. (2023) [21] explored architectures such as ConvNeXt-V2, Vision Transformer (ViT) and EfficientNet on a dataset of 7,666 radiographs (4,621 women and 3,045 men), having the EfficientNet presented the best results.

***Sex classification and age estimation**:* Recent research adopts a multitasking approach to simultaneously estimate age and sex from panoramic radiographs, optimizing the use of shared information. Milošević et al. (2022b) [22] used a dataset composed of 86,495 images of individual teeth, extracted from 2,899 panoramic radiographs of individuals aged between 19 and 86 years, with a predominance of female samples (59.03%). Despite exploring the multitask approach, the results were inferior to those obtained with single-task models, indicating potential challenges in integrating multiple tasks in this scenario. In a different context, He et al.(2024) [24] analyzed cephalometric radiographs with the VGG16-MultiTask model, but their focus was limited to children and adolescents between 3 and 18 years old. Hirunchavarod et al.(2024) [23] presented DeepToothDuo, which employed an EfficientNet-B0 model pre-trained on ImageNet to analyze 5,132 panoramic radiographs, but targeted exclusively at young Thai individuals, aged between 7 and 25 years old and with permanent teeth. Finally, Prado et al. (2024) [16] investigated an approach based on an additional parameter, applying to four variants of an EfficientNet. The EfficientNetV2-L demonstrated the best performance, standing out in sex classification and age estimation.

B. Contributions

We introduce here a Simultaneous Estimation of Sex and Age in a multi-task architecture trained with a Kolmogorov Arnold Network (SESA-KAN). Our network integrates self-supervised ViT pre-training with Mask Autoencoder (MAE), optimized through a KAN to construct an end-to-end deep learning framework. To achieve the MTL synergy, SESA-KAN incorporates a dynamically weighted adaptive loss function that adjusts task weights in real-time, enhancing learning efficiency during KAN training. The results demonstrate superior performance compared to the only method previously reported in the literature for simultaneous sex and age estimation from panoramic radiographs, evaluated on a dataset with broad age and sex distribution [16].

II. SIMULTANEOUS ESTIMATING SEX AND AGE

SESA-KAN is organized into two main steps, as depicted in Fig. 1. In the first step, a ViT encoder was pre-trained as a MAE. This model learns latent representations that capture essential features of the images, in the absence of labels. In the second stage, the radiographic image is segmented into patches and processed by this pre-trained encoder. The output of this ViT encoder, which consists of a latent feature vector, is then directly connected to a final layer comprised of a KAN, which acts as a decoder model, translating latent representations into final age and sex predictions. Ultimately, this layer takes advantage of the extracted representation to integrate the two tasks into a synergistic model.

A. MAE

The backbone of SESA-KAN leverages a ViT model pre-trained on an unlabeled dataset of dental radiographs, using the MAE [24] self-supervised training protocol. This pre-training aims to achieve meaningful intermediate representation in a latent space of reduced dimensionality, building a better and more efficient encoding of information in the MTL-based ViT encoder for the final sex and age estimation.

In the pre-training phase of SESA-KAN, images are divided into patches, with a random subset of input tokens masked. Encoding is applied exclusively to the visible patches, using

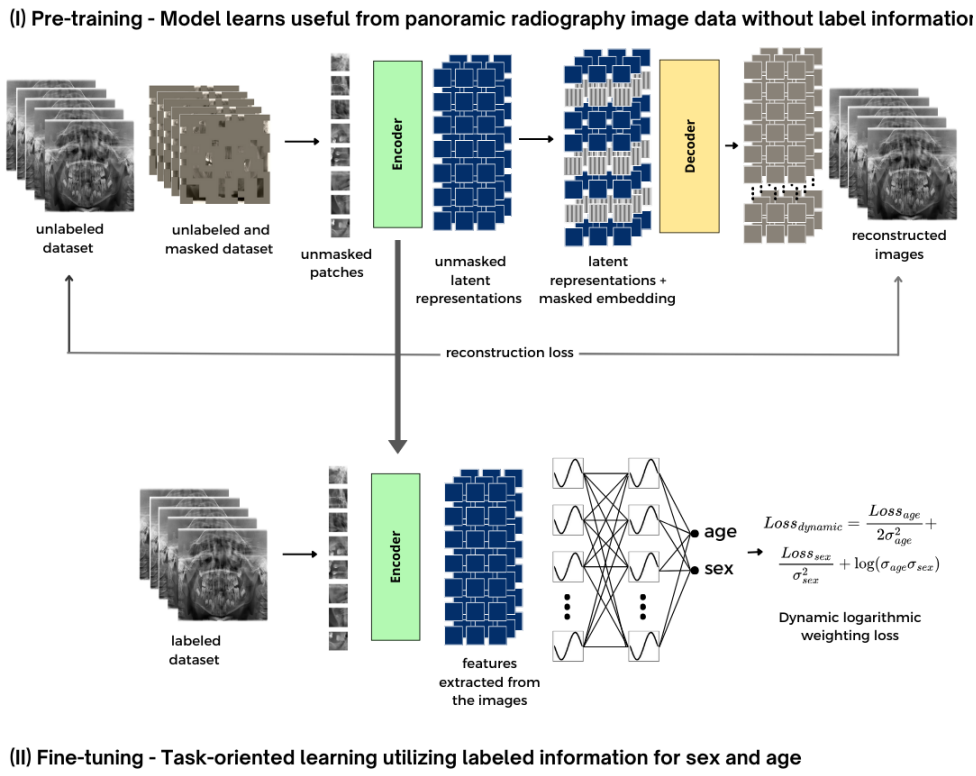


Fig. 1. SESA-KAN architecture. (I) a pre-trained model using dental radiographs and (II) the end-to-end SESA-KAN framework for multi-task age and sex estimation from panoramic radiographs.

Fig. 2. Example of reconstruction of a dental radiograph image. From left to right: original image, masked image, and reconstructed image. The percentage of masked patches was 75%.

ViT blocks to process these tokens and produce a latent representation of the input data. Pre-training is conducted in an unsupervised manner, where the latent representations are processed through a ViT decoder. The reconstruction loss is defined as the mean-squared error (MSE) between the reconstructed and original images.

Figure 2 illustrates the reconstruction process achieved by MAE using a ViT-L(arge)/16 model on a sample from the validation set. The figure displays the original image, the masked version (with 75% of the pixels hidden), and the model-generated reconstruction. The results illustrates that, after pre-trained, MAE captures key structural details, such as tooth contours and essential anatomical features, although certain fine details and contrast variations may not be perfectly recovered.

B. KAN

While multi-layer perceptrons (MLPs) are grounded on the universal approximation theorem (UAT), which asserts that any function from a specific function space, f, can be approximated by a neural network to within an arbitrary error $\epsilon > 0$, given sufficient neurons and a suitable closeness criterion, KANs are based on the Kolmogorov–Arnold representation theorem (KART). The latter states that for $x_i \in [0, 1]$, any multivariate continuous function $f(x_0, x_1, \ldots, x_n)$ can be decomposed into a superposition of continuous 1-dimensional functions such that

$$f(x_0, x_1, \ldots, x_n) = \sum_{q=1}^{2n+1} \Phi_q \left(\sum_{p=1}^{n} \phi_{q,p}(x_p) \right), \quad (1)$$

where $\phi_{q,p} : [0,1] \to \mathbb{R}$ and $\Phi_q : \mathbb{R} \to \mathbb{R}$. This means that instead of relying on approximations of UAT, KART assures that just a polynomial number of univariate functions (Φ_q and $\phi_{q,p}$) should be learned in order to fully represent f. A KAN layer, as defined in [25], with n_{in} input neurons and n_{out} output neurons is defined from a collection of univariate functions ϕ_{ij} - more specifically, the j-th entry of the output vector $\mathbf{y}$ is given by

$$y_j = \sum_{i=1}^{n_{in}} \phi_{ij}(x_i)\,. \quad (2)$$

By stacking Eq. 2, one opens up the possibility of creating deeper and wider networks. Given the input activations $\mathbf{x_l}$, we can define the l-th layer of a KAN as the application of the function matrix $\Phi_l = (\phi_{l,i,j})$:

$$\mathbf{x}_{l+1} = \underbrace{\begin{pmatrix} \phi_{l,1,1}(\cdot) & \phi_{l,1,2}(\cdot) & \cdots & \phi_{l,1,n_l}(\cdot) \\ \phi_{l,2,1}(\cdot) & \phi_{l,2,2}(\cdot) & \cdots & \phi_{l,2,n_l}(\cdot) \\ \vdots & \vdots & & \vdots \\ \phi_{l,n_{l+1},1}(\cdot) & \phi_{l,n_{l+1},2}(\cdot) & \cdots & \phi_{l,n_{l+1},n_l}(\cdot) \end{pmatrix}}_{\Phi_l} \mathbf{x}_l. \quad (3)$$

Thus, we can formally define a L-layered KAN as

$$\mathrm{KAN}(\mathbf{x}) = (\Phi_{L-1} \circ \Phi_{L-2} \circ \cdots \circ \Phi_0)(\mathbf{x}). \quad (4)$$

Originally, $\phi(x) = w_b b(x) + w_s \mathrm{spline}(x)$, where w_b and w_s, are learnable parameters. $b(x)$ is a basis function (in our work, $b(x) = x \cdot \sigma(x)$, where $\sigma(x)$ is the logistic sigmoid), and $\mathrm{spline}(x)$ is a linear combination of B-splines (in our work, splines are cubic), such that

$$\mathrm{spline}(x) = \sum_i c_i B_i(x) \quad (5)$$

where c_i are learnable coefficients. In practice, KANs need fewer input parameters than MLPs while achieving competitive accuracies [25], [26]. Since every function in Eq. 4 is differentiable, KANs can easily be trained with traditional machine learning algorithms such as backpropagation (in our work, we used AdamW [27]).

C. Dynamic logarithmic weighting loss

In MTL, the goal is to optimize multiple loss functions simultaneously, requiring a strategy to combine them into a single value, or to find solutions where all loss functions are optimized together [28]. The most direct way to do that is through a weighted linear sum of the losses of each individual task [29]. The specific losses of each task are summed, resulting in a single scalar loss value. This method presents challenges, mainly due to the sensitivity of the model performance to the choice of weights for each loss. Tuning these hyperparameters can be expensive in terms of computation and time, since the search process requires multiple experiments and manual adjustment of the weights. To tackle this problem, we explored dynamic algorithms capable of automatically adjusting the relative weights of the losses based on task performance, measured by appropriate metrics.

This way, we propose here a variation of the method introduced in [29], an uncertainty-based weighting technique in the context of convolutional neural networks for scene understanding. An additional trainable parameter, σ, was introduced to model the task-dependent uncertainty, known as homoscedastic uncertainty, which is intrinsic to the nature of the task and cannot be reduced simply by increasing the volume of training data. A dynamic logarithmic weighting of the losses was implemented, which allows the adaptive adjustment of the weights between the different classification (sex) and regression (age) tasks. For the KAN decoder weights, the learnable parameters for layer l are the $G + k$ learnable B-spline coefficients $\mathbf{c}^{(l,i,j)} \in \mathbb{R}^{G+k}$ of each edge activation $\phi_{l,i,j}$ - where G is the grid size and k the B-spline order - the weight $w_b^{(l,i,j)}$ of the activation's basis function and the weight $w_s^{(l,i,j)}$ of its associated spline. Therefore, for each layer l with n_l input neurons and n_{l+1} output neurons, we have the collection of learnable weight tensors $\mathbf{W}_l = \{\mathbf{c}^{(l)}, \mathbf{w_s}^{(l)}, \mathbf{w_b}^{(l)}\}$ with $\mathbf{c}^{(l)} \in \mathbb{R}^{n_l \times n_{l+1} \times (G+k)}$ and $\mathbf{w_s}^{(l)}, \mathbf{w_b}^{(l)} \in \mathbb{R}^{n_l \times n_{l+1}}$. Therefore the dynamic weighting can be defined as

$$L(\mathbf{W}_E, \mathbf{W}_0, \ldots, \mathbf{W}_{L-1}, \sigma_1, \sigma_2) = \frac{L_1(\mathbf{W}_E, \mathbf{W}_0, \ldots, \mathbf{W}_{L-1})}{2\sigma_1^2} + \frac{L_2(\mathbf{W}_E, \mathbf{W}_0, \ldots, \mathbf{W}_{L-1})}{\sigma_2^2} + \log(\sigma_1 \sigma_2)\,, \quad (6)$$

$$(7)$$

where $\mathbf{W_E}$ are the weights from the ViT encoder, which is fine-tunned during training, L_1 denotes the loss for the age regression task (managed as the mean squared error) while L_2 is the cross-entropy loss (managed as the F1-score) for the sex classification task, and σ_i is the uncertainty hyperparameter for task i. Essentially, each loss contribution is inversely proportional to the square of the level of intrinsic uncertainty associated to the task, which means that a higher loss has significantly more impact in tasks with low uncertainty and more structure. The final regularization term $\log(\sigma_1 \sigma_2)$ avoids trivial minima that could be achieved as $\sigma_1, \sigma_2 \to +\infty$ and improves convergence.

III. Experimental Analysis

A. Materials

Two distinct datasets of panoramic radiographs were used to train and evaluate the performance of SESA-KAN. Table II summarizes the properties of the datasets used in our work. The first dataset, consisting of 16,824 unlabeled radiographs, was introduced in [19] and employed in our work for self-supervised pre-training of the ViT encoder. This latter dataset

TABLE II
SUMMARY OF THE DATASETS USED IN THE EXPERIMENTAL ANALYSIS.

References	Images	Task	Training and test split
[19]	16,824	Self-supervised pre-training	-
[16]	7,397[a]	Fine-tuning and performance assessment	Training (6.152) and test (1.245)[b]

[a] After a pruning process.
[b] By means of a cross-validation procedure.

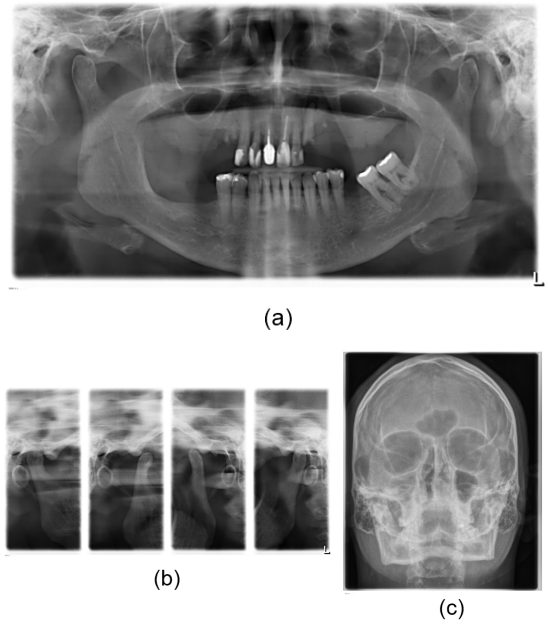


Fig. 3. Examples of images removed after visual analysis: (a) The image is labeled as 6 years old, but does not show deciduous teeth, indicating possible inconsistency; (b) radiograph without adequate coverage of the oral and maxillofacial region; (c) blurred head radiograph, without highlighting the oral and maxillary region;

comprises 6,341 radiographs (37.7%) from male patients and 10,483 (61.3%) from female patients. The second dataset, containing 7,666 radiographs labeled with gender and age information, was introduced in [16]. A pruning process was applied to this dataset to resolve labeling inconsistencies, resulting in the exclusion of 269 images. This process resulted in a final dataset of 7,397 radiographs.

The dataset pruning was conducted by using statistical analysis and visual inspection of the radiographs. The eliminated images included those of low quality, instances with inconsistent age labels, radiographs that failed to meet the criteria for a panoramic image, and those exhibiting excessive noise or significant loss of contrast. Examples of some discarded images are illustrated in Fig. 3.

The pruning process involved training a preliminary model on the dataset of 7,666 images using 5-fold cross-validation (CV). After running the best model obtained in the CV against all images of the dataset, the resulting predictions were analyzed to identify outliers, with residual analysis carried

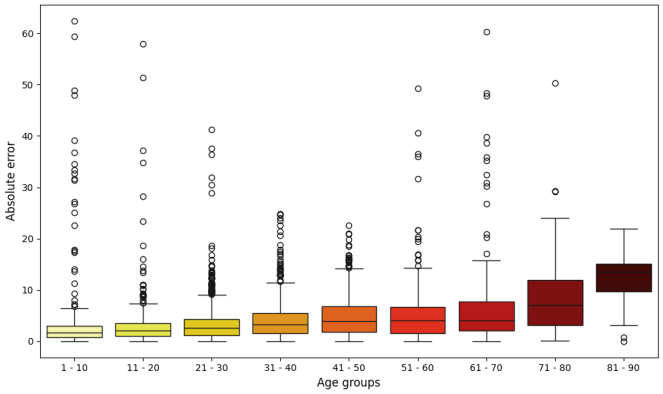


Fig. 4. Segmentation of the original dataset in [16] by age group and absolute error of the early classifier used in the pruning process. A significant number of outliers can be observed up to the 61-70 age range, which highlights the presence of inconsistent samples and the need for a pruning process.

TABLE III
SUMMARY OF THE DATASET OF PANORAMIC RADIOGRAPHS USED ORGANIZED BY AGE GROUP AND SEX.

Age range	Female	Male	Total
1-10	341	339	680
11-20	788	549	1337
21-30	1157	689	1846
31-40	819	505	1324
41-50	615	380	995
51-60	449	250	699
61-70	211	137	348
71-80	74	52	126
81-90	26	16	42
Total	4480	2917	7397

out by age group, according to Fig. 4. The resulting boxplot allowed for the visual inspection and subsequently removal of images with residuals above the third quartile. As a result, 269 images (3.51% of the dataset) were excluded. The pruned dataset included patients aged between 1 and 90 years, with 4,480 images (60.56%) from female patients and 2,917 images (39.45%) from male patients, and an average age of 32.47 years. Age groups above 60 years were underrepresented, a factor to consider during result analysis. Following the pruning process, the dataset was split into a training set with 6,152 images (83.12%) and a test set with 1,245 images (16.88%), reserved for final evaluation using a cross-validation approach. Table III summarizes the pruned panoramic radiograph dataset, detailing the sample distribution across age groups for both male and female patients.

B. Methodology and implementation details

To assess the performance of SESA-KAN, we compared it against two alternative MTL methods: EfficientNetV2-L [16] and ViT-L/16 (MAE) + MLP, which uses the same base architecture as SESA-KAN but substitutes the KAN layer with an MLP. All methods were trained independently using 6,152 images for training and 1,245 images for testing, following a

TABLE IV
BENCHMARK OF THE METHODS CONSIDERING AGE ESTIMATION AND SEX CLASSIFICATION.

Method	Age estimation (mean absolute error)	Sex classification (μF1-score)
EfficientNetV2-L [16]	3.89	86.7%
ViT-L/16 (MAE) + MLP	3.79	93.4%
SESA-KAN	**3.39**	**94.2%**

5-fold cross-validation strategy stratified to maintain proportional sex and age distribution across the dataset.

The experiment based on EfficientNetV2-L followed the methodology outlined in [16], using ImageNet pre-trained weights and fine-tuning them on our pruned dataset. The ViT-L/16 (MAE) + MLP setup was specifically designed to isolate and evaluate the contribution of the KAN layer within the SESA-KAN architecture by replacing it with an MLP layer.

The implementation was carried out using PyTorch. Models were trained with a batch size of 16 on a computational environment comprising a NVidia 4070 GPUs with 12 GB of VRAM, 16 GB of RAM, and an AMD Ryzen 7 processor. As regularization strategies, a weight decay of 0.01 was used to mitigate overfitting by penalizing excessive weights. The training process included 5 warm-up epochs, with an initial learning rate of $1e^{-03}$ and a minimum threshold of $1e^{-06}$.

C. Comparison against other methods

Table IV presents a comparison of the three evaluated methods over the testing dataset (running the 5-fold cross validation procedure). The baseline configuration, based on the EfficientNetV2-L framework described in [16], delivered the weakest performance, with a mean absolute error of 3.89 years for age estimation and an F1-score of 86.7% for gender classification. The inclusion of the self-supervised pre-training strategy resulted in a slight improvement, reducing the error to 3.79 years and increasing the F1-score to 93.4%, highlighting the benefits of using the ViT + MLP model pre-trained with MAE.

The introduction of the KAN layer in the SESA-KAN framework resulted in substantial average performance gains, particularly in age estimation, where the error decreased by 0.5 years, and the F1-score improved by 7.5 percentage points with respect to EfficientNetV2-L. These results underscore the superiority of SESA-KAN, suggesting that the combination of ViT (MAE) with the KAN layer enhances the proposed framework.

D. Task-specific analysis

Figure 5 illustrates the relationship between observed and estimated age by our proposed approach, with the red line representing the ideal line where both ages would be equal. SESA-KAN demonstrated a high correlation between the estimated and the observed ages, with an R^2 of 92.0%. The concentration of most points around the red line highlights the accuracy of the model in generating predictions that are close to the observed values. It is noteworthy a slight dispersion for older ages, which suggests greater variability in the estimates for this age group.

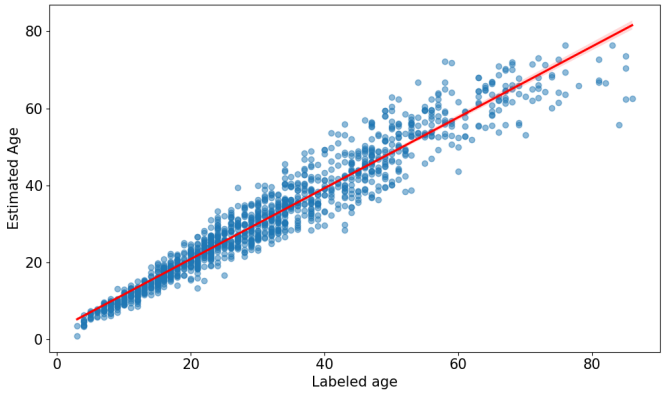


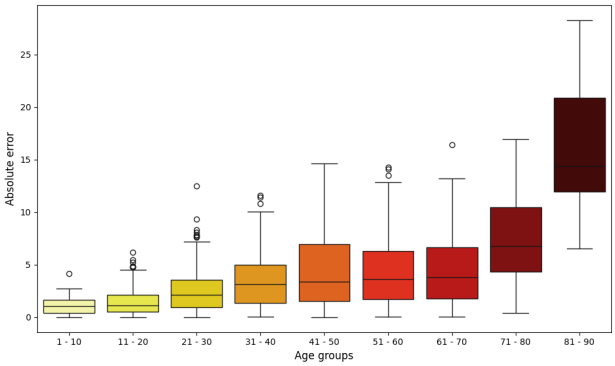
Fig. 5. SESA-KAN's age estimation. The diagonal line represents perfect predictions. While younger samples tend to align closely with the diagonal, older samples exhibit greater variance.

Figure 6(a) presents a summary of the results of the mean absolute error and median by age group on the testing set, highlighting the model's performance in different age groups. The overall mean absolute error was 3.39, with considerable variation between samples. In the groups ranging 1-30, the error was above the average (3.39). It is noteworthy that the worst performance was for ages between 71 and 80, and 81 and 90 years old. The composition of these latter groups is predominantly formed by 126 and 42 training samples (see Table III). This substantially decreased our method's performance in age estimation. Although the increase in mean absolute error with aging is evident in Fig. 6(a), it does not occur in a linear fashion, but with significant jumps in certain age groups. Two significant increases in the mean absolute error were observed during the age transitions. In the transition between 61-70 and 71-80 years, the mean error increased by 2.46 years. The error peak occurred in the transition from 71-80 to 81-90 years, registering an increase of 8.45 years in relation to the previous age group transition.

Figure 6(b) shows the distribution of the absolute error in age predictions for different age groups, highlighting trends and variations in model errors according to age. A significant incidence of outliers is noted in the 11-20 and 21-30 age groups. Furthermore, from the 41-50 age group onward there is a sharp increase in the inter-quartile range, indicating an increase in the variability, in prediction errors. This increase in error can be attributed to a combination of factors, including data bias and the absence of specific dental markers at older ages. After the age of 25, when the formation of permanent teeth is complete, few anatomical changes associated with dental maturation occur. This reduces the availability of indicators to estimate age from radiographs, compromising the accuracy of the model for older age groups [30]. This limitation indicates that the correlation between dental age and chronological age decreases substantially over time, making

Age range	#Samples	μ	y
1-10	77	1.32	1.20
11-20	224	1.76	1.70
21-30	319	2.74	2.66
31-40	247	3.67	3.66
41-50	176	4.64	4.62
51-60	109	4.66	4.66
61-70	61	4.94	5.09
71-80	22	7.40	7.31
81-90	10	15.86	15.71
Total	1245	3.39	2.45

(a)



(b)

Fig. 6. Age estimation distributed by age groups (mean absolute error as μ and the median absolute error as y) (a) and visualization of the mean absolute error (b).

TABLE V
THE RESULTS OF SEX CLASSIFICATION BY AGE GROUP AND BY GENDER IN THE TEST DATABASE.

		F1-score (%)		
Age range	#Samples	Female	Male	Both
1-10	77	76.2	94.7	75.1
11-20	224	97.8	95.1	93.7
21-30	319	99.4	97.1	96.9
31-40	247	99.9	96.3	96.9
41- 50	176	98.4	96.4	95.4
51-60	109	96.9	97.1	94.1
61-70	61	98.6	97.1	96.1
71-80	22	97.6	97.7	95.5
81-90	10	94.7	85.7	83.9
Total	1245	97.7	96.2	94.2

it difficult to differentiate between older individuals. This challenge has also been observed in previous work [14], [20], [21], reinforcing the need for a more balanced dataset and techniques that incorporate other physiological indicators relevant to these age groups.

The performance in sex classification is detailed in Table V, which presents the average F1-score values for SESA-KAN in different age groups, discriminating the results by sex (female and male) and presenting the combination in both sexes. The model obtained an overall F1-score of 94.2%. The best performance was observed in the intermediate age groups, especially between 31 and 40 years old, where the model obtained F1-scores of 99.9% for female samples. While the best performance for male samples was in the 71 to 80 year old range, with an F1-score of 97.7%. In contrast, performance was slightly lower in the extreme age groups, particularly 1–10 years and 81–91 years, where the combined F1-score was 75.1% and 83.9%, respectively, the lowest values among all groups.

Sex classification accuracy followed a pattern distinct from age estimation, with lower performance in the extreme age groups (children, 1–10 years, and elderly, over 80 years), and higher accuracy in adults, corroborating previous studies [16], [31]. In children, the lower efficiency could be attributed to less pronounced morphological differences due to hormonal influence, with more pronounced changes after puberty, especially in females who complete this development earlier [32]. Studies such as in [33] and in [34] point out important morphological criteria in children, such as the pronounced chin, wider and more quadrangular in boys, and the eversion of the gonic region, which is flat in girls and more prominent in boys. At older ages, factors such as physiological changes associated with aging, such as the masculinization of cranial characteristics in women after menopause, and the loss of bone strength in men, make it difficult to distinguish between the sexes [32].

IV. Concluding Remarks

MTL leverages shared representations and inter-task relationships, allowing models to outperform single-task learning by exploiting synergies between related tasks. However, applying MTL in forensic dentistry over panoramic radiographs comes with challenges like balancing tasks of varying complexity and accounting for anatomical variations across different age groups. Additionally, the scarcity of large, well-annotated datasets and the need to model intricate interactions between dental features further complicate implementation of MTL-based architecture. To address these challenges, we introduced the integration of the ViT architecture, KAN layers, and a dynamic logarithmic weighted loss for multitask balancing. This led to the development of SESA-KAN, which achieved a mean absolute error of 3.39 years in age estimation and an F1-score of 94.2% in sex classification.

Future research should investigate the generalizability of the model across diverse populations, as well as the application of interpretability techniques to clinically validate the learned patterns. Moreover, the development of clinical impact studies is essential to quantify potential benefits in diagnostic time, error reduction, and operational costs, thereby supporting the model's adoption in dental and forensic practice.

Acknowledgment

Luciano Oliveira is supported by the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), grant

308580/2021-4.

REFERENCES

[1] G. Berman, M. Bush, P. Bush, A. Freeman, P. Loomis, and R. Miller, "Dental identification," *Manual of Forensic Odontology, Fifth Edition*, pp. 75–128, 1 2013. [Online]. Available: http://www.crcnetbase.com/doi/10.1201/b13744-5

[2] A. Schmeling, R. Dettmeyer, E. Rudolf, V. Vieth, and G. Geserick, "Forensic age estimation: Methods, certainty, and the law," *Deutsches Arzteblatt International*, vol. 113, no. 4, p. 44, jan 2016.

[3] L. Cular, M. Tomaic, M. Subasic, T. Saric, V. Sajkovic, and M. Vodanovic, "Dental age estimation from panoramic X-ray images using statistical models," *International Symposium on Image and Signal Processing and Analysis, ISPA*, pp. 25–30, oct 2017.

[4] A. Demirjian and H. Goldstein, "New systems for dental maturity based on seven and four teeth," *Annals of human biology*, vol. 3, no. 5, pp. 411–421, 1976. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/984727/

[5] W. G. Eckert and N. Garland, "The history of the forensic applications in radiology," *The American journal of forensic medicine and pathology*, vol. 5, pp. 53–56, 1984. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/6369961/

[6] T. Solheim, "Dental age estimation. an alternative technique for tooth sectioning," *The American journal of forensic medicine and pathology*, vol. 5, no. 2, p. 181—184, June 1984. [Online]. Available: http://europepmc.org/abstract/MED/6731411

[7] D. Milošević, M. Vodanović, I. Galić, and M. Subašić, "Estimating biological gender from panoramic dental x-ray images," *International Symposium on Image and Signal Processing and Analysis, ISPA*, vol. 2019-September, p. 105–110, 2019.

[8] G. Jader, J. Fontineli, M. Ruiz, K. Abdalla, M. Pithon, and L. Oliveira, "Deep instance segmentation of teeth in panoramic x-ray images," in *2018 31st SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*, 2018, pp. 400–407.

[9] L. Pinheiro, B. Silva, B. Sobrinho, F. Lima, P. Cury, and L. Oliveira, "Numbering permanent and deciduous teeth via deep instance segmentation in panoramic x-rays," in *17th International Symposium on Medical Information Processing and Analysis*, L. Rittner, E. R. C. M.D., N. Lepore, J. Brieva, and M. G. Linguraru, Eds., vol. 12088, no. 120880C, International Society for Optics and Photonics. SPIE, 2021, pp. 1–10. [Online]. Available: https://doi.org/10.1117/12.2606211

[10] G. Silva, L. Oliveira, and M. Pithon, "Automatic segmenting teeth in X-ray images: Trends, a novel data set, benchmarking and future perspectives," *Expert Systems with Applications*, vol. 107, pp. 15–31, oct 2018.

[11] D. V. Tuzoff, L. N. Tuzova, M. M. Bornstein, A. S. Krasnov, M. A. Kharchenko, S. I. Nikolenko, M. M. Sveshnikov, and G. B. Bednenko, "Tooth detection and numbering in panoramic radiographs using convolutional neural networks," *Dento maxillo facial radiology*, vol. 48, no. 4, 2019. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/30835551/

[12] P. Chu, C. Bo, X. Liang, J. Yang, V. Megalooikonomou, F. Yang, B. Huang, X. Li, and H. Ling, "Using Octuplet Siamese Network for Osteoporosis Analysis on Dental Panoramic Radiographs," *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, vol. 2018-July, pp. 2579–2582, oct 2018.

[13] W. Poedjiastoeti and S. Suebnukarn, "Application of Convolutional Neural Network in the Diagnosis of Jaw Tumors," *Healthcare informatics research*, vol. 24, no. 3, pp. 236–241, jul 2018. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/30109156/

[14] N. Vila-Blanco, M. J. Carreira, P. Varas-Quintana, C. Balsa-Castro, and I. Tomas, "Deep Neural Networks for Chronological Age Estimation from OPG Images," *IEEE Transactions on Medical Imaging*, vol. 39, no. 7, pp. 2374–2384, jul 2020.

[15] A. Heinrich, "Accelerating computer vision-based human identification through the integration of deep learning-based age estimation from 2 to 89 years," *Scientific Reports 2024 14:1*, vol. 14, pp. 1–11, 2 2024. [Online]. Available: https://www.nature.com/articles/s41598-024-54877-1

[16] I. Prado, D. Lima, J. S. Liang, A. Hougaz, B. Peters, and L. R. Oliveira, "Multi-task learning based on log dynamic loss weighting for sex classification and age estimation on panoramic radiographs," in *2024 19th International Conference on Computer Vision Theory and Applications (VISAPP)*, Roma, Italy, 2024.

[17] D. Milosevic, M. Vodanovic, I. Galic, and M. Subasic, "Estimating biological gender from panoramic dental x-ray images," *International Symposium on Image and Signal Processing and Analysis, ISPA*, vol. 2019-September, pp. 105–110, sep 2019.

[18] A. C. M. Ciconelle, R. L. B. da Silva, J. H. Kim, B. A. Rocha, D. G. dos Santos, L. G. R. Vianna, L. G. Gomes Ferreira, V. H. Pereira dos Santos, J. O. Costa, and R. Vicente, "Deep learning for sex determination: Analyzing over 200,000 panoramic radiographs," *Journal of Forensic Sciences*, vol. 68, no. 6, pp. 2057–2064, 2023. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1111/1556-4029.15376

[19] A. Hougaz, D. Lima, B. Peters, P. Cury, and L. Oliveira, "Sex estimation on panoramic dental radiographs: A methodological approach," in *Anais do XXIII Simpósio Brasileiro de Computação Aplicada à Saúde*. Porto Alegre, RS, Brasil: SBC, 2023, pp. 115–125.

[20] D. Milošević, M. Vodanović, I. Galić, and M. Subašić, "A comprehensive exploration of neural networks for forensic analysis of adult single tooth x-ray images," *IEEE*, pp. 70 980–71 002, 2022.

[21] J. Liang, P. Cury, and L. Oliveira, "Revisiting age estimation on panoramic dental images," in *2023 36th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*, 2023, pp. 205–210.

[22] D. Milošević, M. Vodanović, I. Galić, and M. Subašić, "Automated estimation of chronological age from panoramic dental X-ray images using deep learning," *Expert Systems with Applications*, vol. 189, p. 116038, mar 2022.

[23] N. Hirunchavarod, P. Phuphatham, N. Sributsayakarn, N. Prathansap, S. Pornprasertsuk-Damrongsri, V. Jirarattanasopha, and T. Intharah, "Deeptoothduo: Multi-task age-sex estimation and understanding via panoramic radiograph," *2024 IEEE International Symposium on Biomedical Imaging (ISBI)*, pp. 1–5, 5 2024. [Online]. Available: https://ieeexplore.ieee.org/document/10635634/

[24] Y. He, Y. Ji, S. Li, Y. Shen, L. Ye, Z. Li, W. Huang, and Q. Du, "Age and sex estimation in cephalometric radiographs based on multitask convolutional neural networks," *Oral Surgery, Oral Medicine, Oral Pathology and Oral Radiology*, vol. 138, pp. 225–231, 7 2024.

[25] Z. Liu, Y. Wang, S. Vaidya, F. Ruehle, J. Halverson, M. S. Soljači´c, T. Y. Hou, and M. Tegmark, "Kan: Kolmogorov-arnold networks," 4 2024. [Online]. Available: https://arxiv.org/abs/2404.19756v4

[26] S. Sohail, "On training of kolmogorov-arnold networks," 2024. [Online]. Available: https://arxiv.org/abs/2411.05296

[27] I. Loshchilov and F. Hutter, "Decoupled weight decay regularization," 2019. [Online]. Available: https://arxiv.org/abs/1711.05101

[28] O. Sener and V. Koltun, "Multi-task learning as multi-objective optimization," *Advances in Neural Information Processing Systems*, vol. 2018-December, pp. 527–538, 10 2018. [Online]. Available: https://arxiv.org/abs/1810.04650v2

[29] R. Cipolla, Y. Gal, and A. Kendall, "Multi-task learning using uncertainty to weigh losses for scene geometry and semantics," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 7482–7491, 5 2017. [Online]. Available: https://arxiv.org/abs/1705.07115v3

[30] J. L. Prieto, E. Barbería, R. Ortega, and C. Magaña, "Evaluation of chronological age based on third molar development in the spanish population," *International Journal of Legal Medicine*, vol. 119, pp. 349–354, 11 2005. [Online]. Available: https://link.springer.com/article/10.1007/s00414-005-0530-3

[31] I. Ilić, M. Vodanović, and M. Subašić, "Gender estimation from panoramic dental x-ray images using deep convolutional networks," *IEEE*, pp. 1–5, 2019.

[32] U. I. Krenzer, "Compendio de metodos antropologico forenses para la reconstruccion del perfil osteo-biologico cafca," 2006.

[33] J. Cuenca, *Introducción a la antropología forense: análisis e identificación de restos óseos humamos*. Anaconda, 1994. [Online]. Available: https://books.google.com.co/books?id=RU4nMwEACAAJ

[34] H. Schutkowski, "Sex determination of infant and juvenile skeletons: I. morphognostic features," *American Journal of Physical Anthropology*, vol. 90, pp. 199–205, 2 1993.