Multi-Task Learning Based on Log Dynamic Loss Weighting for Sex Classification and Age Estimation on Panoramic Radiographs

Igor Prado, David Lima, Julian Liang, Ana Hougaz, Bernardo Peters and Luciano Oliveira
Intelligent Vision Research Lab, Federal University of Bahia, Brazil
{*igorborja, davidlima, ana.hougaz, bernardo.peters, lrebouca*}*@ufba.br*

Keywords: Multi-Task Learning, Panoramic Radiographs, Sex Classification, Age Estimation.

Abstract: This paper introduces a multi-task learning (MTL) approach for simultaneous sex classification and age estimation in panoramic radiographs, aligning with the tasks pertinent to forensic dentistry. For that, we dynamically optimize the logarithm of the task-specific weights during the loss training. Our results demonstrate the superior performance of our proposed MTL network compared to the individual task-based networks, particularly evident across a diverse data set comprising 7,666 images, spanning ages from 1 to 90 years and encompassing significant sex variability. Our network achieved an F1-score of 90.37%$\pm$0.54 and a mean absolute error of 5.66$\pm$0.22 through a cross-validation assessment procedure, which resulted in a gain of 1.69 percentage points and 1.15 years with respect to the individual sex classification and age estimation procedures. To the best of our knowledge, it is the first successful MTL-based network for these two tasks.

1 INTRODUCTION

In forensic dentistry, specialists provide their proficiency in examining dental records, bite mark analyses, and dental anatomical features to offer support to law enforcement agencies and the judicial system. These experts undertake the task of aligning dental records with unidentifiable remains and scrutinizing bite marks found on victims or objects. Notably, the application of dental radiographs has emerged as an important resource, offering insights, particularly in scenarios involving mass disasters or when conventional identification approaches prove unviable.

In particular, dental radiographs of panoramic type represent a minimally invasive yet highly informative method for extracting pertinent details about an individual. From a single panoramic image, essential information such as sex and age can be deduced, significantly refining the process of narrowing down potential matches and constructing a comprehensive biological profile for unidentified individuals. Recent advancements have witnessed the integration of deep neural networks for the purpose of automating the extraction of information from dental radiographs(Jader et al., 2018; Tuzoff et al., 2019; Silva et al., 2020; Pinheiro et al., 2021; Silva et al., 2023; Hougaz et al., 2023; Liang et al., 2023). However, there remains an unaddressed challenge of effectively integrating the processes related to sex and age estimation within a unified framework, a concern that needs careful investigation and methodological development in the field.

Prevailing approaches often address those tasks as individual networks (Rajee and Mythili, 2021; Ke et al., 2020; Ilić et al., 2019; Hougaz et al., 2023; Milošević et al., 2022a; Vila-Blanco et al., 2020; Liang et al., 2023). However, integrating machine learning techniques, with emphasis on multi-task learning (MTL) models, can substantially enhance operational efficiency by speeding up training and prediction, accuracy, and resource utilization.

Age and sex estimation from panoramic images pose certain challenges. While evaluating young adults with complete dentition simplifies the task, assessing the age and sex of younger and elderly individuals becomes notably intricate, demanding specialized expertise. This complexity arises due to the absence of dimorphic mandibular features in younger individuals, which hinders clear sex identification, and in elderly individuals (Badran et al., 2015), where morphological characteristics used for distinguishing sex and age groups tend to lose reliability over time. Figure 1 illustrates some examples of radiographs used for sex classification and age estimation: Radiographs (a), (c), (e), and (g) belong to females, and (b), (d), (f), and (h) to males. These radiographs are also categorized by age groups: (a) and (b) represent individuals with ages spanning from 0 to 20 years, (c) and (d) are from those aged 21 to 40 years, (e) and (f) to

Table 1: Comparison of works on sex classification and age estimation using panoramic radiographs.

Reference	Year	Task	Pre-processing	Classifier	Data set	Performance
(Ilić et al., 2019)	2019	Sex	Masking	VGG-16	4,155 (images)	ACC = 94.3%
(Ke et al., 2020)	2020	Sex	Image adjustments + GradCAM	VGG-16	19,976 (images)	ACC = $94.6\% \pm 0.58$
(Rajee and Mythili, 2021)	2021	Sex	Image adjustments	ResNet-50	1,000 (images)	F1 $\cong$ 66.55% ACC = 98.27%
(Hougaz et al., 2023)	2023	Sex	GradCAM	EfficientNetV2-Large	16,824 (images)	F1 = $91.43\% \pm 0.67$
(Vila-Blanco et al., 2020)	2020	Age	Image adjustments	DASNet	2,289 (images)	MAE = 2.84 ± 3.75
(Milošević et al., 2022a)	2022	Age	Image adjustments + Augmentation	VGG-16	4,035 (images)	MAE = 3.96
(Liang et al., 2023)	2023	Age	Image adjustments	EfficientNet-B7	7,666 (images)	MAE = 4.46
(Vila-Blanco et al., 2022)	2022	Age and Sex	Bounding box for tooth detection	Faster R-CNN architecture with ResNet-50	1,746 (images)	F1$\cong$91.80% ACC = 91.82% and MAE = 0.97
(Milošević et al., 2022b)	2022	Age and Sex	Bounding box for tooth detection	VGG-16	86,495 (teeth)	F1=74.90% ACC = 76.41% and MAE = 4.94
Ours	2023	Age and Sex	Image adjustments	EfficientNet V2-L	7,666 (images)	F1 = $90.37\% \pm 0.54$ and MAE = 5.66 ± 0.22

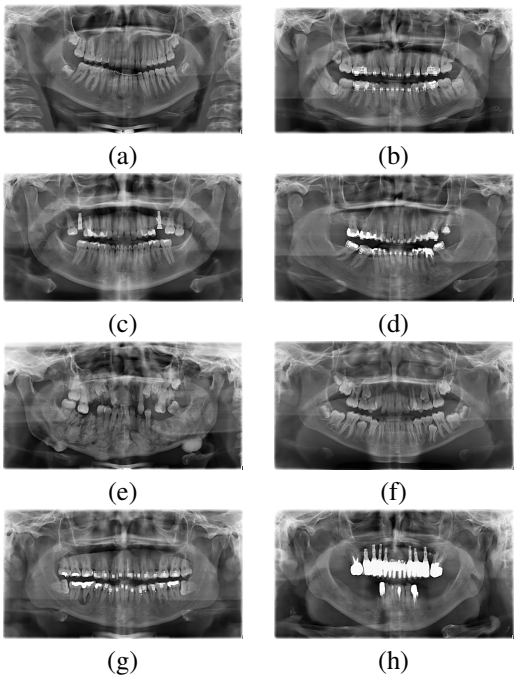

Figure 1: Examples of radiographs distributed by sex and age. Radiographs (a), (c), (e), and (g) belong to females, while (b), (d), (f), and (h) belong to males. These radiographs are also categorized by age groups: (a) and (b) represent individuals aged 0 to 20 years, (c) and (d) are from those aged 21 to 40 years, (e) and (f) to 41 to 60 years age group, and (g) and (h) represent individuals aged from 61 to 80 years. Over time, morphological features that can be used to distinguish between different sex and age groups tend to become less reliable indicators.

41 to 60 years age group, and (g) and (h) individuals aged from 61 to 80 years. While individuals get older, morphological features that can be used to distinguish between different sex and age groups tend to become less reliable indicators.

Taking all of these factors into account, we introduce in this paper a novel method to dynamically weigh the parameters used for each task-based loss in an end-to-end deep network. The goal is to leverage multi-task learning for concurrent sex classification and age estimation. It is noteworthy that the nature of the tasks relies on classification and regression operations, which are concurrent with each other.

1.1 Related Work

Table 1 provides a comprehensive overview of the studies addressing the challenges of sex classification and/or age estimation from panoramic radiographs. The selected columns aim to emphasize key attributes (reference, year, task, pre-processing, classifier, data set, and performance), facilitating an at-a-glance benchmark among these works.

Sex Classification Only. Some studies have developed classification models for sex classification based on dental images. Ilić *et al.* (2019) presented their results with a distribution across age groups, showcasing that the age group between 40 and 50 exhibited the highest accuracy, while the age group over 80 displayed the lowest performance. Wenchi *et al.* (2020) enhanced the accuracy of VGG16 by incorporating a multiple feature fusion (MFF) model; this approach yielded an accuracy of $94.6\% \pm 0.58$ through a modified cross-validation technique on a data set consisting of 19,976 images. Rajee et al. (2021) introduced a gradient-based recursive threshold segmentation (GBRTS) method for segmenting dental radiographic images, which contributed to achieving an accuracy of 98.27% within the age range of 20 to 60 years, even with a considerably smaller data set of 1,000 images. Hougaz *et al.* (2023) explored EfficientNet architectures to perform sex classification, having the EfficientNet V2-L and EfficientNet B0 as the best models via cross-validation, resulting in an accuracy of $91.43\% \pm 0.67$ and $91.30\% \pm 0.47$, respectively, over a data set containing 16,824 images. Only Hougaz *et al.* make a public data set available for the academic community.

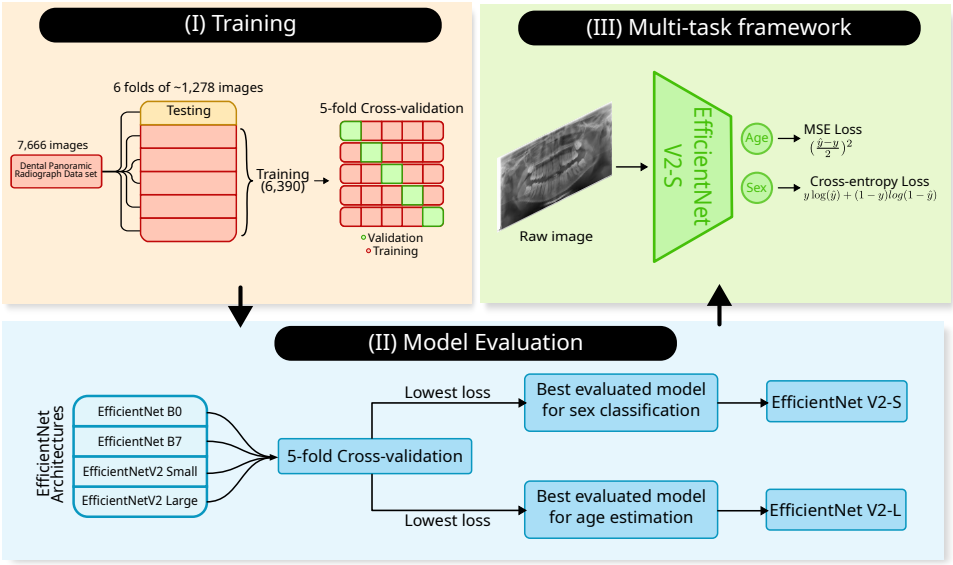

Figure 2: A visual representation of each step taken when creating the multi-task framework. (I) Data division and training method (5-fold Cross-validation), (II) evaluation of four different EfficientNet models, and (III) MTL-based network and loss function optimizations.

Age Estimation Only. Vila-Blanco *et al.* (2020) introduced the DASNet architecture; despite incorporating sex classification in the proposed architecture, the primary objective of the network was to share resources between two backbones to ultimately enhance age estimation; the mean absolute error (MAE) achieved was 2.84 years, over a small data set containing 2,289 images. Milošević *et al.* (2022) conducted six experiments using different pre-trained feature extractors and achieved the best performance with VGG16, using a data set comprising 4,035 images; the final result was an MAE of 3.96 years. Liang et al. (2023) explored the use of the EfficientNet, ConvNeXt, and ViT architectures, employing the largest publicly available data set comprising 7,666 images spanning age ranges from 1 to 90 years; the MAE achieved was 4.46 years using the EfficientNet-B7 architecture. The only available data set is provided in (Liang et al., 2023), which we used for performance assessment in our work.

Sex Classification and Age Estimation. Vila-Blanco *et al.* (2022) proposed tooth detection using a rotated R-CNN to extract oriented bounding boxes for each tooth. Subsequently, the image features within these tooth boxes feed a second CNN module designed to generate age and sex probability distributions per tooth; an uncertainty-aware policy is employed to aggregate these estimated distributions. The proposed approach achieved an MAE of 0.97 years among individuals aged 5 to 25 years and an accuracy of 91.8% in sex classification in the age group ranging from 16 to 60 years over 1,746 images. Although age estimation presents the smallest error compared to other works, it focused only on the age range of young individuals, and the MTL approach did not outperform the single classifiers. Milošević *et al.* (2022a) evaluated different single-task models for both sex classification and age estimation, obtaining an accuracy of 76.41% and an MAE of 4.94, respectively; they experimented with the possibility of a joint model using MTL, but tests yielded results that underperformed in comparison to the single-task models. None of these works provide a public data set for benchmark purposes.

1.2 Contributions

To date, the challenge of concurrently classifying sex and estimating age on dental panoramic radiographs through MTL has remained largely unaddressed. In response, this study introduces a novel method involving dynamic log weighting of each task's loss to optimize an MTL-based end-to-end deep network. The core principle is to optimize parameters that enhance the mean F1-score of classification while minimizing the mean MAE of regression (age estimation). Our findings demonstrate the efficacy of this approach, outperforming the individual networks employed for each task, which has not been achieved by any other work before.

2 OUR PROPOSED APPROACH

Constructing our MTL-based network involved three steps, depicted in Fig. 2. Initially, we gathered a publicly available data set from (Liang et al., 2023), employing the same methodological protocol for performance assessment. Subsequently, these images were partitioned into six folds, with each one encompassing around 1,278 images, totaling 7,666 images. One fold was earmarked for testing purposes, while the remaining five were allocated for training. Training procedures were conducted utilizing a 5-fold cross-validation methodology (see Fig. 2-I).

Four different EfficientNet models were evaluated for sex classification and age estimation, selecting the best-performing model for each task based on its loss. In our experiments, the V2-S architecture yielded the best results for sex classification, while its larger counterpart was the best-performing model for age estimation (see Fig. 2-II). These experiments are detailed in Section 3. Finally, our MTL-based end-to-end network was comprised of an EfficientNet V2-S and two neurons in the last layer, ultimately trained by considering the integration of two types of losses: mean square error (MSE) and cross-entropy (see Fig. 2-III).

2.1 Dynamic Log-Loss Weighting

In traditional MTL approaches to machine learning problems, each loss function corresponding to each task is attributed a weight, and the resulting products are summed to generate the final MTL loss. However, the manual tuning of weights constitutes a time and resource-consuming task, depending on the precise predefined increments in the search process and the number of different tasks. In this context, arises the need for a dynamic algorithm capable of adjusting the relative weights for each loss according to their performance, measured using the relevant metrics.

A novel approach to this problem, although in the context of scene understanding, was described in (Cipolla et al., 2018), where an additional learnable parameter σ is introduced to represent task-dependent (homoscedastic) uncertainty, a type of uncertainty correlated to the task's nature, and thus not necessarily avoidable through increasing amount of training data. In the case where the tasks are sex classification and age estimation, based on a training data set of dental radiographic images, this uncertainty measure can represent the randomness linked to external factors, such as patient's genetics and lifestyle, which can cause the existence of patients with different sex or significantly different age and similar radiographic images in the data set.

For regression tasks, given a model $\mathbf{f}$ with a set of weights $\mathbf{W}$, we define, as in (Cipolla et al., 2018), the likelihood $p(y|\mathbf{f}^{\mathbf{W}}(\mathbf{x}))$ of input image $\mathbf{x}$ having label y as a normal probability distribution on the y variable, with mean equal to the model's output, $\mathbf{f}^{\mathbf{W}}(\mathbf{x})$ and standard deviation equal to the task's homoscedastic uncertainty parameter σ:

$$p(y|\mathbf{f}^{\mathbf{W}}(\mathbf{x})) \sim \mathcal{N}(\mathbf{f}^{\mathbf{W}}(\mathbf{x}), \sigma^2)\,. \tag{1}$$

Therefore, we have that the log-likelihood is given by

$$\log p(y|\mathbf{f}^{\mathbf{W}}(\mathbf{x})) = -\frac{1}{2\sigma^2}(y-\mathbf{f}^{W}(\mathbf{x}))^2 - \log\sigma + \varepsilon\,, \tag{2}$$

where ε is some constant term independent of the model parameters, the images, and the labels. It follows that the expected value of this log-likelihood over a batch of n random samples $(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_n, y_n)$, consisting each of an image and a label, is simply given by

$$\mathbb{E}[\log p(y|\mathbf{f}^{\mathbf{W}}(\mathbf{x}))] = -\frac{1}{2\sigma^2}\cdot\left(\frac{1}{n}\sum_{i=1}^{n}(y_i - \mathbf{f}^{\mathbf{W}}(\mathbf{x}_i))^2\right) - \log\sigma + \varepsilon\,, \tag{3}$$

where $\frac{1}{n}\sum_{i=0}^{n}(y_i - \mathbf{f}^{\mathbf{W}}(\mathbf{x}))^2$ is exactly the MSE.

In a similar manner, the log-likelihood relative to a classification task is given by a scaled version of the logSoftmax of the activations obtained from the last layer of the network. This scaled version uses the task's own uncertainty parameter σ to encode the classifier's confidence in its predictions. Indeed, we have a multivariate probability distribution $p(y|\mathbf{f}^{\mathbf{W}}(\mathbf{x}))$ which can be written as

$$p(y|\mathbf{f}^{\mathbf{W}}(\mathbf{x})) = \log\mathrm{Softmax}\left(\frac{1}{\sigma^2}\mathbf{f}^{\mathbf{W}}(\mathbf{x})\right)\,, \tag{4}$$

where an entry i corresponds to the probability $p(y = i|\mathbf{f}^{\mathbf{W}}(\mathbf{x}))$ that image $\mathbf{x}$ belongs to the i-th class. Therefore, as $\sigma \to \infty$, the maximum difference between the entries of the probability vector $p(y|\mathbf{f}^{\mathbf{W}}(\mathbf{x}))$ tend to 0, which represents that the classifier is completely uncertain of its predictions. Similarly, as $\sigma \to 0$, each probability vector tends to be a one-hot vector representing maximum confidence.

Introducing the approximation $\frac{1}{\sigma}\sum_i \exp(\frac{1}{\sigma^2}\mathbf{f}_\mathbf{i}^{\mathbf{W}}(\mathbf{x})) \approx \left(\sum_i \exp(\mathbf{f}_\mathbf{i}^{\mathbf{W}}(\mathbf{x}))\right)^{\frac{1}{\sigma^2}}$ (Cipolla

et al., 2018) which holds when $\sigma \to 1$, we have that

$$\log p(y=i|\mathbf{f^W}(\mathbf{x})) = \log \frac{\exp(\frac{1}{\sigma^2}\mathbf{f_i^W}(\mathbf{x}))}{\sum_j \exp(\frac{1}{\sigma^2}\mathbf{f_j^W}(\mathbf{x}))}$$
$$\approx \log \frac{\left(\exp(\mathbf{f_i^W}(\mathbf{x}))\right)^{\frac{1}{\sigma^2}}}{\sigma\left(\sum_j \exp(\mathbf{f_j^W}(\mathbf{x}))\right)^{\frac{1}{\sigma^2}}}$$
$$= \frac{1}{\sigma^2}\log \frac{\exp(\mathbf{f_i^W}(\mathbf{x}))}{\sum_j \exp(\mathbf{f_j^W}(\mathbf{x}))} - \log\sigma \quad (5)$$

and therefore

$$\log p(y=i|\mathbf{f^W}(\mathbf{x})) \approx \frac{1}{\sigma^2}\log \frac{\exp(\mathbf{f_i^W}(\mathbf{x}))}{\left(\sum_j \exp(\mathbf{f_i^W}(\mathbf{x}))\right)} - \log\sigma\,. \quad (6)$$

From Eqs. 5 and 6, it is noteworthy that by repeating this process in a batched manner, the first term of the resulting sum will be the cross-entropy loss, assuming as ground truth the sample belongs to the *i*-th class, multiplied by a constant factor of $\frac{1}{\sigma^2}$.

Therefore, when learning simultaneously to classify sex and estimate age from dental radiographic images, the joint probability distribution, assuming independent, separate network outputs for both tasks, will be equal to the product of the individual probability distributions. It follows that the negative log-likelihood loss for this MTL process can be written as

$$L(W,\sigma_1,\sigma_2) = \frac{L_1(W,\sigma_1)}{2\sigma_1^2} + \frac{L_2(W,\sigma_2)}{\sigma_2^2} + \log(\sigma_1\sigma_2)\,, \quad (7)$$

where L_1 denotes the loss for the regression task while L_2 is the loss for the classification task.

In the experimental analysis, we observed that by minimizing the loss from Eq. 7, it takes to a numerically unstable process. Due to the discrete nature of the steps taken in optimization algorithms, such as stochastic gradient descent, it is possible for one of the uncertainty parameters σ_1,σ_2 to reach negative values, which escapes the domain of the logarithm function and makes it impossible to calculate the loss function in the next training step correctly. Because of this instability, we have chosen to train the logarithm of the parameters σ_1,σ_2, which differs from (Cipolla et al., 2018), where the logarithm of the variances σ_1^2,σ_2^2 was trained instead. This solves the domain problem, as the term $\log(\sigma_1\sigma_2)$ becomes a sum of the new trainable parameters $\sigma_1' = \log\sigma_1$ and $\sigma_2' = \log\sigma_2$.

2.2 MTL-Based Network and Relevant Hyperparameters

Our proposed MTL-based deep network is comprised of an EfficientNet architecture with weights pre-trained on the ImageNet data set - and a fully-connected layer on top, mapping the space of features generated by the convolutional neural network to two outputs between 0 and 1, according to Fig. 2-III. The first output node represents the network prediction for the sex attribute, which was chosen as 0 for female and 1 for male. The second output node represents the normalized estimation of the patient's age, in which the interval from 0 to 100 years was linearly mapped to the interval from 0 to 1.

The EfficientNet model used as the backbone was the V2-Small, selected for being the best model according to experiments performed in each task separately, as described in Section 3.2. Additionally, the fully-connected layer, working as a decoder for the features generated by the EfficientNet network, is comprised of 1280 nodes, using dropout with probability $p = 0.2$.

Finally, the initial value of the uncertainty parameters, σ_1,σ_2, were defined to satisfy

$$2\sigma_1^2 = \sigma_2^2 = 1\,, \quad (8)$$

in order to attribute equal initial values to the weights of the individual losses in Eq. 7. Since the network is trained considering the logarithm of the loss weights, it implies the initial values of the uncertainty parameters $\sigma_1' = \log\sigma_1$ and $\sigma_2' = \log\sigma_2$ were $\sigma_1' = -0.5\log 2$ and $\sigma_2' = 0$.

3 EXPERIMENTAL ANALYSIS

3.1 Materials

Data Set. The data set comprises anonymized panoramic radiographs labeled by age and sex, totaling 7,666 panoramic radiographs, available from (Liang et al., 2023). Among these, there are 4,621 samples from women and 3,045 from men. These images have non-standard dimensions due to being captured with different equipment. Consequently, images were resized to match EfficientNet requirements. Notably, this data set does not incorporate any artificial enhancements or synthetic images. This approach was chosen to ensure real-world representation and authenticity. Therefore, the panoramic radiographs in the database exhibit a range of dental conditions, including dental implant placements, cavities, periodontitis, dental plaque, natural tooth loss, and jaw

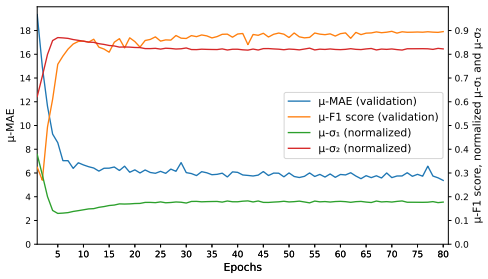

Figure 3: Behavior of the uncertainty parameters during the training with respect to the (normalized) mean F1-score and MAE.

skeletal structure damage. This diversity enables a comprehensive evaluation of the sex classification and age estimation network's performance across various dental issues.

The average age of individuals in the data set is 32.47 years, categorizing our overall sample as adults in terms of dental development. It is worth mentioning that age groups over 60 years old have a notably smaller sample size, which should be considered during the result analysis. The age range within the database spans from 1 to 90 years old, with a distribution of 60.12% women and 39.87% men. Consequently, the sample distribution by age and sex is non-uniform, with more images of younger individuals and a slight majority of women.

The ablation study considered four networks from the EfficientNet series of deep convolutional architectures, aiming to obtain the one with the highest performance to be used in our multi-task learning framework. The rationale was to evaluate the lightest and the heaviest of each version. The choice of the EfficientNet was motivated by its known efficiency (Huang et al., 2019), faster training times, and optimized default hyperparameters, which contribute to shortening the tuning process necessary to achieve better generalization and more robust results over the test data set. Additionally, using the lightest and heaviest models from both EfficientNet versions provides a quantitative way to measure the impact of network size and depth in the performance of the sex classification and age estimation tasks.

3.2 Experimental Methodology

In order to evaluate our MTL-based network and compare its performance with standard single-task neural networks, a series of experiments were conducted, consisting of separately training multiple single-task architectures to classify sex and estimate age of patients from the data set of 7,666 panoramic radiographs. These experiments provided a quantitative comparison between EfficientNet and EfficientNetV2, which aided in the selection of the appropriate architecture for the MTL network. This methodology resulted in 8 separate training processes, obtaining the results shown in Table 2, which details the performance metrics achieved, averaged across the cross-validation process and presented with their respective standard deviations. According to this table, concerning age estimation, it was observed that the B0 and B7 networks exhibited relatively lower efficiency compared to the other architectures. In contrast, the second-generation architectures within the EfficientNet series achieved the best performances. Specifically, V2-L achieved the lowest MAE of 6.34 years, while V2-S obtained an MAE of 6.81 years. In the case of gender classification networks, similar to age estimation, second-generation architectures also outperformed the others. However, the V2-S achieved a higher F1-score of 88.6%, surpassing the V2-L, which achieved an F1-score of 87.5%.

Table 2: Comparison of EfficientNet and EfficientNetV2 architectures on sex classification and age estimation in single-task models. μ and σ denote the mean and standard deviation obtained via a 5-fold cross-validation procedure.

Network	Sex class. (μF1-score$\pm\sigma$)	Age estim. (μMAE$\pm\sigma$)
B0	87.15 ± 0.57	7.63 ± 0.09
B7	85.83 ± 0.87	7.02 ± 0.11
V2-S	88.68 ± 0.58	6.81 ± 0.13
V2-L	87.60 ± 0.15	6.34 ± 0.05

Finally, the best model generated through this pipeline, according simultaneously to F1-score and MAE over the test fold, was used for training our MTL-based network. However, as seen in Table 2, the best models for each task were different, as the EfficientNetV2-Small performed better in the sex classification (with an F1-score of $88.6\% \pm 0.5\%$), and the EfficientNetV2-Large obtained the best MAE, of 6.43 ± 0.05, in the age estimation, with its smaller variation coming second with a MAE of 6.81 ± 0.13. Therefore, considering the performance and cost-benefit analysis of computational resources, V2-S was selected for training our network in the MTL-based framework due to its superiority for the sex classification task, while performing very close to the best V2-L network in age estimation. The V2-S outperformed V2-L by 1.1% in F1-score, and it exhibited only a 0.47-year difference in age estimation MAE compared to V2-L, while the latter had 5.45 times more training parameters than its smaller version.

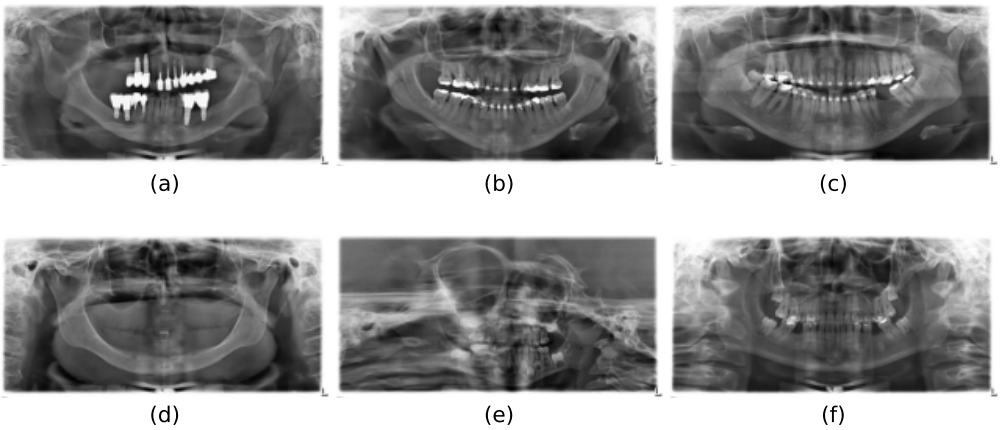

Figure 4: Some of the best, (a), (b) and (c), and worst predictions, (d), (e) and (f), over the test set: (a): Sex: Female. Prediction: Female. Age: 67 years. Prediction: 67.02 years. (b): Sex: Female. Prediction: Female. Age: 37 years. Prediction: 37.98 years. (c): Sex: Female. Prediction: Female. Age: 34 years. Prediction: 34.04 years. (d): Sex: Female. Prediction: Male. Age: 83 years. Prediction: 60.26 years. (e): Sex: Female. Prediction: Male. Age: 2 years. Prediction: 23.70 years. (f): Sex: Female. Prediction: Male. Age: 18 years. Prediction: 30.36 years.

3.3 Result Analysis

Our framework was trained for 80 epochs over the data set of 7,666 panoramic radiographs, which are the same conditions for the single-task experiments described in Section 3.2. The results were an F1-score of 90.37%$\pm$0.54, and a mean absolute error of 5.66$\pm$0.22. This score is the average over the 5 folds defined in the cross-validation procedure, with standard deviation calculated relative to this mean.

As summarized in Table 2, our MTL framework outperforms all selected single-task networks in both tasks of sex classification and age estimation. This demonstrates the effectiveness of sharing task-related knowledge in the training process, the basis principle of MTL, as well as the positive impact that introducing learnable weights for each task's contribution to the loss function had in the final performance. Figure 3 outlines this relationship in more detail, correlating the mean F1-score and MAE over the validation set with the normalized mean uncertainty parameters σ_1, σ_2. The uncertainty parameters' values at each epoch were averaged over all 5 iterations of the cross-validation procedure, and the resulting parameters $\sigma_1^{(mean)}, \sigma_2^{(mean)}$ were then divided by their sum.

It is also possible to observe from Fig. 3 that our MTL-based network converges quickly, since most of the performance metrics (F1-score and MAE) evolution was accomplished in the first few epochs.

3.4 Qualitative Results

Fig. 4 exhibits some of our proposed MTL-based network's best and worst predictions over the test data set via cross-validation. The samples were chosen according to simultaneous performance quality in both tasks. According to this criteria, the best examples were those simultaneously presenting the minimum age difference between ground truth and prediction and that classified the sex label correctly, while the worst examples were those that maximized the age difference and missed the correct sex label.

The images in Fig. 4 suggest the best network performance in both sex classification and age estimation tasks on panoramic radiographs of adults and young adults, especially between the ages 30 and 40, while underperforming in the extreme points of the age range - children and seniors. It also suggests that missing teeth negatively impact network accuracy and age estimation error, as most negative examples register the absence of some teeth.

4 CONCLUDING REMARKS

By learning tasks concurrently, the model can benefit from the complementary nature of tasks, uncovering hidden patterns and relationships that might be missed in single-task learning scenarios. However, it is not always trivial to benefit from MTL in scenarios where the loss of a single task overshadows the loss of the other tasks. If one task has a much higher loss than others, the model might focus too much on minimizing that particular loss, thereby neglecting the other tasks. An alternative in such cases is to search for fixed weights that will result in better metrics for each task. This option has two disadvantages: (i) It needs some search to reach the optimal fixed weights, and

(ii) it does not consider that the losses vary considerably during training. To address these challenges, we proposed dynamically computing weights for losses at every training stage. This method efficiently circumvents the need for intensive parameter searches and adjusts weights in real time, reflecting the evolving nature of training losses. We conducted experiments on a dental panoramic radiograph data set to prove our method's efficiency. Future work includes experimenting with other strategies to integrate features from the component tasks, synergistically.

ACKNOWLEDGMENT

Brazilian National Council for Scientific and Technological Development supported Luciano Oliveira and Igor Prado under grants 308580/2021-4 and 118442/2023-6. Fundação de Apoio à Pesquisa do Estado da Bahia supported Bernardo Silva and David Lima under grants BOL0569/2020 BOL1383/2023.

REFERENCES

Badran, D. H., Othman, D. A., Thnaibat, H. W., and M., A. W. (2015). Predictive accuracy of mandibular ramus flexure as a morphologic indicator of sex dimorphism in jordanians. *International Journal of Morphology*, 33(4):1248–1254.

Cipolla, R., Gal, Y., and Kendall, A. (2018). Multi-task learning using uncertainty to weigh losses for scene geometry and semantics. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7482–7491.

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

Huang, Y., Cheng, Y., Bapna, A., Firat, O., Chen, D., Chen, M., Lee, H., Ngiam, J., Le, Q. V., Wu, Y., and Chen, z. (2019). Gpipe: Efficient training of giant neural networks using pipeline parallelism. In Wallach, H., Larochelle, H., Beygelzimer, A., d'Alché-Buc, F., Fox, E., and Garnett, R., editors, *Advances in Neural Information Processing Systems*, volume 32, pages 1–11. Curran Associates, Inc.

Ilić, I., Vodanović, M., and Subašić, M. (2019). Gender estimation from panoramic dental x-ray images using deep convolutional networks. In *IEEE EUROCON 2019 -18th International Conference on Smart Technologies*, pages 1–5.

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

Ke, W., Fan, F., Liao, P., Lai, Y., Wu, Q., Du, W., Chen, H., Deng, Z., and Zhang, Y. (2020). Biological Gender Estimation from Panoramic Dental X-ray Images Based on Multiple Feature Fusion Model. *Sensing and Imaging*, 21(1):1–11.

Liang, J. S., Cury, P., and Oliveira, L. R. (2023). Revisiting age estimation on panoramic dental images. In *2023 36th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*.

Milošević, D., Vodanović, M., Galić, I., and Subašić, M. (2022a). Automated estimation of chronological age from panoramic dental x-ray images using deep learning. *Expert Systems with Applications*, 189(116038):1–23.

Milošević, D., Vodanović, M., Galić, I., and SubašIć, M. (2022b). A comprehensive exploration of neural networks for forensic analysis of adult single tooth x-ray images. *IEEE Access*, 10:70980–71002.

Pinheiro, L., Silva, B., Sobrinho, B., Lima, F., Cury, P., and Oliveira, L. (2021). Numbering permanent and deciduous teeth via deep instance segmentation in panoramic x-rays. In Rittner, L., M.D., E. R. C., Lepore, N., Brieva, J., and Linguraru, M. G., editors, *17th International Symposium on Medical Information Processing and Analysis*, volume 12088, pages 1–10. International Society for Optics and Photonics, SPIE.

Rajee, M. and Mythili, C. (2021). Gender classification on digital dental x-ray images using deep convolutional neural network. *Biomedical Signal Processing and Control*, 69(102939):1–13.

Silva, B., Pinheiro, L., Oliveira, L., and Pithon, M. (2020). A study on tooth segmentation and numbering using end-to-end deep neural networks. In *2020 33rd SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*, pages 164–171.

Silva, B. P. M., Pinheiro, L. B., Sobrinho, B. P. P., Lima, F. P., Sobrinho, B. P. P., Abdalla Buzar Lima, K., Pithon, M. M., Cury, P. R., and Oliveira, L. R. d. (2023). Boosting research on dental panoramic radiographs: a challenging data set, baselines, and a task central online platform for benchmark. *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, pages 1–21.

Tuzoff, D. V., Tuzova, L. N., Bornstein, M. M., Krasnov, A. S., Kharchenko, M. A., Nikolenko, S. I., Sveshnikov, M. M., and Bednenko, G. B. (2019). Tooth detection and numbering in panoramic radiographs using convolutional neural networks. *Dento maxillo facial radiology*, 48(4):1–10.

Vila-Blanco, N., Carreira, M. J., Varas-Quintana, P., Balsa-Castro, C., and Tomas, I. (2020). Deep Neural Networks for Chronological Age Estimation from OPG Images. *IEEE Transactions on Medical Imaging*, 39(7):2374–2384.

Vila-Blanco, N., Varas-Quintana, P., Ángela Aneiros-Ardao, Tomás, I., and Carreira, M. J. (2022). Xas: Automatic yet explainable age and sex determination by combining imprecise per-tooth predictions. *Computers in Biology and Medicine*, 149:1–10.