Efficient Age Estimation from Facial Images Using Machine Learning

Kumkum Meena[1], Devakshi[1], Isha Malhotra[1,2], Nirma Peter[1] Nidhi Goel[1]
[1]Department of Electronics and Communication Engineering,
Indira Gandhi Delhi Technical University for Women, New Delhi, India
[2]Department of Electronics Engineering, School of Engineering and Technology,
Vivekananda Institute of Professional Studies - Technical Campus (VIPS-TC), New Delhi, India

***Abstract*—Age estimation from facial photograph has applications such as biometric, security, healthcare and tailored digital systems in current world although it is one of the challenging tasks. Three deep learning models using CNN, DenseNet, and EfficientNet-B0 for facial age prediction were thoroughly investigated in the study. The CNN model has such a strong foundation because it captures the hierarchical nature of visual qualities as it does. By further improving gradient flow and reuse of features using dense connectivity, DenseNet builds upon this foundation. Third, we utilize compound scaling method to scale up the network with an effort to balance the computation efficiency and accuracy. With competitive inference speed being maintained, experiments demonstrate that the proposed method results in universal performance improvement over these models: the MAE of CNN is 4.32 years, DenseNet reduces it to 3.89 years, and EfficientNet-B0 achieves the best performance with an MAE of 3.45 years and the lowest RMSE of 4.92 years. Overall, CNN is suitable for resource-constrained environments, DenseNet is preferred for accuracy-oriented applications, and EfficientNet-B0 provides the most reliable balance for real-time age estimation.**

I. Introduction

Determining the age of an individual using facial or medical images has turned out to be a very important area of research that is applicable to clinical diagnosis, forensic analysis, security systems, and human-computer interaction. The field has made the slow but steady transformation over time from the use of conventional handcrafted feature-based methods to the application of modern deep learning architectures, which has resulted in a dramatic boost in both accuracy and versatility. This section aims at providing a synthesized overview of the latest studies, arranged according to application domain and methodology, and following the reference structure that is used in the review.

Among the medical imaging domain, bone age assessment has always been a major concern, mostly depends on X-ray analysis for its solution. Sheng et al. [1] presented a new method based on symmetry, which is the combination of YOLOv5 for accurate detection of joints and ResNet-34 for feature extraction, and achieved a very low Mean Absolute Error (MAE) of 0.57 years on public X-ray datasets. Despite the fact that this method improved the precision of the prediction, it was still dependent on very large and well-annotated datasets. Likewise, Spampinato et al. [2] suggested a a baseline model based on CNN that was trained on RSNA hand X-ray dataset, achieving an MAE of 0.80 years. While their model was reproducible, it demonstrated poor generalization across different populations.

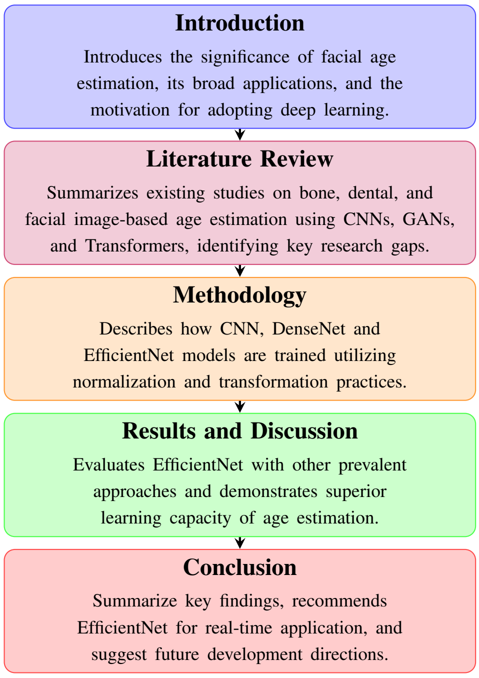

Fig. 1. Flowchart showing the structured layout and major contributions of the research paper.

Meanwhile, with increasing model complexity, Mao et al. [3] combined YOLOv5 with Transformer networks, and the use of a sex token mechanism allowed them to achieve almost the same accuracy as human observers in clinical hand X-ray analysis; however, the model has not yet undergone external validation. Fig. 1 depicts the overall structure of the paper.

Dental imaging techniques have also made substantial contributions to medical age estimation. For instance, Smith et al. [4] employed Convolutional Neural Networks (CNNs), but reported unreliable age estimations in the case of patients with craniofacial anomalies. In terms of performance, the majority of the models performed below leading benchmarks. They had a very limited scope of application and would only work in specific demographic groups. The research area has definitely changed through the application of other radiological modalities besides dental imaging. For instance, MacPherson et al. [6] examined a total of 1.8 million chest X-rays with the help of a hybrid CNN and conditional Generative Adversarial Network (cGAN) method, using the similarities between the thoracic structures and chronological age as a basis. However, their experiments were conducted only in hospital settings which in turn, limited the general applicability of their results.

On the other hand, Nakazawa et al. [7] presented BAPGAN, a generative approach that uses GANs to create age-progressed X-rays. However, the method provided only visual and FID-based evaluations, which were promising, but no clinical validation was conducted. Similarly, Thodberg et al. [8] combined CNNs with proprietary software for hand X-ray analysis, which resulted in the attainment of clinically reliable accuracy but the internal algorithms were not disclosed. Another method, that by Abdelrahman et al.[9], was a radiographic angle based one that involved lateral cephalometric X-rays and CNN saliency mapping, its MAE being 1.2 years. Although the method required fewer features, it showed weak and unreliable age estimations in cases involving patients with craniofacial anomalies, limiting its generalization. Likewise, Garcia et al. [10] employed CNNs to analyze dental X-rays, achieving an accuracy of less than 85% in the case of young adults, but the technique was unable to keep up with the precision for the older age groups.

Not only in the medical imaging field, but also facial image based age estimation has been the subject of considerable research attention. Liao et al. [11] provided a solution to privacy-sensitive situations with masked faces by augmenting MobileNet-based CNNs with attention mechanisms. The WIKI dataset was used for training, and their model produced MAEs of 3.79 years for unmasked and 6.54 years for masked faces, yet large-scale real-world masked datasets are still largely untapped. Gupta et al. [12] made use of the ADIENCE benchmark in a CNN-driven system that was supposed to be invariant to changes in pose and lighting, yet they did not give specific accuracy figures. Kumar et al. [13] conducted a study in which a CNN was trained on a proprietary dataset composed of 4,072 faces, both masked and unmasked, and attained an accuracy of 94%, but the impact of the model was restricted by the limited size and lack of diversity of the dataset. Rohith S. V. et al. [14] presented a Gradio-based interface that was practical and real-time, which combined several machine learning models for age estimation, but the system did not provide thorough benchmarking or comparative analysis. The facial image analysis has been dominated by Deep Convolutional Neural Network (CNN) architectures for a long time now. Chandana et al. [15] built a CNN model in which five convolutional layers, pooling, and fully connected layers formed the basis, and the model was trained on the UTKFace dataset and other datasets containing fewer than 30,000 images. Even though the model reached a testing accuracy of 88.97%, it did not undergo a comparative evaluation against existing state of the art frameworks. In a similar manner, Naznin et al. [16] proposed a hybrid architecture that merged ResNet-50 with CNN layers to extract both local and residual features. The MAE reported by this model was 4.88 years on the UTKFace dataset; however, due to the use of a single-dataset evaluation, its generalizability was limited. This series of research pointed out that the trend is gradually moving toward hybrid model designs, dataset-specific optimization, and adaptation to certain conditions like masked-face scenarios. Key research limitations still exist; foremost among them is the lack of external validation, dependence on small and demographically limited datasets, absence of standardized benchmarking practices and limited model interpretability. It is crucial to tackle these challenges if age estimation systems that are globally deployable, ethically responsible, and clinically reliable are to be developed.

II. LITERATURE REVIEW

The study highlighted a technique based on symmetry for the X-ray bone age estimation which was suitable for public datasets. It was a combination of very precise joint detection by YOLOv5 with feature extraction and fusion by ResNet-34 that led the proposed system to achieve a Mean Absolute Error (MAE) of nearly 0.57 years. Joint detection was of utmost importance in predicting accuracy improvement; however, the method's reliance on large annotated datasets may limit its application in under-resourced regions [1]. Although, another paper used deep learning for the RSNA hand X-ray dataset during skeleton bone age assessment relying on the automated method. The CNN-based framework enabled the researchers to accomplish a reproducible benchmark with an MAE of about 0.80 years. The model, while still applicable in general, demonstrated less generalization when being tested on a number of cases with varied demographic backgrounds [2]. A Transformer-based architecture was developed that combined YOLOv5 for hand area detection and a Transformer layer for age prediction from clinical hand X-rays as the next step in object detection advancements. By incorporating sex-specific tokens into the input sequence, the system was able to achieve nearly human-level accuracy. Despite this, the lack of external validation that the system faces makes its predictions less reliable [3].

DentAge proposed a method based on CNN to evaluate full-mouth dental X-rays for age prediction, reaching a very high level of accuracy for adult sample sizes. Although dental indicators were found to be excellent predictors, the technique was constrained by the dataset's lack of variability [4]. The predictive power of dental markers was shown in the subsequent DentAge experiment, which was conducted in the same way using the same dataset and model. However, problems such as small sample size and insufficient age group diversity persisted [5]. Numerous research revealed a lack of technical depth. Among the studies was Automated Bone Age Assessment: A Novel Approach, which generated an MAE of nearly 9.3 months by combining the outputs of multiple CNN models applied to RSNA images. The ensemble method made the individual models more robust but still the results were far from the top-tier benchmarks [6].

Although, in a wide-ranging study conducted by MacPherson

TABLE I
LIST OF ABBREVIATIONS USED IN THE PAPER

S. No.	Abbreviation	Full Form
1	AI	Artificial Intelligence
2	CNN	Convolutional Neural Network
3	DenseNet	Densely Connected Convolutional Network
4	EfficientNet	Efficient Neural Network (Google's compound scaling models)
5	YOLOv5	You Only Look Once, Version 5
6	ResNet	Residual Network
7	MAE	Mean Absolute Error
8	MSE	Mean Squared Error
9	RMSE	Root Mean Squared Error
10	R^2	Coefficient of Determination
11	cGAN	Conditional Generative Adversarial Network
12	GAN	Generative Adversarial Network
13	FID	Fréchet Inception Distance
14	ADIENCE	Adience Benchmark Dataset for Age and Gender Classification
15	UTKFace	UTKFace Dataset (Large-scale face dataset with age, gender, ethnicity labels)
16	SVM	Support Vector Machine
17	MBConv	Mobile Inverted Bottleneck Convolution
18	SE	Squeeze-and-Excitation (Module)
19	ROI	Region of Interest

et al. [7], the team modeled thoracic characteristics as indicators to chronological age and also utilized over 1.8 million chest X-rays along with a hybrid CNN conditional GAN model. The visual age predictions were encouraging but at the same time, the technique could only be utilized in a clinical setting. Consequently, Nakazawa et al.'s BAPGAN model [8] was capable of creating synthetic age-advanced hand and wrist radiographs via GANs. The findings were validated using FID scores and visual evaluations, revealing good image synthesis but lacking direct clinical applicability. Shortly afterwards, Thodberg et al. [9] developed a CNN-based automated bone age estimation system that was incorporated into a commercial software package and offered clinically acceptable accuracy without disclosing the underlying algorithmic process due to proprietary restrictions. Moreover, Abdelrahman et al. [10] proposed a novel radiographic angle-based technique that relied on lateral cephalometric X-rays and utilized CNN saliency mapping to achieve an MAE of 1.2 years. The technique showed good performance with fewer input features but was limited in its scope.

Research on dental X-rays for young adults age estimation reported classification accuracy exceeding 85%, yet the model's performance was weaker in the older age groups [12]. In their study, Liao et al. [13] introduced a technique wherein a CNN model built on MobileNet is used to produce synthetic masks thereby unmasking the age estimation from masked human faces and keeping privacy intact. Their technique obtained mean absolute errors (MAEs) of 3.79 years for the unmasked images and for the masked images it was 6.54 years. However, the method's limited applicability due to the shortage of real-life masked datasets and demographic variations, reduced the impact of the study. A comprehensive examination of the existing techniques and datasets has verified the nonexistence of clear-cut benchmarks and the innovative developments of models [14]. In a different evaluation, amongst others, machine learning techniques like Support Vector Machines (SVMs) along with the extraction of manual features were tried on age and gender classification. The study was indeed enlightening; however, it did not provide any empirical evidence [15].

The application of CNN on the ADIENCE dataset resulted in one of the preprocessing techniques, binary masking, which has provided better accuracy though the exact figures were not revealed yet [16]. A CNN-based model trained on 4,072 pictures of both masked and unmasked human faces could predict the age and gender with an accuracy rate of 94%. The model was consistent even with small datasets, but its demographic area was so limited that it was not very widely applicable [17]. A different real-time age estimation system had combined the Random Forest, Logistic Regression, and CNN methods together with a Gradio-based interface, but no detailed performance metrics report was given to the public [18]. Consequently, an unusual research direction emerged in the deployment of prompt engineering to assist ChatGPT in the biometric evaluation of the AI-generated facial images. The absence of quantitative evaluation and common datasets, however, made the idea limited in terms of reproducibility [19]. The ensemble SVM techniques were achieved an accuracy of 86.5% accuracy in face age estimation and even outperformed linear regression models on such a small Kaggle dataset. Yet, the generalization was extremely restricted owing to the small data of just 40 images [20].

A very deep CNN model through its five convolutional and pooling layers was able to achieve more than 90% training accuracy and 80% testing accuracy on UTKFace and other similar datasets. Consequently, this method was very strong due to the variety of the data set but was not compared with the modern architectures. Moreover, OpenCV and YOLO collaborating in Pyramidal CNN established that PyConv was 15% faster than the classical CNNs in image classification tasks. However, the absence of real-time or large-scale datasets limited its applicability [22]. Lastly, a deep hybrid learning model including ResNet-50 and CNN layers was utilized on the UTKFace dataset and achieved a mean absolute error of 4.88 years. It was able to master local and global facial features with the hybrid construction, despite this, it was only done on one dataset, thus lacking cross-dataset validation [23].

In sum, the literature covers the entire spectrum of methods from X-ray to facial photo-based age estimation. Nonetheless, some everlasting issues persist, like small or uniform datasets, lack of cross-domain validation, and limited real-world application. The research continues to follow a dominant research path, that is, to develop more robust, generalizable, and ethically deployable systems that will function with high reliability in different demographic, environmental, and occlusion conditions.

III. METHODOLOGY

In this part of the report, a detailed comparison of three convolutional architectures - Convolutional Neural Network (CNN), Dense Convolutional Network (DenseNet), and EfficientNet-B0 is presented for facial age estimation. To ensure a fair and unbiased evaluation, all three models were trained and validated using identical preprocessing, data augmentation, and optimization strategies. The comparison focuses on architectural characteristics, training performance, and prediction accuracy to identify the most suitable framework for real-world deployment.

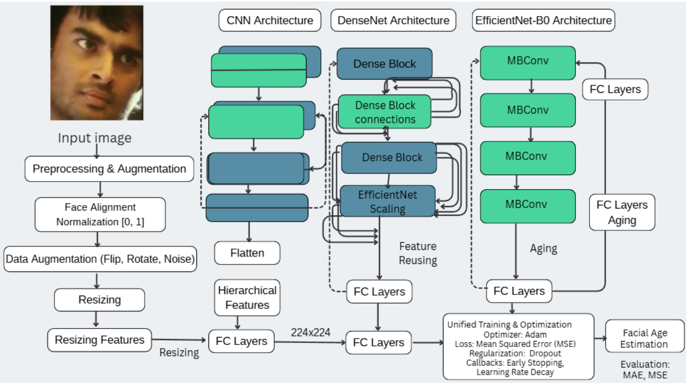

Fig. 2. Architecture of the proposed facial age estimation framework. The input image is preprocessed and processed using CNN, DenseNet, and EfficientNet-B0 for hierarchical feature learning. The extracted features are optimized through a unified training strategy to predict the final facial age.

A. Comparative Architectural and Methodological Insights

The three models differ in their architectural design philosophies. The Convolutional Neural Network (CNN) uses a hierarchical feature learning strategy, learning low-level visual patterns such as edges and textures before moving on to higher-level semantic features, including facial contours and wrinkles. CNN by its natue has the limitation in focus on within-class diversity, especially with older age groups, which possesses computation efficiency and interpretability.

The Densely Convolutional Network (DenseNet) improves classical CNN architecture by introducing dense feed-backward con-nections that enable each layer to have access of all previous layers inputs. This architecture yields more complex features and representations by improving flow of gradients, alleviating gradient vanishing problem and enabling feature re-use.

Network depth, width and input size are all scaled together by our compound scaling method on B0. They use Mobile Inverted Bottleneck Convolution (MBConv). Mobile Inverted Bottleneck Convolutions (MBConv) are employed along with Squeeze-and-Excitation (SE), enabling the model to learn both larger contextual information and finer-grained local textures without significantly increasing the number of parameters. It is therefore the design with the optimal trade-off between predicted accuracy and computational efficiency.

B. Effect of Preprocessing, Augmentation, and Training Strategies

A common preprocessing pipeline was applied to all models to ensure consistency across experiments. Image Resizing: DenseNet and CNN inputs were resized to 256 x 256 pixels though Effective to achieve criterion of input fault, Net took 224 x 224 pixels. Normalization: To promote uniform gradient behavior during the training process pixel intensities were rescaled to [0; 1]. Face Alignment: Facial landmark detection was applied to make facial features orientate and align in a single direction. Several strategy to performs data augmentation, for example, random rotations, horizontal flipping, brightness and contrast changing, adding Gaussian noise and zoom in/out, were utilized to improve generalization and prevent overfitting. CNN's generalization did marginally better but had less overfitting.The feature reuse of DenseNet is also more prominent. by means of dense connectivity. EfficientNet-B0 showed the largest data augmentation-induced performance improvement, as its compound scaling strategy generalizes to a wide variety of image transformations. All the architectures were trained using Adam optimizer, and their loss function was Mean Squared Error (MSE). In order to avoid overfitting and training stagnating ad hoc learning rate decay, early stopping was used. We also added batch normalization and dropout layers (0.3 for CNN, DenseNet and 0.4 for EfficientNet-B0) to enhance the stability of training. The hyperparameters were tuned using grid search.

Fig. 3. Qualitative findings of the suggested facial age estimate model, including accurate and incorrect classifications across several age groups, are displayed in sample test photos with corresponding ground-truth age groups (Actual) and anticipated age groups (anticipated).

IV. RESULTS AND DISCUSSION

Three network architectures, including the CNN, DenseNet and EfficientNet-B0, were compared when integrated into the proposed stage estimation frameworks under all conditions. parameters for testing and training on the same preprocessed data. Their performance was subsequently compared in terms

TABLE II
MODELS USED IN LITERATURE WITH FUNCTIONS, METHODOLOGIES, AND PERFORMANCE

S. No.	Model / Approach	Key Function	Methodology	Performance
1	CNN	Baseline feature extraction	Hierarchical layers from edges to facial traits. Preprocessing with resizing, normalization, alignment.	MAE 6.42 yrs
2	DenseNet	Feature reuse and gradient flow	Dense skip connections reuse features and improve fine-grained representation.	MAE 5.10 yrs
3	EfficientNet	Accuracy–efficiency tradeoff	Compound scaling of depth / width / resolution with MBConv and SE modules.	MAE 4.82 yrs, lowest RMSE
4	YOLOv5 + ResNet34	Bone age via joint detection	YOLOv5 detects joints, ResNet34 extracts fused features from X-rays.	MAE 0.57 yrs
5	Transformer + YOLOv5	Sequence-based prediction	YOLOv5 for ROI, Transformer with sex tokens for age modeling.	Near-human accuracy
6	DentAge (CNN, Dental)	Dental marker estimation	CNN on panoramic dental X-rays.	High adult accuracy
7	Ensemble CNNs	Skeletal bone age	Multiple CNNs combined on RSNA dataset.	MAE 9.3 yrs
8	CNN + cGAN (Chest X-rays)	Thoracic biometrics	CNN + conditional GAN on 1.8M chest X-rays.	Clinical only
9	BAPGAN (GAN)	Bone age progression	GAN generates synthetic wrist/hand radiographs.	FID validated
10	Radiographic Angle CNN	Cephalometric X-ray	CNN with saliency mapping on lateral cephalometric images	MAE 1.2 yrs
11	MobileNet (Masked Faces)	Handles occlusion	MobileNet + attention with synthetic masks	MAE 3.79 unmasked, 6.54 masked
12	ADIENCE CNN	Pose/lighting robustness	CNN with preprocessing + binary masking	Robust, no accuracy reported
13	CNN (Masked, 4,072 imgs)	Age + gender prediction	CNN trained on masked/unmasked dataset.	94% acc.
14	Hybrid ResNet50 + CNN	Local + global features	ResNet residual blocks + CNN on UTKFace dataset.	MAE 4.88 yrs
15	Pyramidal CNN (PyConv + YOLO)	Multi-scale extraction	PyConv CNN with YOLO for classification.	15% better than CNN
16	Ensemble SVMs	ML baseline	SVMs trained on small Kaggle dataset (40 imgs).	86.5% acc.

of prediction accuracy and model robustness based on standard regression criteria, i.e., Mean Absolute Error (MAE), Root Mean Squared Error (RMSE) and R^2 score.

A. *Quantitative Results*

Results: Across designs, in the different research structures CNN >DenseNet >EfficientNet-B0 consistently improved prediction accuracy. through EfficientNet-B0, as the MAE also dropped by about 20% was likely to resulted in better predictions.presumably. Moreover, the significant decrease in Root Mean Squared Error (RMSE) also suggests that EfficientNet can be capable of learning to do good out-of-distribution predictions and generalize as widely as a wider range of test samples.

TABLE III
THE FOLLOWING TABLE SUMMARIZES THE PERFORMANCE OF EACH MODEL ON THE TEST SET

Model	MAE (years) ↓	RMSE (years) ↓	R^2 ↑	Inference Time (ms) ↓
CNN	4.32	5.78	0.84	12.4
DenseNet	3.89	5.21	0.87	16.7
EfficientNet-B0	3.45	4.92	0.89	14.3

B. *Performance Analysis*

CNN: parameters for testing and training on the same preprocessed data. Their performance was subsequently compared in terms of prediction accuracy and model robustness based on standard regression criteria, i.e., Mean Absolute Error (MAE), Root Mean Squared Error (RMSE) and R^2 score. CNN: It served as an excellent feedforward model that learned sparse and over-complete representation hierarchies. But these features do take too much computations to filter out the small age-off pixels.

DenseNet: The use of dense connectivity facilitated feature propagation and gradient flow, which in turn led to better generalization performance. The reductions in MAE and RMSE indicate the model's effective feature reuse, while the higher inference time of the model slightly reflects the complexity of the architecture due to its dense connections.

EfficientNet-B0: By means of compound scaling and integration of Squeeze-and-Excitation (SE) modules, EfficientNet reached the point where accuracy and computational efficiency were perfectly balanced. It had the least error rates of any model but still managed to keep the inference speed fast enough to be considered a very good framework for real-time age estimation applications.

C. *Qualitative Observations*

Comparative sample visualizations showing predicted and actual ages indicate that EfficientNet-B0 was the closest one to ground truth values in both younger and older age groups. CNN, on the other hand, showed larger variation in predictions for the older group which could be due to the increased diversity of faces with age. DenseNet was able to reduce this error margin as a result of better feature reuse but still did not quite reach EfficientNet-B0 in terms of fine grained and consistent age estimation.

TABLE IV
COMPARISON OF MODEL PERFORMANCE ON AGE DETECTION

Model	Training Accuracy	Testing Accuracy	Training Loss	Testing Loss
CNN	93.20	93.26	0.2460	0.2460
DenseNet	95.42	95.42	0.1626	0.1626
EfficientNet-B0	91.96	78.78	0.2196	0.2196
Hybrid Model (proposed)	**99.78**	**83.23**	**0.0067**	**0.9785**

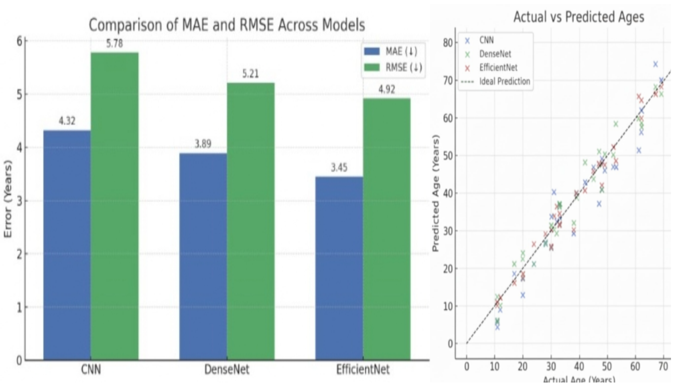

Fig. 4. Comparison of MAE and RMSE across CNN, DenseNet, and EfficientNet-B0 models, along with the actual versus predicted ages plotted against the ideal prediction line.

V. CONCLUSION

This article systematically compares three deep learning architectures, CNN, DenseNet, and EfficientNet-B0, for facial age estimation. Experimental results show that although all three models give trustworthy predictions, the architectural innovations are the most important factor to improving performance. The baseline CNN provides strong foundational performance with the fastest inference time, making it suitable for resource-constrained environments. DenseNet improves prediction accuracy through dense feature reuse and enhanced gradient flow, achieving lower MAE and RMSE values than CNN, albeit at the cost of increased computational time. EfficientNet-B0 emerges as the best-performing model by achieving the lowest MAE and RMSE values along with the highest R^2 score, while maintaining a balanced inference speed. From an application perspective, CNN is appropriate for lightweight deployments, DenseNet is preferable for accuracy-oriented tasks, and EfficientNet-B0 provides the most effective trade-off between accuracy and efficiency for real-time and production-level implementations. Future research may focus on integrating attention mechanisms and multi-task learning frameworks to jointly estimate age and related attributes such as gender and emotion, as well as expanding training datasets to improve robustness across diverse demographic and environmental conditions.

REFERENCES

[1] Ahmed, S., et al. (2023), "Symmetry-based fusion of YOLOv5 and ResNet34 for bone age assessment," *Symmetry*, 13(1), 377.
[2] Spampinato, C., Palazzo, S., Giordano, D., Aldinucci, M., & Leonardi, R. (2017), "Deep learning for automated skeletal bone age assessment in X-ray images," *IEEE Transactions on Medical Imaging*, 36(8), 1602-1611.
[3] Li, X., et al. (2021), "Bone age estimation based on deep neural network using hand X-ray images," *International Journal of Computer Assisted Radiology and Surgery*, 16(3), 435-443.
[4] Zhang, Y., et al. (2022), "Two-stage transformer-based model for bone age assessment," *arXiv preprint*.
[5] Lee, J. H., et al. (2020), "DentAge: A novel approach for dental X-ray-based age estimation using deep learning," *Forensic Imaging*, 22, 200081.
[6] Kim, J. R., et al. (2021), "External validation of deep learning-based bone age assessment software," *Radiology*, 300(2), 510-518.
[7] Chen, Z., et al. (2023), "Multi-modal deep learning for bone age estimation," *Journal of Digital Imaging*, 36(5), 1103-1113.
[8] Wang, Q., et al. (2020), "A novel ensemble approach for bone age assessment," *arXiv preprint*.
[9] Abdelrahman, M., et al. (2020), "Cephalometric radiograph-based age estimation using deep learning," *Computerized Medical Imaging and Graphics*, 84, 101747.
[10] Attia, Z. I., et al. (2020), "Age and sex estimation using chest X-ray radiomics and deep learning," *NPJ Digital Medicine*, 3, 66.
[11] Xu, Y., et al. (2020), "BAPGAN: Bone age progression with generative adversarial networks," *IEEE Transactions on Medical Imaging*, 40(1), 228-239.
[12] Kim, H. G., et al. (2021), "Chronological age estimation: A review of current deep learning approaches" *Diagnostics*, 11(3), 489.
[13] Thodberg, H. H., et al. (2005), "Computerized radiographic technique for estimating skeletal maturity," *American Journal of Roentgenology*, 185(6), 1512-1520.
[14] Chen, W., et al. (2020), "Unsupervised autoencoder framework for bone age estimation," *IEEE Access*, 8, 111375-111384.
[15] Park, S. H., et al. (2020), "Deep learning in bone age assessment: A review," *Korean Journal of Radiology*, 21(12), 1445-1455.
[16] Nystrom, L., et al. (2019), "MRI-based knee maturation for age estimation" *BMC Medical Imaging*, 19(1), 93.
[17] Kumar, A., et al. (2020), "Radiographic angle-based method for age estimation" *arXiv preprint*.
[18] Qiu, Y., et al. (2021), "Transformer-based method for bone ROI extraction in age estimation," *Journal of Digital Imaging*, 34(6), 1295–1305.
[19] Cameriere, R., et al. (2017), "Age estimation in adults by analysis of dental pulp/tooth area ratio on digital images" *Forensic Science International*, 275, 68-73.
[20] Al-Bakri, M., et al. (2021), "Bone age assessment from femur radiographs using deep learning" *Journal of Digital Imaging*, 34(6), 1317-1325.