

Classifying Age-Specific Stages of Cervical Vertebral Maturation Using Deep Feature Analysis

Banu Akkuş*, Shireen Al-Ofary*, Hamza Osman İlhan*, Yasin Yaşa†

*Yıldız Technical University, Istanbul, Turkey

Email: banu.akkus@std.yildiz.edu.tr, shireen.alofari@std.yildiz.edu.tr, hoilhan@yildiz.edu.tr

†Ordu University, Ordu, Turkey

Email: yasinyasa@odu.edu.tr

Abstract—The cervical vertebrae play a critical role in the structural and functional integrity of the human body. Their morphology is closely linked to spinal pathologies, injuries, and developmental conditions. This study investigates the classification of cervical vertebral maturation (CVM) stages using three convolutional neural networks (CNNs): ShuffleNet, AlexNet, and DenseNet. A dataset of 1018 cephalometric radiographs from patients aged 6 to 56 years was analyzed and categorized into three age-based stages. Each image was annotated to ensure accurate labeling of C2, C3, and C4 vertebrae. Features extracted by the CNNs were further processed using Principal Component Analysis (PCA) and classified with Support Vector Machines (SVM). The highest performance was achieved with ShuffleNet features combined with PCA and SVM, reaching an accuracy of 72.69%.

Keywords—Deep Feature Analysis, SVM, PCA, Cervical Vertebrae Maturation (CVM)

I. INTRODUCTION

The timing of orthodontic or orthopedic treatment is critical for addressing craniofacial skeletal abnormalities. An accurate assessment of skeletal maturity helps identify the optimal treatment period. Rather than relying on chronological age, radiographic methods such as cervical vertebral maturation (CVM) staging on lateral cephalometric radiographs offer a more reliable estimate of skeletal development [1]. Numerous studies have validated the reliability of CVM staging for estimating growth potential [2], [3], [4].

Cervical vertebrae are key anatomical structures that support head movement and spinal protection. In CVM analysis, morphological changes in C2, C3, and C4, particularly at their superior and inferior borders, are critical indicators of development. These vertebrae are distinct for their small size and the presence of transverse foramina, which carry blood to the brain. Age estimation holds significant value in forensic science for identifying deceased individuals and ensuring legal clarity on age status [5]. It also plays a vital role in medicine, supporting orthodontic diagnosis and treatment planning during growth spurts, and aiding the detection of metabolic and endocrine disorders. Methods include evaluating skeletal and dental development, as well as secondary sexual characteristics.

The powerful pattern recognition ability of the human brain has inspired artificial intelligence techniques aimed at

automating tasks such as age, gender, and race classification using medical images [6]. Machine learning, especially deep learning, has become an effective tool in estimating age based on bone morphology, and has also been applied for chronological age estimation from lateral cephalometric radiographs [7].

This study uses lateral cervical radiographs to classify vertebral development stages of C2, C3, and C4 in individuals aged 6 to 56 years in three age categories. Three pre-trained CNN models, ShuffleNet, AlexNet, and DenseNet, were used to extract features. These features were refined using Principal Component Analysis (PCA) and classified using both neural networks and Support Vector Machines (SVM), with SVM yielding the most accurate results.

II. RELATED WORK

Atici et al. developed a deep learning approach using directional filters to emphasize vertebral borders in cephalometric X-ray images. Using 1,018 radiographs, their model achieved an accuracy of 84.63% for the classification of five classes of CVM and 75.11% for the classification of six classes [2].

Seo et al. compared six CNN models (ResNet18, MobileNet-v2, ResNet-50/101, Inception-v3, and Inception-ResNet-v2) for CVM stage classification using Grad-CAM for visualization. All models exceeded the accuracy 90%, with Inception-ResNet-v2 showing the highest performance [3].

Kim et al. evaluated an automated bone age assessment system using hand radiographs of 200 patients (ages 3–17). Despite the limited sample size, the system showed 69.5% concordance with the reference ages, supporting the feasibility of deep learning in clinical settings [8].

Rahimi et al. trained a transfer learning model on 890 cephalograms, focusing on vertebrae C2–C4 to classify stages of CVM into three and six categories. Their model achieved 82.8% accuracy for the three-class problem and 61.62% for six-class CVM classification [4].

Al-Ofary and İlhan applied transfer learning to classify colon and lung cancer datasets. They leveraged pre-trained lightweight convolutional neural networks, namely ShuffleNet, AlexNet, and DenseNet. In three separate scenarios, the first step involved using pre-trained models individually to classify the datasets. Subsequently, the features extracted by these

pre-trained networks served as feature extractors, and these extracted features were fed into a Support Vector Machine (SVM) classifier. Lastly, feature selection was performed using Principal Component Analysis (PCA) on the extracted features from the pre-trained models to reduce the number of features and enhance classification performance. Using PCA-based feature selection with pretrained CNNs, the highest classification accuracies—99.93% for colon and 97.92% for lung cancer images—were achieved [9]. This study demonstrated the effectiveness of the deep feature analysis of CNNs in biomedical images.

III. MATERIALS AND METHODS

Three pretrained CNN models were utilized to extract features, which were then classified after significant features were selected using PCA. This process was carried out using a data set comprising cephalometric images of the C2, C3 and C4 cervical vertebrae. Patients were grouped into age categories according to the developmental status of these vertebrae.

A. Dataset Information

In this study, cephalometric X-ray images of cervical vertebrae C2, C3, and C4 were used to classify patients according to their stage of vertebral maturation. The data set was gathered from the Faculty of Dentistry of Ordu University. In addition, the data set consisted of an Excel table containing patient details such as name, age, sex, and vertebrae maturation stage. Ethical approval for this study was granted by the Clinical Research Ethics Committee of Ordu University under Decision No: 58.

The data set comprised three classes (Class 1, Class 2, and Class 3), which were classified according to the maturation stage of the vertebrae. In the CVM literature, six stages (Stage 1–Stage 6) are defined to describe vertebral maturation. However, due to class imbalance and the limited number of samples in certain stages, these six stages were grouped into three broader categories in this study: Stage 1 and Stage 2 were combined as Class 1, Stage 3 and Stage 4 were combined as Class 2, and Stage 5 and Stage 6 were combined as Class 3. This grouping was adopted to ensure a balanced distribution of samples across classes while maintaining clinical relevance in the classification. However, the data set exhibited an imbalance, with Class 1 containing 280 images, Class 2 containing 655 images, and Class 3 containing 2090 images. To rectify this imbalance, the data set was split into five cross-validations, with each class having both training and test sets. The training data set was augmented from Class 1 to Class 3 to ensure a balanced representation in all classes.

The images were already cropped around the vertebral region, focusing exclusively on the regions of C2, C3 and C4. However, the resolution of the images was low, which posed limitations in obtaining high-resolution results. Example images for each class are shown in Fig. 1.

B. Transfer Learning

The study of cervical vertebrae is vital for diagnosing spinal conditions through CT and X-ray imaging, which is essential

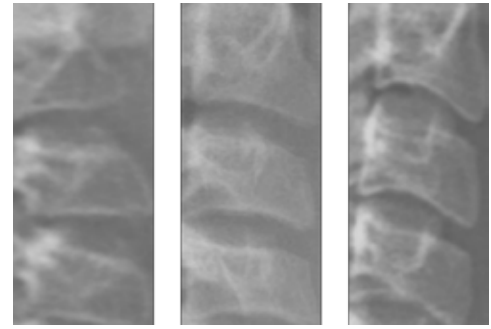


Fig. 1: Example images for each class - 8 years old class 1 (left), 14 years old class 2 (middle), 19 years old class 3 (right)

for fields like physical therapy, neurology, and orthopedics. This research employed three CNN architectures—ShuffleNet, AlexNet, and DenseNet201—to extract vertebral maturation data, utilizing transfer learning with pretrained model weights from other domains.

These models were chosen due to their complementary architectural characteristics: ShuffleNet is known for its computational efficiency and suitability for mobile and embedded devices; AlexNet introduced fundamental deep learning concepts and remains a baseline for CNN performance; and DenseNet201 ensures feature reuse through dense connections, leading to improved information flow and parameter efficiency. By comparing these architectures, the study aimed to explore how model depth and connectivity influence feature extraction and classification performance in medical imaging tasks.

ShuffleNet improves efficiency with group convolution and channel shuffle, AlexNet provides a simple baseline with fundamental CNN concepts, and DenseNet201 enhances feature reuse through dense connections.

These models were selected due to their balance of computational efficiency (ShuffleNet), foundational relevance and historical significance in CNN research (AlexNet), and feature reuse capabilities with low computational cost (DenseNet). This allows a comparison across different network complexities and architectural innovations. For the training process, hyperparameters were set empirically based on preliminary experiments. Specifically, the learning rate was set to 0.001, batch size to 32, and the number of epochs to 100. No automated hyperparameter optimization method such as grid search or Bayesian optimization was applied. These values were selected to ensure model convergence while avoiding overfitting.

C. Principal Components Analysis (PCA)

PCA was applied to reduce the dimensionality of the extracted features while preserving 95% of the variance. This step ensured that essential information was retained and computational complexity reduced before classification.

D. Performance Metrics

Evaluation of the model incorporated several metrics, namely accuracy, precision, recall, and the F1 score. These

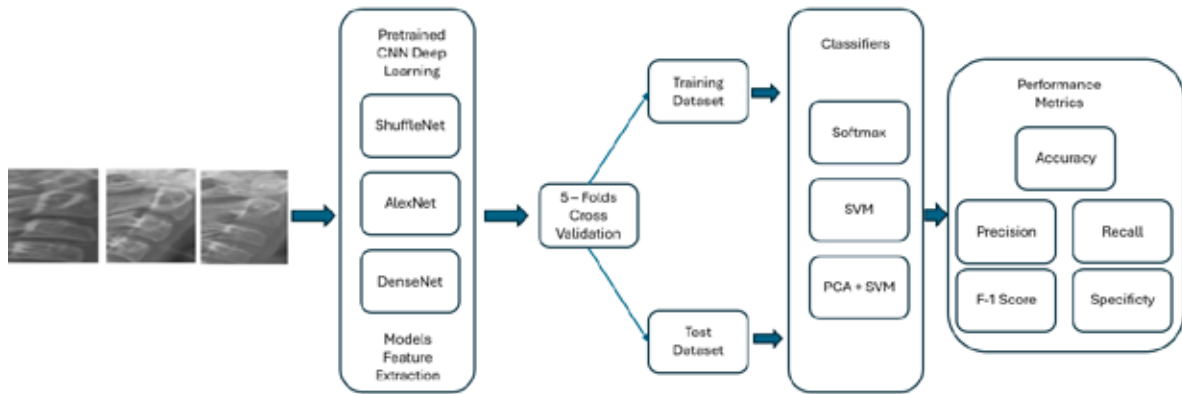


Fig. 2: Flowchart of the study

metrics were derived from an analysis of true positives, true negatives, false positives, and false negatives within the dataset. Notably, the F1 score proves particularly advantageous when it comes to balancing precision against recall, a critical consideration in the comparative assessment of various models.

E. Methodology

In this study, three convolutional neural networks were used to classify a set of cephalometric X-ray images of cervical vertebrae C2, C3, and C4 and categorize a group of patients by their ages based on the development of these vertebrae into three categories. Each category covers a range of ages from 6 years to 56 years.

In the first stage, three convolutional neural networks were used to classify the data set. The features were extracted within these networks. As no modifications were made to the network, features were taken from the last convolutional layer and the data were classified into three categories.

In the second stage, only the features extracted in the previous step were utilized to classify the data differently using classical machine learning, specifically employing the Support Vector Machine (SVM) technique. To compare the results of applying deep learning and classical machine learning, SVM demonstrated its ability to classify rapidly and with high accuracy, particularly given that the features were obtained from convolutional neural networks, renowned for their effectiveness as feature extractors.

In the third step, Principal Component Analysis (PCA) was applied to extract only the significant features from the images. PCA selected the important features from images without affecting the original data by transforming a new dataset containing the important features; this technique helps reduce time, not wasting effort on irrelevant or redundant features in the image and focusing on the most informative ones.

These features were then reused for classification using the SVM technique, and the best results were obtained when this approach was applied in the final scenario. The general flow chart of the study is shown in Figure 2.

IV. EXPERIMENTAL RESULTS

The Google Colab environment was utilized for all experiments with the Python Keras library. Table I presents the accuracy scores of the models under the three scenarios on the cervical vertebrae dataset.

TABLE I: The classification performances of the models under the three scenarios: i) Individual CNN Results, ii) SVM results over CNN derived Features, iii) PCA applied reduced Feature classification results

Accuracy (%)	SoftMax	SVM	SVM + PCA
ShuffleNet	68.32	69.70	72.69
AlexNet	68.73	64.36	71.04
DenseNet	69.14	70.19	71.16

Upon examining the individual performance of pre-trained networks, based on the SoftMax classification results presented in Table I, ShuffleNet, AlexNet, and DenseNet, achieved accuracy scores of 68.32%, 68.73%, and 69.14% for cervical vertebrae classification, respectively. The confusion matrices for the first scenario is presented in Figure 3.

In the second experimental case, the feature extractor was a network, and then the SVM classifier was supplied with the features. The accuracy scores for ShuffleNet, AlexNet and DenseNet are 69.7%, 64.36%, and 70.19%, respectively.

In the last case, uninformative features were eliminated using PCA. Then, similar to the second case, the condensed feature sets were provided to SVM. The highest scores in this scenario were attributed to categorization performance, which achieved accuracy rates of 72.69%, 71.04%, and 71.16%, respectively.

TABLE II: SVM Classification Results of PCA reduced ShuffleNet Features

%	Precision	Recall	F1-Score	Support
Class 1	68.32	69.70	72.69	280
Class 2	68.73	64.36	71.04	655
Class 3	69.14	70.19	71.16	2090
Overall Accuracy	72.69			

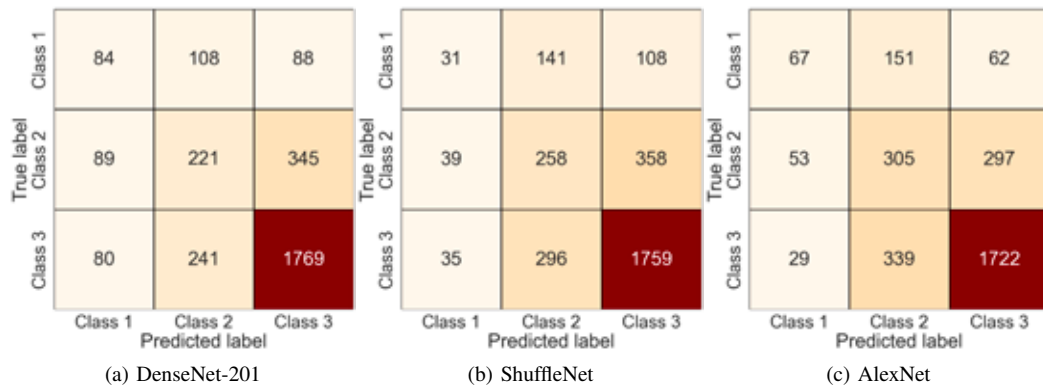


Fig. 3: Confusion matrices of individual network classification results.

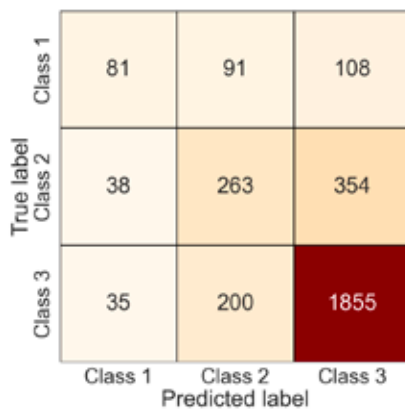


Fig. 4: Confusion Matrix of Best Result (SVM classification over PCA reduced ShuffleNet Features)

Figure 4 presents the confusion matrices of the best outcomes that ShuffleNet obtained using SVM and PCA. Furthermore, Table II lists additional performance metrics, including precision, recall, and F1 scores derived from the corresponding confusion matrix.

V. CONCLUSION

In this study, three deep learning networks—ShuffleNet, AlexNet, and DenseNet—were used to classify the developmental stages of cervical vertebrae (C2, C3, C4) in the cervical spine for a group of patients. The objective was to observe how these vertebrae develop with age. Due to limitations in the number of patients in certain age groups, the age ranges were grouped. These networks were employed as pre-trained models using transfer-learning techniques.

In the first study, convolutional neural networks (CNN) functioned as classifiers for radiographic images of the cervical vertebrae and each model showed strong classification performance, as indicated by SoftMax scores. Then, only the features extracted from the images by the pre-trained models were collected and entered into a Support Vector Machine (SVM) classifier. Finally, Principal Component Analysis (PCA) was applied to the features extracted from the networks to reduce

their dimensionality. Following this feature selection process, the reduced feature sets were provided to the SVM to improve classification accuracy. Remarkably, the final scenario yielded the best results when SVM with PCA was applied, using features extracted by ShuffleNet, reaching an overall accuracy of 72.69% (see Table I and Table II).

These findings highlight that the proposed framework offers not only competitive accuracy but also practical advantages. In particular, the use of ShuffleNet with PCA and SVM demonstrates a lightweight and computationally efficient solution, which is valuable for real-world clinical settings where fast, low-cost, and resource-friendly analysis is required. Future work will fix data imbalance/resolution limits, add tuning and AUC, and test decision-layer classification.

REFERENCES

- [1] E.-G. Kim, I.-S. Oh, J.-E. So, J. Kang, V. N. T. Le, M.-K. Tak, and D.-W. Lee, "Estimating cervical vertebral maturation with a lateral cephalogram using the convolutional neural network," *J. Clin. Med.*, vol. 10, no. 22, pp. 5400, 2021.
- [2] S. F. Atici, R. Ansari, V. Allareddy, O. Suhaym, A. E. Cetin, and M. H. Elnagar, "Fully automated determination of the cervical vertebrae maturation stages using deep learning with directional filters," *PLoS One*, vol. 17, no. 7, pp. e0269198, 2022.
- [3] H. Seo, J. Hwang, T. Jeong, and J. Shin, "Comparison of deep learning models for cervical vertebral maturation stage classification on lateral cephalometric radiographs," *J. Clin. Med.*, vol. 10, no. 16, pp. 3591, 2021.
- [4] H. Mohammad-Rahimi et al., "Deep learning for the classification of cervical maturation degree and pubertal growth spurts: A pilot study," *Korean J. Orthod.*, vol. 52, no. 2, pp. 112–122, 2022.
- [5] A. Schmeling, A. Olze, W. Reisinger, and G. Geserick, "Age estimation of living people undergoing criminal proceedings," *The Lancet*, vol. 358, no. 9276, pp. 89–90, 2001.
- [6] Y.-C. Guo, M. Han, Y. Chi, H. Long, D. Zhang, J. Yang, Y. Yang, T. Chen, and S. Du, "Accurate age classification using manual method and deep convolutional neural network based on orthopantomogram images," *Int. J. Legal Med.*, vol. 135, pp. 1589–1597, 2021.
- [7] N. Liu, "Chronological age estimation of lateral cephalometric radiographs with deep learning," *arXiv preprint arXiv:2101.11805*, 2021.
- [8] J. R. Kim et al., "Computerized bone age estimation using deep learning based program: evaluation of the accuracy and efficiency," *Am. J. Roentgenol.*, vol. 209, no. 6, pp. 1374–1380, 2017.
- [9] S. Al-Ofary and H. O. Ilhan, "Classification of PCA based Reduced Deep Features by SVM for Diagnosing Lung and Colon Cancer," *Proc. 5th Int. Congr. Human-Computer Interact., Optimization Robot. Appl. (HORA)*, pp. 1–5, 2023.