SIIM CMIMI24
Oct 21-22 | Boston, MA
Conference On Machine Intelligence In Medical Imaging

Iterations on a Classic: A Robust Hand Bone Age Algorithm Resistant to Computational Stress

Kelly Horst, MD, Mayo Clinic; Bradley Erickson, MD, PhD, CIIP, FSIIM; Shahriar Faghani, MD

Introduction/Background

Anterior-posterior images of the left hand have been traditionally used to estimate skeletal age for decades. More recently, deep learning (DL) algorithms have been used to estimate skeletal age. The winning algorithm from the 2017 RSNA challenge was recently republished with relatively poor performance when a validation set with varied clinical image appearances was used to test the model under computational stress. We sought to train a model with improved results, with more robust performance to extensive variations in clinical image appearance.

Methods/Intervention

A multimodal DL model was developed and adapted for pediatric bone age assessment. This model includes the DenseNet121 architecture as an imaging feature extractor and two shallow neural networks: one for transforming patient sex into the imaging feature space, and the other for combining all features. The 2017 RSNA bone age dataset was utilized, with 12,611 training images and 1,425 reserved for the validation set. There are 5,778 female and 6,833 male subjects, ranging from 1 to 228 months in age, with a mean bone age of 127.2 months and a standard deviation of 41.7 months. Resizing to 512 x 512, foreground cropping, normalization of image intensities using train set statistics to zero mean and unit standard deviation, random histogram shift, flipping, random affine transformations including rotation, translation and scaling, and Gaussian noise addition were used to enhance model generalizability during training. To perform computational stress test, the same data augmentation pipeline was applied during inference on the validation set. Mean absolute error (MAE) was reported as the performance metric.

Results/Outcome

A MAE of 5.6 months was achieved with a batch size of 16, learning rate of 0.001, and 500 epochs. which is an improvement on the previously published winning model performance of 6.8 months. Importantly, the model achieved this result with extensive variations in clinical image appearance.

Conclusion

A deep neural network can accurately estimate bone age from radiographs of left hand among pediatric patients up to 21 years of age, with robust performance under computational stressors.

Statement of Impact

Training and testing algorithms with computational stress will enhance real world performance. This should be confirmed prospectively with clinical application.

Keywords

Bone age; Computational stress; CNNs