International Journal of Scientific Research and Engineering Development-– Volume 6 Issue 3, May-June 2023

Available at www.ijsred.com

RESEARCH ARTICLE OPEN ACCESS

A Comparative Analysis of various Deep Learning Architectures for Bone Age Detection

Rinzing Namgyal Rai

Department of Computer Sciences and Engineering, Sikkim Manipal Institute of Technology
Sikkim Manipal University, Majitar, East-Sikkim
Email: rinzingnamgyal20@gmail.com

----------**********----------

Abstract:

An accurate assessment of bone age is of vital essence in pediatric radiology for the diagnosis of developmental disorders and evaluation of growth. Recently remarkable performance is shown by deep learning models in various tasks pertaining to image analysis. The purpose of this study is to find differences in learning outcomes between the four Machine Learning Models, namely CNNs (Convolutional Neural Networks), ResNet-50, VGG-19 and Inception v3 to accurately determine the age of the person based on the image of the hand radiograph that it is provided and compare the results of all to determine which is more effective. The primary focus of the study is to evaluate the effectiveness of these deep learning models in accurately predicting bone age from hand radiographs.

.

Keywords **—CNNs (Convolutional Neural Networks), ResNet-50, VGG-19, Inception v3, Neural Network.**

----------**********----------

I. INTRODUCTION

Bone age detection is a task of high importance in pediatric radiology. It provides us with a whole host of valuable data regarding a child's development [1]. Bone age determination is typically performed by trained professionals manually which on top of being time-consuming is further subject to inter-observer variability. With the recent advancements in computer vision and deep learning however, bone age detection using deep learning models has emerged as a promising solution.

The focus of this study is to evaluate how effectively and accurately these deep learning models predict the bone age from hand radiographs. A dataset which comprises of diverse age ranges and populations is used. It is split into training, validation, and testing sets.

The performance for each model is evaluated using various metrics such as mean absolute error and root mean square error. The aim of the comparative analysis is to find the model that achieves the best performance overall in terms of accuracy.

The analysis performed provides valuable insights into the strengths and weaknesses of each model as far as bone age detection is concerned. Moreover, this study is successful in highlighting the potential of deep learning techniques in automated bone age detection.

II. METHODOLOGY

The methodology of this research involved using CNNs (Convolutional Neural Networks), ResNet-50, VGG-19 and Inception v3 to accurately determine the age of the person based on the image of the hand radiograph. The size of the radiographs was reduced to 250x250. The mean absolute error

 ©IJSRED:All Rights are Reserved

and root mean square error (both in months) was calculated to judge the performance of the models.

A. CNNs (Convolutional Neural Networks):

CNNs (Convolutional Neural Networks) are a category of deep learning models which are designed keeping in mind the need to process and analyze visual data, such as images and videos [2]. CNNs have been revolutionary in terms of performing various computer vision activities which include but not limited to image segmentation, image classification and object detection. Convolutional layers make up the fundamental blocks of CNNs. These layers apply a set of learnable filters/kernels to the input data in order to extract all the relevant features. Using a sliding window approach these filters learn the spatial dependencies and patterns within the data. Activation functions, such as ReLU (Rectified Linear Unit) or Softmax usually succeed the convolutional layers which introduces non-linearity in the model thus increasing its expressive power. Pooling layers such as max poolingor average pooling is incorporated in order to reduce the spatial dimensionality of the features that are extracted and assist in capturing large-scale patterns. These layers aggregate the information from neighbouring regions and down-sample the feature maps.

The architecture of a Convolutional Neural Network is often made up of multiple convolutional and pooling layers stacked together which progressively extracts more complex and abstract features. One or more fully connected layers make up the final layers of a CNN which are also referred to as dense layers. These layers then integrate the extracted features to produce the desired result like class probabilities in the case of image classification.

The CNN's training comprises of optimizing their parameters using backpropagation. Backpropagation is a process wherein the weights are repeatedly adjusted using an optimization approach such as Stochastic Gradient Descent. The aim is to minimize the loss function which quantifies the error between the predicted output and the actual output. On top of having outstanding performance in various computer vision tasks by outperforming traditional feature extraction methods, CNNs additionally have also benefited from advancements such as transfer learning. Transfer learning involves leveraging knowledge from models trained on large-scale datasets, such as ImageNet, and fine-tuning them on specific tasks with smaller datasets. This approach makes the training process much faster and improves generalization.

In short, CNNs have become a cornerstone in computer vision research and have found widespread applications across numerous domains, including healthcare, autonomous vehicles, surveillance, and more. The ability to automatically extract meaningful features from visual data has contributed greatly to the advancement of image analysis.

B. ResNet-50:

ResNet-50 is a deep learning architecture which belongs to the family of Residual Neural Networks [3]. They are particularly renowned for their ability to address the vanishing gradient problem and effectively train deep neural networks. The introduction of residual connections called Skip Connections or Shortcut Connections is the key innovation in ResNet-50. These connections enable the network to pass the information directly from the initial layers to later layers and in doing so enable the network to learn the residual mappings as opposed to directly approximating the underlying mapping. This eases the degradation problem. Degradation problem is the problem that occurs when deep network start performing worse as compared to shallow networkswhich are attributed to difficulties in training. Residual Networks include convolutional layers, pooling layers, fully connected layers, and residual blocks. A residual block usually consists of multiple convolutional

 ©IJSRED: All Rights are Reserved

layers along with shortcut connections that might bypass one or more than one layers. The role of the shortcut connections is to facilitate the flow of gradients during training which thus enables effective optimization of the network. ResNet-50 comprises of 50 layers (48 convolutional layers, one MaxPool layer, and one average pool layer).

During training, gradient descent optimization methods such as Stochastic Gradient Descent is used to update the model's weights. Also, techniques like batch normalization are commonly used in order to prevent overfitting and to improve the network's overall generalization.

The deep architecture of ResNet-50 allows it to learn rich and discriminative features which in turn enable state-of-the-art performance on large datasets. Further, ResNet-50 has been widely utilized in transfer learning, where pre-trained models are trained on large datasets and are fine-tuned on certain specific tasks with limited labeled data.

C. VGG-19:

VGG-19 is a deep learning architecture which belongs to the family of Visual Geometry Group models [4]. It is mostly known for the simplicity it offers and its uniformity in design which makes it easy to understand and implement. The key characteristic of VGG-19 is its emphasis on using a series of stacked convolutional layers with small receptive fields (3x3 filters) along with max pooling layers for downsampling. The architecture consists of 19 layers (16 convolution layers, 3 Fully connected layer, 5 MaxPool layers and 1 SoftMax layer). The convolutional layers are designed to enable the learning of increasingly complex features, to capture local patterns, edges, and textures in the input images. The pooling layers help in reducing spatial dimensions, allowing the network to focus on more high-level features.

During training, gradient descent optimization methods such as Stochastic Gradient Descent is used to update the model's weights. Regularization techniques like dropout and weight decay are commonly used to prevent overfitting and improve generalization. The simplicity and uniformity of VGG-19's architecture makes it easy to understand and implement.

The architecture of VGG-19 leads to many parameters compared to other models. However, it also offers benefits like improved interpretability and transferability. The uniformity in its design facilitates seamless transfer learning by leveraging pre-trained models on large-scale datasets like ImageNet and fine-tuning them for specific tasks with limited labeled data.

D. Inception v3:

Inception V3 is a deep learning model based on Convolutional Neural Networks, which is used for image classification [5]. The inception V3 is a superior version of the basic model Inception V1 which was introduced as GoogLeNet in 2014. As the name suggests it was developed by a team at Google.

Introduction of the inception module, which facilitates the parallel extraction of features at different scales is the key innovation in Inception v3. The inception module comprises of multiple convolutional layers with different filter sizes (1x1, 3x3, and 5x5) and a max pooling layer (3x3). By combining filters of different sizes, the model can capture both local and global features, enabling it to learn more comprehensive representations.

Inception v3 also makes use of other techniques to enhance its performance. Most notably it uses factorized convolutions. Factorized Convolutions is where large filters are decomposed into small ones in order to reduce the computational complexity. It also incorporates batch normalization, which normalizes the input data within each mini-batch and applies regularization techniques like dropout to prevent overfitting.

 ©IJSRED: All Rights are Reserved

During training, gradient descent optimization methods such as Stochastic Gradient Descent is used to update the model's weights. Transfer learning is also commonly used with Inception v3.

Overall, Inception v3's architecture demonstrates the effectiveness of the inception module and other optimization techniques in improving the performance of deep learning models for computer vision tasks. Its combination of multi-scale feature extraction, factorized convolutions, and regularization techniques has contributed to its success and made it a valuable tool in the field of computer vision.

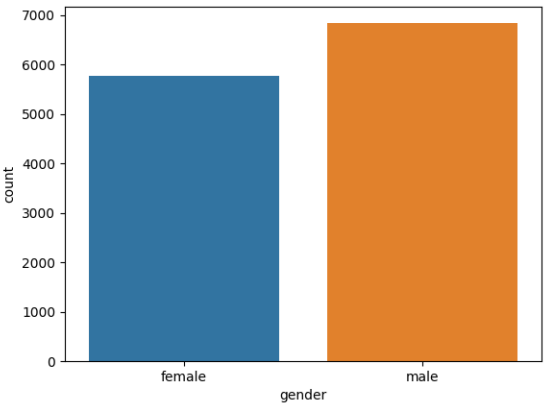

Fig. 1Total Count of Males and Females in the dataset

E. Dataset:

The training dataset consists of 12611 images of hand radiographs and the test dataset comprises of 200 images. The datasets have been obtained from Radiological Society of North America (RSNA). The training dataset was further split into training and validation sets in a 80 – 20 split which resulted in 10088 training images and 2523 validation images.

Upon performing analysis of the dataset it was found that there were 6833 images of male radiographs and 5778 images of female radiographs. The maximum age in the dataset was 228 months while the minimum age was 1 month. The mean age was 127.3207517246848 months and the median age was 132.0 months.

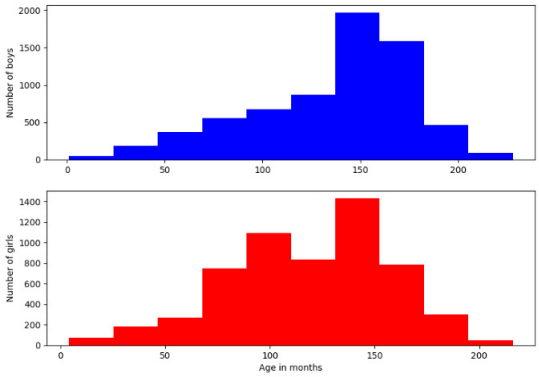

Fig.2 Age in Months vs Number of Males/Females

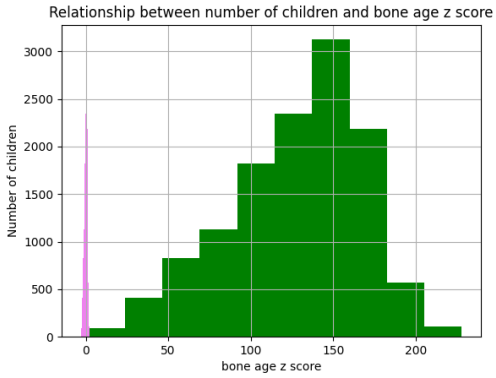

Fig.3 Bone Age Z score vs Number of Children

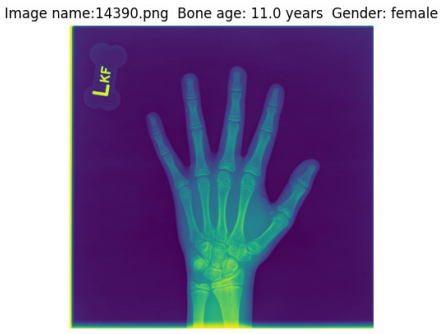

Fig.4Sample Image in dataset

III. RESULTS

In this section, we present the results of the bone age detection as obtained by the four models. The assessment of the models was done based the model's Mean Absolute Error and Root Mean Square Error. The Mean Absolute Error and Root Mean Square Error were calculated in months.

For the first model, i.e., a CNN the results obtained were as follows:

Mean Absolute Error: 41.7664
Root Mean Square Error: 42.15439755196913

For the second model, i.e., ResNet-50 the results obtained were as follows:

Mean Absolute Error: 33.9984
Root Mean Square Error: 44.62955851514074

For the third model, i.e., VGG-19 the results obtained were as follows:

Mean Absolute Error: 26.2518
Root Mean Square Error: 51.16760444236032

For the final model, i.e., Inception v3 the results obtained were as follows:

Mean Absolute Error: 26.3930
Root Mean Square Error: 53.88215358623807

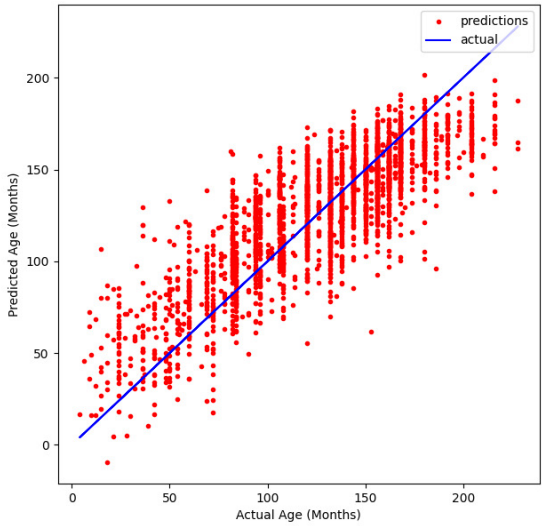

Fig.5 Performance of Inception v3

IV. CONCLUSIONS

In this Analysis we see that the models perform in the order of:

CNN < ResNet-50 < Inception v3 < VGG-19

in terms of Mean Absolute Error. VGG-19 gives us the lowest value i.e., 26.2518 while a standard CNN gives us the highest i.e., 41.7664.

In terms of Root Mean Square Error however, we see that the CNN gives us the lowest value i.e., 42.15439755196913 while Inception v3 gives us the highest value i.e., 53.88215358623807. The order of the performance of the models in this case is:

Inception v3 < VGG-19 < ResNet-50 < CNN

TABLE I
PERFORMANCE OF THE MODELS

Sl. No	Model	Mean Absolute Error	Root Mean Squared Error
1	CNN	41.7664	42.1543
2	ResNet-50	33.9984	44.6295
3	VGG-19	26.2518	51.1676
4	Inception v3	26.3930	53.8821

We therefore see that CNN and Inception v3 perform the worst among the four in terms of Mean Absolute Error and Root Mean Squared Error respectively. CNN does perform the best in terms of Root Mean Squared Error however its relatively high Mean Absolute Error means it is not generalizing as expected. Inception v3 performs relatively better in terms of Mean Absolute Error but its high Root Mean Squared Error means it too is not generalizing as expected.

VGG-19 is the best performing model in terms of Mean Absolute Error and its Root Mean Squared Error isn't too terrible. ResNet-50 however, shows the most promising results as it shows a good balance across the board. It ranks 3rd and 2nd respectively in terms of Mean Absolute Error and Root Mean Squared Error.

ACKNOWLEDGMENT

I would like to express my sincere gratitude to the Department of Computer Science and Engineering at Sikkim Manipal Institute of Technology for their immense support during the course of this research.

I would also like to thank the Radiological Society of North America (RSNA) for providing the dataset on which the analysis was performed. Without their efforts to create the dataset this project would not have been possible.

Finally, I would like to thank the open-source community for providing me with the necessary tools and resources to conduct this research. Without their contributions, this project would not have been possible.

REFERENCES

[1] X. Ren et al., "Regression Convolutional Neural Network for Automated Pediatric Bone Age Assessment From Hand Radiograph," in IEEE Journal of Biomedical and Health Informatics, vol. 23, no. 5, pp. 2030-2038, Sept. 2019, doi: 10.1109/JBHI.2018.2876916.

[2] Y H. -C. Shin et al., "Deep Convolutional Neural Networks for Computer-Aided Detection: CNN Architectures, Dataset Characteristics and Transfer Learning," in IEEE Transactions on Medical Imaging, vol. 35, no. 5, pp. 1285-1298, May 2016, doi: 10.1109/TMI.2016.2528162

[3] Wen, L., Li, X. & Gao, L. A transfer convolutional neural network for fault diagnosis based on ResNet-50. Neural Comput&Applic 32, 6111–6124 (2020). https://doi.org/10.1007/s00521-019-04097-w.

[4] L. Wen, X. Li, X. Li and L. Gao, "A New Transfer Learning Based on VGG-19 Network for Fault Diagnosis," 2019 IEEE 23rd International Conference on Computer Supported Cooperative Work in Design (CSCWD), Porto, Portugal, 2019, pp. 205-209, doi: 10.1109/CSCWD.2019.8791884.

[5] Lin, C., Li, L., Luo, W., Wang, K. C. P., Guo, J. (2019) "Transfer Learning Based Traffic Sign Recognition Using Inception-v3 Model", PeriodicaPolytechnica Transportation Engineering, 47(3), pp. 242–250.

 ©IJSRED: All Rights are Reserved