International Journal of Scientific Research and Engineering Development-– Volume 6 Issue 4, July- Aug 2023

Available at www.ijsred.com

RESEARCH ARTICLE OPEN ACCESS

A Comparative Analysis of Various Deep Learning Architectures for Chest Radiograph Classification

Shuvam Ganguli

Department of Computer Sciences and Engineering, Sikkim Manipal Institute of Technology

Sikkim Manipal University, Majitar, East-Sikkim

shuvam.ganguli@gmail.com

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

Abstract:

Anaccurate assessment of chest radiographs is of vital essence in radiology for the diagnosis of thoracic diseases. Recently the performance shown by deep learning models in various tasks pertaining to image analysis has been amazing. The purpose of this study is to find differences in learning outcomes between the four Machine Learning Models, namely AlexNet, ResNet-50, VGG-16, and MobileNet v2 to accurately determine the presence of one of five diseases based on the image of the chest radiographs 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 classifying the radiograph into no finding or one of the five diseases it has been trained on.

Keywords **—AlexNet, ResNet-50, VGG-16, MobileNet v2, Neural Network.**

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

I. INTRODUCTION:

Radiograph classification is a task of high importance in radiology. It 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, radiograph classification 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 classify the radiographs into one of six classes of one no finding and five diseases. The NIH ChestXray-14 dataset which comprises of data of diverse populations is used with a mix of AP and PA radiographs, male and female patients, which has been separated into 14 diseases as per its name. It is split into training, validation, and testing sets.

The performance for each model is evaluated using the metrics of accuracy. 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 chest radiograph classification is concerned. Moreover, this study is successful in highlighting the potential of deep learning techniques in automated radiograph classification.

II. METHODOLOGY:

The methodology of this research involved using AlexNet, ResNet-50, VGG-16 and MobileNet v2 to accurately determine the class of thoracic disease based on the image of the radiograph. The size of the radiographs was set to 224x224. The accuracy for each class was calculated to judge the

performance of the models. All models were trained using Adam optimizer and a batch size of 32.

A. AlexNet:

AlexNet [4] is a convolutional neural network (CNN) architecture that managed to revolutionize the field of image classification. It was developed by Alex Krizhevsky, Ilya Sutskever, and Geoffrey Hinton in 2012, it then gained significant attention for its outstanding performance in the ImageNet Large Scale Visual Recognition Challenge (ILSVRC), 2012.

The architecture is characterized by its depth and utilization of convolutional layers. It has eight layers in total, comprising five convolutional layers, three pooing layers, and three fully connected layers. It is also notable that it introduced the concept of using rectified linear units (ReLU) as activation functions, which helped address the vanishing gradient problem and accelerate training. Additionally, it employs techniques such as dropout regularization to prevent overfitting and data augmentation to expand the training set.

Overall, AlexNet's architectural innovations and outstanding performance established the significance of deep learning for image classification. Its success laid the groundwork for subsequent advancements in deep neural networks and continues to inspire researchers in the field.

B. ResNet-50:

ResNet-50 [1] is a deep learning architecture belonging to the family of Residual Neural Networks. It gained prominence for effectively addressing the vanishing gradient problem and enabling the training of deep neural networks. The key innovation is the introduction of residual connections, skip connections, or shortcut connections. These connections allow information to flow directly from the initial layers to the later layers, enabling the network to learn residual mappings instead of directly approximating the underlying mapping. This allows for alleviation the degradation problem, where deep networks perform worse than shallow networks due to training difficulties.

Residual Networks consist of convolutional layers, pooling layers, fully connected layers, and residual blocks. A residual block typically comprises multiple convolutional layers along with shortcut connections that can bypass one or more layers. These shortcut connections facilitate the flow of gradients during training, enabling effective optimization of the network. ResNet-50 specifically, as per its name suggests, has a total of fifty layers, including forty-eight convolutional layers, a max pool layer, and an average pool layer.

During training, gradient descent optimization methods, such as Stochastic Gradient Descent, are used to update the model's weights. While, techniques like batch normalization are commonly employed to prevent overfitting and enhance the network's generalization capabilities.

ResNet-50's architecture allows it to learn rich and discriminative features, leading to state-of-the-art performance on large datasets. It has also seen extensive use in transfer learning, where pre-trained models are trained on large datasets and then fine-tuned for specific tasks with limited labeled data. This utilization of pre-trained models helps in leveraging the already learned representations and thus significantly improves performance on similar new tasks.

C. VGG-16:

VGG-16 [6], short for the Visual Geometry Group 16-layer model, is a deep convolutional neural network architecture that has gained widespread popularity and usage in the field of computer vision. It was introduced by the Visual Geometry Group at the University of Oxford in 2014. It is highly regarded for its simplicity and strong performance in image classification tasks.

The architecture of VGG-16 is characterized consisting of 16 layers, including 13 convolutional layers and 3 fully connected layers. The convolutional layers primarily utilize 3x3 filters with a stride of 1 and a padding of 1, thus, enabling the network to learn hierarchical representations of increasing complexity.

One significant aspect of VGG-16 is its uniformity in design. It maintains a consistent configuration throughout the network, utilizing the same number of filters and the same filter size in each convolutional layer. This simplicity makes it easier to implement and interpret, making it a popular choice for researchers and practitioners.

VGG-16 has proven to be highly effective in image classification tasks, often achieving top-tier performance on benchmark datasets. Its deep architecture allows it to capture intricate features and patterns in images, enabling state-of-the-art accuracy.

In professional usage, VGG-16 is frequently employed as a baseline architecture and a benchmark for evaluating the performance of new models and techniques. It serves as a reference point due to its simplicity, strong performance, and widely recognized characteristics. Furthermore, the pre-trained weights of VGG-16 on large-scale datasets are often used for transfer learning, where the learned representations are fine-tuned on specific tasks with limited labeled data, providing a boost in performance and reducing training time and resources required.

D. MobileNet v2:

MobileNetV2 [2] is a deep learning architecture specifically designed for efficient inference on mobile and embedded devices with limited computational resources. It helps to address the challenge of deploying neural networks on devices with restricted memory and processing power, while maintaining high accuracy in image classification tasks.

The key innovation in MobileNetV2 is the use of depth wise separable convolutions. Inverted residuals use expansion layers to increase the channel layer quantity, then apply the depth wise seperable convolution, providing a richer representation space. This is then followed by a pointwise convolution. The depth wise convolution applies a single filter to each input channel separately, helping reduce the computational cost significantly. The pointwise convolution then combines the outputs of the depth wise convolution, increasing the network's capacity to capture complex features.

MobileNetV2 also introduces linear bottleneck blocks, which consist of a sequence of 1x1 pointwise convolutions and 3x3 depth wise separable convolutions. These bottleneck blocks allow the network to efficiently capture spatial and channel-wise correlations, thus helping reduce both the model size and computational requirements, and to ensure that the gradients flow effectively during training.

During training, MobileNetV2 employs standard techniques like stochastic gradient descent for weight updates. Additionally, techniques like batch normalization and regularization methods such as dropout or weight decay are used to prevent overfitting and improve generalization.

MobileNetV2 achieves a good balance between model size, computational efficiency, and accuracy. It has been widely adopted for various mobile and embedded vision applications, where real-time inference and resource constraints are critical factors. Pre-trained MobileNetV2 models are also available, enabling transfer learning for specific tasks with limited labeled data on mobile and edge devices.

E. Dataset:

The dataset consists of images of chest radiographs from ChestXray-14 [5] dataset from NIH, which is the update of ChestXray-8 dataset. Only 5 classes were taken from the dataset – Atelectasis, Consolidation, Effusion, Infiltration, and No Finding.

The distribution of images among these classes was as follows: 4,215 for Atelectasis, 1,310for Consolidation, 3,955 for Effusion, 9,547 for Infiltration, and 60,361 for No Finding.

Furthermore, the dataset comprised 44,837 images of male radiographs and 34,551 images of female radiographs. The age of the patients varied within the dataset, with the maximum age recorded as 95 years and the minimum age as 1 year. Additionally, there were 49,214 images captured in the PA (Posterior-Anterior) view position and 30,174 images captured in the AP (Anterior-Posterior) position.

The dataset was split, 500 images from each class were allocated for testing, while the remaining images were divided into an 80:20 ratio for the training and validation datasets, respectively.

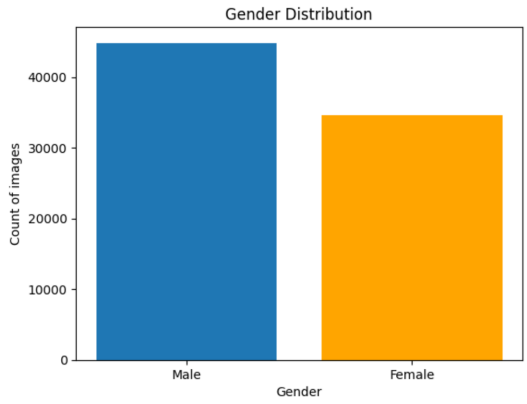

Figure 1. Total Count of Males and Females in the dataset

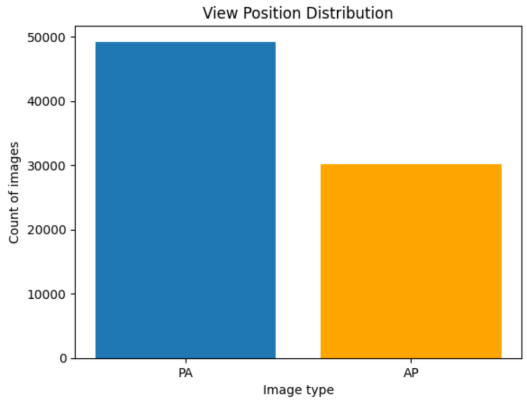

Figure 2. Total Count of AP and PA View positions in the dataset

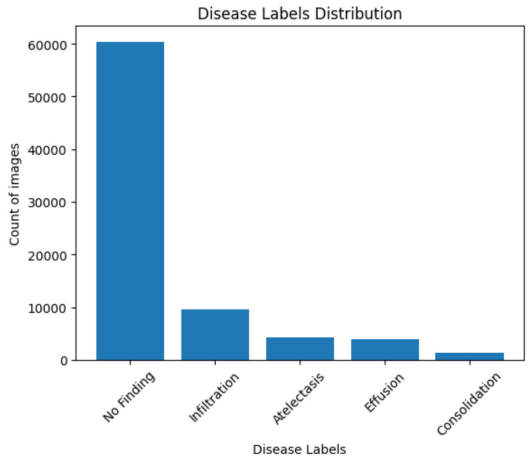

Figure 2. Number of images per class

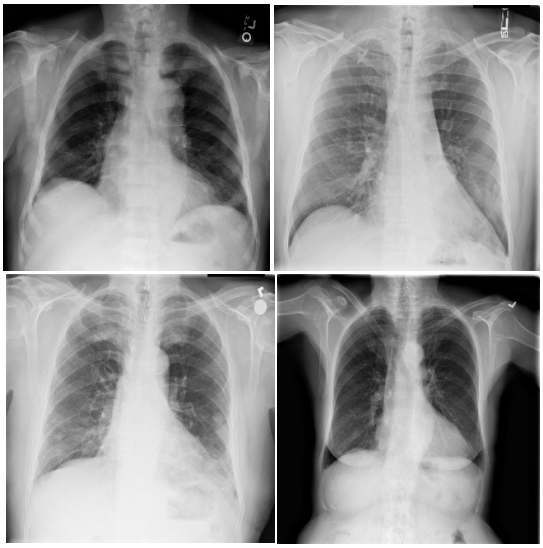

Figure 3. Sample Images from the dataset

III. RESULTS:

In this section, we present the results of the image classification as obtained by the four models. The assessment of the models was done based the model's accuracy per class.

TABLEI
Testing accuracy per class of the models

Model	AlexNet	ResNet-50	VGG-16	MobileNet v2
Atelectasis	72.8%	84.9%	82.3%	82.7%
Consolidation	74.3%	82.2%	78.1%	80.2%
Effusion	84.1%	87.1%	81.9%	81.1%
Infiltration	81.9%	82.6%	80.7%	78.6%
No Finding	88.8%	95.7%	89.2%	89.9%
Total	80.4%	86.5%	82.4%	82.5%

The table of testing accuracy reveals that ResNet-50 achieves the highest accuracy of 86.5%, while AlexNet demonstrates the lowest accuracy at 80.4% for the multiclass image classification task with 5 classes. While, VGG-16 and MobileNetV2 exhibit similar accuracies of 82.4% and 82.5% in total and even across all classes, indicating comparable performance.

Notably, the classes are very unevenly distributed in the dataset, with a significant skew towards the No Finding class, which has the highest number of images. Consequently, No Finding

emerges as the most accurately predicted class in all models, likely due to the ample availability of training examples.

Overall, ResNet-50 stands out as the best-performing model with the highest accuracy, while AlexNet sees the worst performance. VGG-16 and MobileNetV2 demonstrate similar and competitive accuracies across all classes, indicating their comparable performance on the multiclass image classification task.

IV. CONCLUSION

In this analysis, the models exhibit varying levels of performance in terms of accuracy, with the order being AlexNet< VGG-16 < MobileNetV2 < ResNet-50.

It is noteworthy that ResNet-50 achieves the highest accuracy, demonstrating its strong capability in the multiclass image classification task. MobileNetV2 follows closely in performance, showcasing its impressive efficiency while being lightweight enough to be trained even on mobile devices. This makes MobileNetV2 a promising choice for resource-constrained scenarios.

On the other hand, VGG-16 and AlexNet, although historically significant in the development of neural networks, do not demonstrate as strong a performance in this analysis. It is possible that utilizing upgraded and larger versions such as VGG-19 could potentially yield improved results for these architectures.

Considering the overall results, MobileNetV2 shows considerable promise due to its commendable performance and efficient design.

ACKNOWLEDGEMENT

I would like to express my sincere gratitude to the Department of Computer Science and Engineering at Sikkim Manipal Institute of Technology as this research was conducted to their specialization program.

I would also like to thank the National Institutes of Health, USA for providing the ChestX-ray14 dataset on which the analysis was performed, as without their efforts to create the dataset this research paper would not have been possible.

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

REFERENCES

[1] 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.

[2] Q. Xiang, X. Wang, R. Li, G. Zhang, J. Lai, and Q. Hu, "Fruit Image Classification Based on MobileNetV2 with Transfer Learning Technique," Proceedings of the 3rd International Conference on Computer Science and Application Engineering. ACM, Oct. 22, 2019. doi: 10.1145/3331453.3361658.

[3] H. Alshazly, C. Linse, E. Barth, and T. Martinetz, "Ensembles of Deep Learning Models and Transfer Learning for Ear Recognition," Sensors, vol. 19, no. 19. MDPI AG, p. 4139, Sep. 24, 2019. doi: 10.3390/s19194139.

[4] A. A. Almisreb, N. Jamil, and N. M. Din, "Utilizing AlexNet Deep Transfer Learning for Ear Recognition," 2018 Fourth International Conference on Information Retrieval and Knowledge Management (CAMP). IEEE, Mar. 2018. doi: 10.1109/infrkm.2018.8464769.

[5] Wang, Xiaosong, Yifan Peng, Le Lu, Zhiyong Lu, Mohammadhadi Bagheri, and Ronald M. Summers. "Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases." In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 2097-2106. 2017.

[6] A. Inés, C. Domínguez, J. Heras, E. Mata, and V. Pascual, "Biomedical image classification made easier thanks to transfer and semi-supervised learning," Computer Methods and Programs in Biomedicine, vol. 198. Elsevier BV, p. 105782, Jan. 2021. doi: 10.1016/j.cmpb.2020.105782.

