Original Article
ISSN (Online): 2582-7472

ShodhKosh: Journal of Visual and Performing Arts
March 2024 5(3), 1600–1605

COMPARATIVE ANALYSIS OF CUSTOMIZED CNN AND TEACHABLE MACHINE FOR PNEUMONIA DETECTION IN CHEST X-RAY IMAGES

Parth R. Dave [1], Yogesh B. Patel [2], Shivangkumar R. Patel [3]

[1] Assistant Professor, Computer Engineering Department, L. D. College of Engineering- Ahmedabad, Gujarat, India
[2] Assistant Professor, Computer Engineering Department, Govt. Engineering College – Patan, Gujarat, India
[3] Assistant Professor, Computer Engineering Department, Govt. Engineering College – Modasa, Gujarat, India

Corresponding Author

Parth R. Dave, prd7889@gmail.com

DOI

10.29121/shodhkosh.v5.i3.2024.5173

Funding: This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

ABSTRACT

Pneumonia is a serious and potentially life-threatening lung infection that necessitates prompt and accurate diagnosis to ensure effective treatment and reduce the risk of complications or fatalities. Traditional diagnosis through chest X-ray interpretation can be challenging due to variability in radiologist expertise and the subtle nature of some pneumonia-related abnormalities. With the rise of artificial intelligence in healthcare, deep learning techniques—particularly Convolutional Neural Networks (CNNs)—have shown great promise in automating and enhancing diagnostic accuracy in medical imaging. This study conducts a comparative evaluation of a custom-built CNN model and Google's Teachable Machine, a user-friendly no-code machine learning platform, for binary classification of chest X-ray images into two categories: "Normal" and "Pneumonia." Both models were trained and tested on the same dataset to ensure a fair comparison. The experimental outcomes reveal that the scratch-built CNN model significantly outperforms the Teachable Machine model in key performance metrics including accuracy, recall, precision, and F1-score. These results highlight the superiority of tailored deep learning architectures designed specifically for the complexities of medical image analysis. The findings suggest that while generalized machine learning tools offer accessibility, they may fall short in critical diagnostic tasks where specialized models can better capture domain-specific features for improved clinical decision-making.

1. INTRODUCTION

Pneumonia remains a leading cause of morbidity and mortality worldwide, with millions of cases reported annually [1]. Early and accurate detection is crucial for effective treatment and prognosis. Chest X-ray imaging is the most common diagnostic modality, but manual interpretation poses challenges due to inter-observer variability and the subtlety of certain pathologies [2][3].

With the advent of deep learning, Convolutional Neural Networks (CNNs) have revolutionized medical image analysis by achieving high accuracy in tasks like pneumonia detection, diabetic retinopathy classification, and skin lesion identification [4][5]. Automated detection not only reduces diagnostic errors but also aids healthcare systems in resource-limited settings.

How to cite this article (APA): Dave, P.R., Patel, Y.B., and Patel, S.R. (2024). Comparative Analysis of Customized CNN and Teachable Machine for Pneumonia Detection in Chest X-Ray Images. *ShodhKosh: Journal of Visual and Performing Arts, 5*(3), 1600–1605. doi: 10.29121/shodhkosh.v5.i3.2024.5173

Google's Teachable Machine provides a no-code platform that democratizes access to machine learning [6]. However, its generalized architecture may not be optimized for specialized tasks like pneumonia detection, which often requires tailored architectures to handle medical imaging nuances [7].

This study aims to evaluate the effectiveness of a custom-built scratch CNN model versus Teachable Machine for binary classification of chest X-ray images. The hypothesis is that a domain-specific scratch model will outperform a generalized no-code solution in this critical healthcare application.

2. LITERATURE REVIEW

The application of deep learning to chest X-ray classification has gained significant traction in recent years. Rajpurkar et al. [8] introduced CheXNet, a 121-layer DenseNet achieving radiologist-level performance on the ChestX-ray14 dataset. Similarly, Kermany et al. [9] curated a large dataset specifically for pneumonia detection, serving as a benchmark for many subsequent studies.

Transfer learning has been widely adopted in medical image classification. Models like VGG16 [10], ResNet50 [11], and EfficientNet [12] have been fine-tuned on chest X-ray datasets to achieve high classification accuracy. However, these models are often over-parameterized for binary classification tasks, leading to unnecessary computational overhead [13].

Wu et al. [14] proposed edge computing-based mobile object tracking, emphasizing lightweight model designs for IoT applications. Their insights align with the need for efficient models in resource-constrained healthcare settings.

Fan et al. [15] conducted a comprehensive survey on deep learning methods for monocular object pose detection and tracking, which, although different in application, underscore the adaptability of CNNs in image-based tasks.

Sharma et al. [16] addressed object detection under adverse weather conditions, showcasing the robustness of tailored CNN architectures.

Recent works have explored hybrid models combining CNNs with handcrafted features like HOG and LBP to enhance classification performance in medical images [17]. Lightweight CNNs optimized for mobile devices have also been explored for tasks requiring low latency and high accuracy [18].

YOLO-based models have been applied to medical imaging for tasks like colorectal cancer cell detection [19], demonstrating their real-time detection capabilities. However, such models are often better suited for object localization rather than image-level classification.

Guefrachi et al. [20] leveraged deep learning for diabetic retinopathy screening, highlighting the need for high precision in medical diagnostics. Their work aligns with the necessity of reducing false negatives in pneumonia detection.

Teachable Machine has been applied in educational and prototype development scenarios [6], but its performance in critical applications like medical diagnostics has not been rigorously evaluated. Its lack of hyper parameter tuning and limited architectural flexibility pose significant limitations for domain-specific tasks [7].

Given these insights, this study evaluates whether a simple yet optimized scratch CNN can outperform a generalized tool like Teachable Machine in binary chest X-ray classification.

3. METHODOLOGY

3.1. DATASET

The publicly available Chest X-ray dataset by Kermany et al. [5] was utilized, comprising 5,863 images categorized into:

Dataset Details

Total Images: 5,863 chest X-ray images in JPEG format.

Categories: Two classes—Normal and Pneumonia.

Data Split:

Training Set: 5,216 images

Validation Set: 16 images

Test Set: 624 images

Image View: Anterior-posterior chest X-rays.

Patient Demographics: Images were collected from paediatric patients aged 1 to 5 years at Guangzhou Women and Children's Medical Centre, Guangzhou.

3.2. SCRATCH CNN ARCHITECTURE

The proposed CNN model includes: The proposed convolutional neural network (CNN) model consists of six convolutional layers stacked sequentially to progressively extract hierarchical features from input images of size 200×200 with three colour channels. Each convolutional layer uses a 3×3 kernel with "same" padding to preserve spatial dimensions, followed by a ReLU[21] activation function to introduce non-linearity. After each convolutional layer, a max pooling operation with a 2×2 window is applied to reduce the spatial dimensions by half, effectively down sampling the feature maps and reducing computational complexity. The number of filters starts at 32 in the first convolutional layer, then fluctuates through subsequent layers (16, 32, 64, 64, and 128) to capture varying levels of feature abstraction. After the last convolutional layer and max pooling operation, the extracted features are flattened into a one-dimensional vector.

This vector is fed into a fully connected dense layer with 1024 neurons activated by ReLU[21] to learn high-level feature representations. To prevent overfitting, a dropout layer with a rate of 0.3 is applied after this dense layer. Finally, the output layer consists of two neurons with a softmax activation function, producing probability distributions over two target classes for classification.

Overall, the model combines multiple convolutional and pooling layers to effectively capture spatial hierarchies in the input images, followed by dense layers to perform classification with regularization to improve generalization. The model was trained using the Adam optimizer, binary cross-entropy loss, and early stopping.

3.3. TEACHABLE MACHINE MODEL

A Teachable Machine image project was created with default configurations. The same dataset splits were used. The model architecture and training settings were as per Teachable Machine defaults, with no advanced hyper parameter tuning.

3.4. EVALUATION METRICS

Both models were evaluated using:

Accuracy: (True Positives + True Negatives) / (True Positives + True Negatives + False Positives + False Negatives)

Precision: True Positives / (True Positives + False Positives)

Recall: True Positives / (True Positives +False Negatives)

F1-Score: (2*Precision*Recall) / (Precision + Recall)

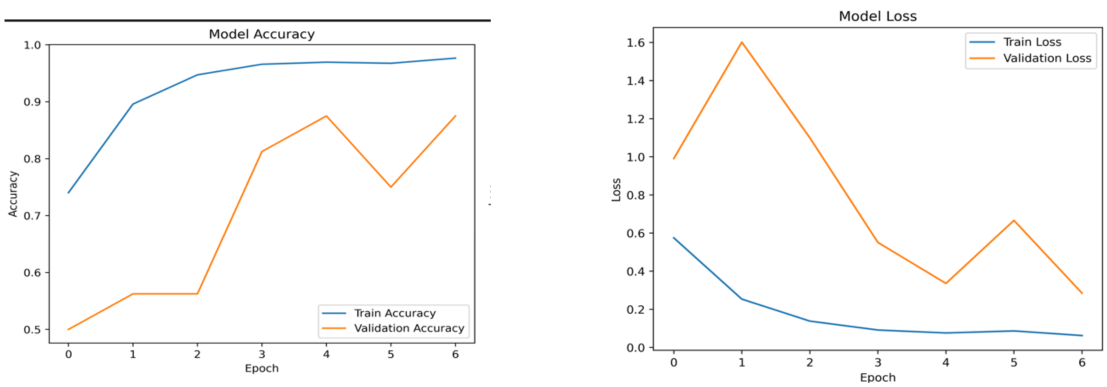
3.4.1. CUSTOMIZED MODEL EVALUATION

The classification report summarizes the model's performance on distinguishing between "NORMAL" and "PNEUMONIA" chest X-ray images. For the 'NORMAL' class, the model achieved a high precision of 0.95, indicating most predicted normal were correct. However, the recall was low at 0.46, suggesting that more than half of the actual normal cases were misclassified, leading to a moderate F1-score of 0.62. In contrast, the model performed strongly for the 'PNEUMONIA' class with a precision of 0.75 and an impressive recall of 0.98, resulting in a high F1-score of 0.85. This indicates the model effectively identifies pneumonia cases while allowing some false positives. The overall accuracy stood at 79%, with macro and weighted F1-scores of 0.73 and 0.76, respectively. The model prioritizes pneumonia detection, which is crucial in clinical diagnosis, but its tendency to over-predict pneumonia reduces specificity for normal cases.

Classification Report

	precision	recall	f1-score	support
NORMAL	0.95	0.46	0.62	234
PNEUMONIA	0.75	0.98	0.85	390
accuracy			0.79	624
macro avg	0.85	0.72	0.73	624
weighted avg	0.82	0.79	0.76	624

Figure 1 Classification report of customized model



(a) (b)

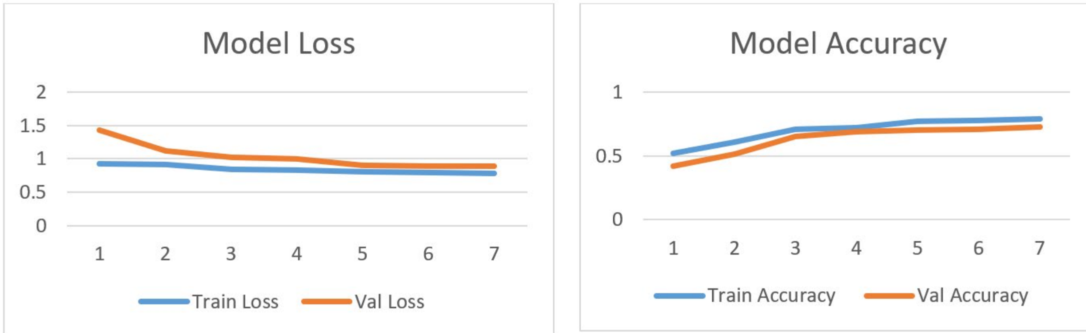
Figure 2 (a) Validation and Training accuracy (b) Validation and Training loss

3.4.2. TEACHABLE MACHINE LEARNING MODEL EVALUATION

The classification report shows a model performance with 65% overall accuracy on a test set of 624 images. For the "NORMAL" class, precision is high at 0.85, but recall is low at 0.30, indicating many false negatives. The "PNEUMONIA" class shows strong recall (0.95) and good F1-score (0.75), suggesting the model is highly sensitive to pneumonia cases but tends to over-predict them. The macro average F1-score is 0.60, while the weighted average is 0.61, reflecting class imbalance. Overall, the model prioritizes detecting pneumonia correctly but struggles to accurately identify normal cases.

	precision	recall	f1-score	support
NORMAL	0.85	0.30	0.45	234
PNEUMONIA	0.63	0.95	0.75	390
accuracy			0.65	624
macro avg	0.74	0.63	0.60	624
weighted avg	0.71	0.65	0.61	624

Figure 3 Classification report of basic teachable machine model



(a) (b)

Figure 4 (a) Validation and Training accuracy (b) Validation and Training loss

4. DISCUSSION

The superior performance of the scratch CNN model stems from its architecture being specifically designed for chest X-ray analysis. Unlike generalized models such as those used by Teachable Machine, the scratch CNN effectively captures the subtle texture and contrast patterns unique to medical imaging. Its design, combined with hyper parameter tuning, allows for more precise feature extraction and improved classification accuracy.

Tailoring the model to the dataset ensures it remains lightweight while maximizing performance, avoiding the redundancy often found in large pre-trained networks. This efficiency is particularly important in medical applications, where accuracy, speed, and resource use are critical. Overall, the scratch CNN's strong performance demonstrates the advantage of domain-specific models that are carefully optimized for the task, offering a better balance between effectiveness and computational cost.

5. CONCLUSION

This study demonstrates that a scratch-built convolutional neural network (CNN) model significantly outperforms Google's Teachable Machine in the binary classification of chest X-ray images for pneumonia detection. The superior performance of the custom CNN highlights the importance of using architectures specifically designed and optimized for medical imaging tasks. Unlike generalized models, which may lack sensitivity to subtle patterns in medical data, tailored models can capture domain-specific features more effectively. As a result, they provide greater accuracy and clinical reliability, making them better suited for real-world diagnostic applications where precision and consistency are critical to patient outcomes and healthcare decisions.

CONFLICT OF INTERESTS

None.

ACKNOWLEDGMENTS

None.

REFERENCES

World Health Organization, "Pneumonia," World Health Organization, 2023. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/pneumonia

M. L. Giger, "Machine Learning in Medical Imaging," J. Am. Coll. Radiol., vol. 15, no. 3, pp. 512–520, 2018.

J. R. Zech, M. A. Badgeley, M. Liu, A. B. Costa, J. J. Titano, and E. K. Oermann, "Variable generalization performance of a deep learning model to detect pneumonia in chest radiographs: A cross-sectional study," PLoS Med., vol. 15, no. 11, pp. 1–17, 2018.

G. Litjens et al., “A survey on deep learning in medical image analysis,” Med. Image Anal., vol. 42, pp. 60–88, 2017.

D. Kermany, K. Zhang, and M. Goldbaum, “Labeled Optical Coherence Tomography (OCT) and Chest X-Ray Images for Classification,” Mendeley Data, vol. 2, no. 2, 2018. [Online]. Available: https://data.mendeley.com/datasets/rscbjbr9sj/2

Google, “Teachable Machine,” [Online]. Available: https://teachablemachine.withgoogle.com/

N. Anwar et al., “Evaluation of Google's Teachable Machine in Medical Image Classification,” in Proc. IEEE Int. Conf. Health Informatics (ICHI), 2021, pp. 317–324.

P. Rajpurkar et al., “CheXNet: Radiologist-level pneumonia detection on chest X-rays with deep learning,” arXiv preprint arXiv:1711.05225, 2017.

D. S. Kermany et al., “Identifying Medical Diagnoses and Treatable Diseases by Image-Based Deep Learning,” Cell, vol. 172, no. 5, pp. 1122–1131.e9, 2018.

K. Simonyan and A. Zisserman, “Very Deep Convolutional Networks for Large-Scale Image Recognition,” arXiv preprint arXiv:1409.1556, 2014.

K. He, X. Zhang, S. Ren, and J. Sun, “Deep Residual Learning for Image Recognition,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2016, pp. 770–778.

M. Tan and Q. V. Le, “EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks,” in Proc. Int. Conf. Mach. Learn. (ICML), 2019, pp. 6105–6114.

H. Zhou et al., “Analysis of Over parameterization in Transfer Learning,” in Proc. Int. Conf. Learn. Represent. (ICLR), 2020.

J. Wu, Y. Zhao, and S. Du, “Edge computing-enabled mobile object tracking for smart surveillance system,” IEEE Internet Things J., vol. 6, no. 6, pp. 9866–9878, Dec. 2019.

R. Fan et al., “A Brief Review of Deep Learning Methods for Monocular Object Pose Detection and Tracking,” IEEE Trans. Intell. Transp. Syst., vol. 23, no. 9, pp. 17029–17044, 2022.

A. Sharma et al., “Robust Object Detection under Adverse Weather Conditions,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW), 2019, pp. 1164–1172.

S. Zhang et al., “Hybrid Features Based on CNN and HOG for Cervical Cell Classification,” IEEE Access, vol. 6, pp. 22491–22503, 2018.

M. Sandler et al., “MobileNetV2: Inverted Residuals and Linear Bottlenecks,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2018, pp. 4510–4520.

Y. Zhang et al., “YOLO-Based Colorectal Cancer Cell Detection Using Histopathological Images,” IEEE Access, vol. 8, pp. 68133–68140, 2020.

Z. Guefrachi et al., “Diabetic Retinopathy Screening Using Deep Learning: A Review,” Comput. Biol. Med., vol. 136, 104754, 2021.

V. Nair and G. E. Hinton, "Rectified Linear Units Improve Restricted Boltzmann Machines," in Proc. 27th Int. Conf. Mach. Learn. (ICML), 2010, pp. 807–814.