Kalpa Publications in Computing
Volume 19, 2024, Pages 445–454
Proceedings of 6th International Conference
on Smart Systems and Inventive Technology

Experimentally Enhancing ResNet50 Performance on the Intel Dataset Through Architectural Modifications

Ketone Agasti
ketone.agasti@gmail.com
, Kisor.G
kisorg01@gmail.com
, Maanav Thalapilly
thalapillymaanav@gmail.com
, and

Pranathi.M
pranathi02m@gmail.com

Vinitha Panicker J*
vinithapanicker@am.amrita.edu

Department of Computer Science and Engineering
Amrita School of Computing, Amrita Vishwa Vidyapeetham, Amritapuri, India

Abstract

This study focuses on enhancing the performance of the ResNet50 model on the Intel dataset, a collection of images depicting diverse natural scenes under various environmental conditions. While ResNet50 has shown remarkable performance in image classification tasks, its application to the Intel dataset reveals certain limitations in accurately discerning subtle features within scenes. To address this, proposed architectural modifications to ResNet50 aimed at capturing intricate features specific to the Intel dataset. Four distinct modifications are introduced, tailored to exploit different aspects of scene complexity present in the dataset. Through extensive experimentation and evaluation, we demonstrate the effectiveness of these modifications in improving the model's classification accuracy on the Intel dataset. the findings not only contribute to advancing deep learning methodologies for image analysis but also underscore the importance of tailored model design for specific task domains.

Keywords* — *ResNet50, Architectural modifications, Spatial Pyramid Pooling, Transfer learning, Scene classification, Neural networks

1 Introduction

The Intel dataset, a collection of images capturing various natural scenes under diverse environmental conditions, presents a rich and challenging landscape for image classification tasks [1]. With over thousands of images spanning different categories such as forests, mountains, buildings, and rivers, this dataset reflects real-world scenarios where accurate scene classification is crucial for applications like autonomous navigation systems, environmental monitoring, and urban planning [2]. However, the dataset's complexity lies in its diverse scenes, varying lighting conditions, and occlusions, which pose significant challenges for traditional classification algorithms [3]. Understanding and effectively classifying scenes in this dataset not only advances the field of computer vision but also has tangible implications for improving the performance of automated systems operating in dynamic and unpredictable environments.

Despite the remarkable performance of the baseline ResNet50 model in various image classification tasks [4], its application to the Intel dataset reveals certain limitations, particularly in accurately discerning subtle features within scenes under diverse conditions [5]. The motivation behind incorporating additional layers into the ResNet50 architecture stems from the need to enhance its capability to capture and leverage intricate features specific to the Intel dataset [6]. By introducing supplementary layers, we aim to enable the model to extract more nuanced representations of scenes, thereby improving its ability to discriminate between different categories and increasing overall classification accuracy.

The architectural modifications are designed to address specific challenges posed by the Intel dataset, such as variability in lighting, texture, and scene composition [5], ultimately enhancing the robustness and adaptability of the model to real-world scenarios. In this study, we propose four distinct modifications to the ResNet50 architecture, each tailored to exploit different aspects of scene complexity present in the Intel dataset [7].

To evaluate the performance of the modified architectures, we conducted extensive experiments following a rigorous methodology [8]. We preprocessed the Intel dataset to standardize image sizes, normalize pixel intensities, and augment data to increase diversity and robustness [9]. The modified architectures were initialized with pre-trained weights from the baseline ResNet50 model and fine-tuned on the training set using stochastic gradient descent with momentum [1]. We employed techniques such as learning rate scheduling, weight decay, and dropout regularization to prevent overfitting and improve generalization [10]. Model performance was evaluated on a separate test set using standard classification metrics such as accuracy, precision, recall, and F1-score, allowing for comprehensive assessment and comparison of the different architectures [8].

The experimental results demonstrate the effectiveness of the proposed architectural modifications in enhancing the performance of the ResNet50 model on the Intel dataset [6]. Across various subsets and evaluation metrics, we observed consistent improvements in classification accuracy compared to the baseline model [11]. Specifically, the architectures incorporating attention mechanisms and feature fusion techniques achieved notable gains in discerning fine-grained details and capturing contextual information, leading to significant enhancements in scene classification accuracy [10]. Moreover, the spatial pyramid pooling modification proved effective in accommodating variations in scene scale and composition, further boosting the model's robustness and adaptability [1]. These findings underscore the importance of architectural design in addressing dataset-specific challenges and highlight the potential for improving deep learning models' performance through targeted modifications.

Extending the evaluation of architectural modifications to encompass a wider range of environmental conditions and scenarios would provide a more comprehensive understanding of their robustness. Real-world environments are inherently dynamic and unpredictable, presenting challenges that may not be fully captured by existing evaluation protocols. By expanding the scope of evaluation to include real-world deployment scenarios and considering factors such as domain shift and adversarial attacks, we can better assess the practical utility and limitations of modified architectures.

The implications of the findings extend beyond the realm of scene classification, shedding light on the importance of architectural design in adapting deep learning models to specific datasets and tasks [3]. By systematically analyzing the impact of different architectural modifications on model performance, we provide valuable insights into the mechanisms underlying effective feature representation and discrimination in complex visual environments [6]. While this study focuses on the Intel dataset, the principles and methodologies outlined can be generalized to other image analysis tasks and datasets, leading to the development of more robust and adaptive deep learning models [1]. However, it's essential to acknowledge the limitations of the study, including the potential trade-offs between model complexity and computational efficiency, as well as the need for further validation on additional datasets and real-world applications [8]. Nonetheless, the research underscores the potential of architectural innovations in advancing the state-of-the-art in computer vision and lays the groundwork for future investigations into tailored model design for specific task domains.

2 Related Works

Recent research emphasizes the significance of remote sensing image scene classification, fueled by deep neural networks' robust feature learning capabilities. However, a comprehensive review of deep learning methods in this context remains scarce. This article [1] fills this gap by systematically surveying over 160 papers, addressing challenges and categorizing methods into autoencoder-based, convolutional neural network-based, and generative adversarial network-based approaches. Additionally, it highlights benchmark datasets and evaluates the performance of representative algorithms across commonly used benchmarks. Finally, it outlines promising avenues for future research in this rapidly evolving field.

In [2] the authors presents a comprehensive survey of recent advancements in scene classification using deep learning, covering over 200 publications and addressing challenges, benchmark datasets, taxonomy, and performance comparisons. It concludes with a list of promising research opportunities, offering valuable insights for researchers in the field.

Several similar studies have tackled the difficulty of training deep neural networks by investigating various architectural alterations and training methodologies. One significant technique is the use of skip connections proposed by the authors [5], as seen in residual networks (ResNets), which ease the vanishing gradient problem by allowing for the direct flow of information between layers. This breakthrough has cleared the path for training far deeper networks with hundreds of layers while retaining good gradient flow. Furthermore, approaches such as batch normalization, which normalizes activations within each mini-batch during training, and adaptive optimization algorithms like Adam have helped to stabilize and speed up deep network training. Other research lines have focused on network initialization methods, regularization strategies, and innovative activation functions to improve deep architecture training.

Recent works [5] highlight the complexity of neural networks, with deep architectures facilitating the expression of intricate functions. However, challenges such as vanishing gradients hinder effective training, a limitation mitigated by the introduction of residual networks, which enable deeper architectures.

In [5] authors proposed the scene classification tasks by introducing a novel categorization approach. By re-categorizing the dataset into previously unexplored classifications (e.g., natural scenes vs. real scenes), the proposed model's adaptability and accuracy are further highlighted.

Several recent studies underscore the importance of remote sensing image scene classification, leveraging the robust feature learning capabilities of deep neural networks. However, a comprehensive review [6] focusing on recent advancements in deep learning methods for this task is currently absent in literature. To address this gap, this paper conducts a systematic survey encompassing over 160 research articles. It delves into the key challenges of remote sensing image scene classification, categorizing surveyed methods into autoencoder-based, convolutional neural network-based, and generative adversarial network-based approaches. Furthermore, it introduces benchmark datasets commonly employed in this domain and provides a comprehensive performance summary of more than two dozen representative algorithms. Lastly, it outlines promising directions for future research in this rapidly evolving field.Significant research using deep learning approaches has been carried out in the last several years [12, 13, 14, 15, 16, 17]

3 Material And Methods

3.1 Data Set

This study uses the Intel Image Classification dataset, obtained from Kaggle, as the foundation for the research. This dataset has six unique classes: Forest, Glacier, Mountain, Sea, Buildings, and Streets. Each class reflects a distinct category of nature or urban settings, adding to the dataset's diversity and depth.

By using the Intel dataset, hope to examine and assess the effectiveness of various deep learning architectures in picture classification tasks. The dataset's inclusion of varied scene categories allows for a thorough examination of model performance in a variety of visual situations, ranging from natural landscapes to urban surroundings.

Furthermore, the dataset's availability on Kaggle assures accessibility and repeatability, allowing other academics and practitioners to check and extend the findings. Through thorough testing and analysis of this well-curated dataset, wanted to provide useful insights to the larger area of computer vision and deep learning research.

3.2 Model Architecture and Training

Employed a transfer learning approach utilizing the ResNet50 architecture as the base model for the image classification task. The pre-trained ResNet50 model was obtained from a widely used deep-learning library. Adapted ResNet50 architectures utilize pooling layers for easier feature extraction and complexity management, enhancing performance in scene categorization tasks on the Intel dataset. Softmax activation in the output layer transforms raw scores into class probabilities, crucial for multi-class scene classification. To tailor the model for specific task, removed the original output layers and appended additional layers to create three distinct models for comparative analysis.

3.2.1 Proposed Model 1 Architecture

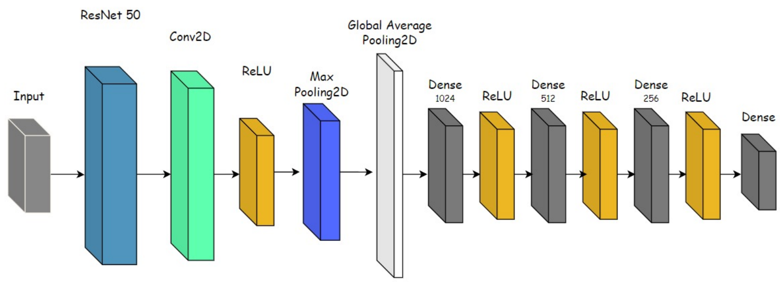

Fig.1 Proposed Model 1

The first proposed model (Fig. 1) comprises additional layers after the base model. Added 1 convolutional layer, 3 dense layers, and 1 output layer. Specifically, added 1 convolutional layer with 64 filters of size (3, 3), followed by ReLU activation and max-pooling. Additionally, included 3 dense layers with 1024, 512, and 256 units, respectively, each followed by ReLU activation. The output layer consists of 6 nodes with softmax activation to match the six-class classification problem.

3.2.2 Proposed Model 2 Architecture

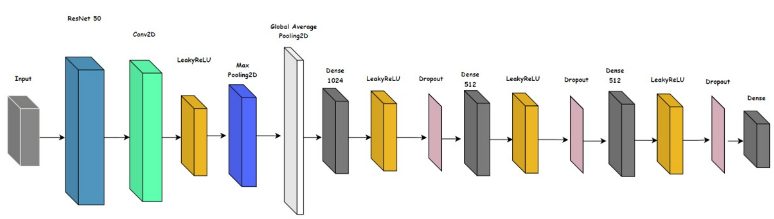

Fig.2 Proposed Model 2

The second proposed model (Fig. 2) features a unique set of added layers, including 1 convolutional layer, 3 dense layers, and 3 dropout layers. Specifically, added 1 convolutional layer with 64 filters of size (3, 3), followed by LeakyReLU activation and max-pooling. Following this, included 3 dense layers with 1024, 512, and 256 units, respectively, each followed by LeakyReLU activation. Additionally, 3 dropout layers with a dropout rate of 0.5 were inserted for regularization. The final layer utilizes softmax activation to produce class predictions for the six target classes.

3.2.3 Proposed Model 3 Architecture

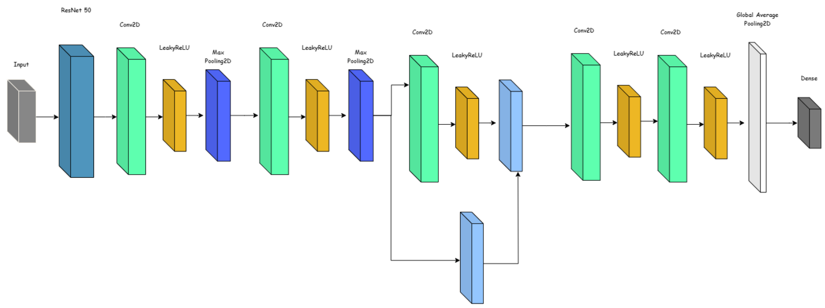

Fig.3 Proposed Model 3

For the third proposed model (Fig. 3), a residual connection was integrated following the Convolution layers atop ResNet50, after the removal of the output layer. The model comprises of convolutional layers with LeakyReLU activation and max-pooling, as well as residual connections to facilitate gradient flow during training. Specifically, added 3 convolutional layers with 64 filters of size (3, 3) each, followed by LeakyReLU activation and max-pooling. Additionally, a residual connection was established after the second convolutional layer. Subsequently, included 2 convolutional layers with 128 filters of size (3, 3) each are added by LeakyReLU activation. The network concludes with a global average pooling layer and 2 dense layers, with 150 and 6 units, respectively, utilizing softmax activation for multi-class classification.

4 Experimental and Result Analysis

4.1 Training Procedure

All proposed models were trained using a suitable optimization algorithm, such as stochastic gradient descent (SGD), and compiled with categorical cross-entropy loss. Applied L2 regularization to mitigate overfitting. The dataset was split into training and validation sets for model training and a testing set, and hyperparameters were fine-tuned to achieve optimal performance.

4.2 Hardware and Software

An Intel i7 core CPU (2.8 GHz) with 16GB RAM was used for the testing. Version 2.15.0 of the TensorFlow deep learning framework and Version 2.15.0 of Keras were utilized in the code implementation to construct the models. To ensure smooth integration and reproducibility of the studies, extra libraries were used for data manipulation and evaluation metrics computation, such as Scikit-learn version 1.2.2 and Matplotlib version 3.7.1.

4.3 Loss Graph Analysis

The analysis of the loss graph for the ResNet50 model and its derivatives provides valuable information about their training behavior and ability to generalize. All proposed models were trained for 25 epochs, and early stopping criteria were used to prevent overfitting. This criterion stops the training when the testing loss starts to plateau or increase, ensuring that the models do not memorize the training data excessively.

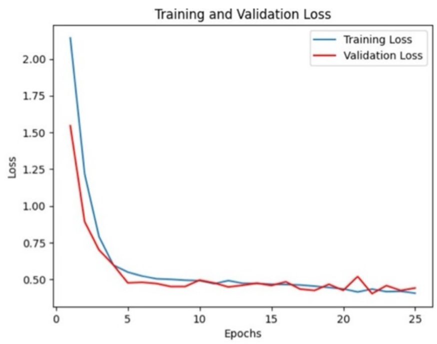

Fig.4 Proposed Model 1

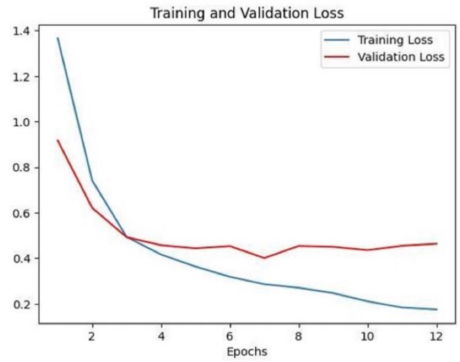

Fig.5 Proposed Model 2

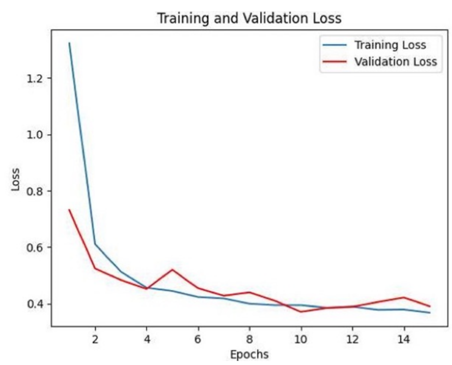

Fig.6 Proposed Model 3

The Proposed Model 1 (Fig. 4) finished 25 epochs with no overfitting. It consistently performed well during training, using all epochs effectively to improve generalization. Meanwhile, Model 2 (Fig. 5) stopped after just 12 epochs. The early stopping criteria kicked in, likely due to diminishing returns or overfitting risk. Still, The Proposed Model 2 (Fig. 5) performed remarkably well despite the shorter training time. It used the available epochs efficiently, generalizing the task effectively.

The Proposed Model 3 (Fig. 6) stopped training after 15 epochs because it met the early stopping criteria. The model's performance probably didn't improve anymore or started getting worse, so training stopped. This prevented overfitting to the training data. Even though The Proposed Model 3 (Figure. 6) trained for less time than The Proposed Model 1 (Figure. 4), it could still generalize well. The early stopping mechanism kept the model from overfitting while letting it learn relevant features. Early stopping is important for achieving optimal performance. It balances model complexity and generalization ability.

4.4 Results

Table 1: COMPARISON OF TRAINING AND TESTING ACCURACY FOR VARIOUS MODELS

	Training And Testing Accuracy		
S.No	***Models***	***Training***	***Testing***
1	ResNet50	93.19	89.89
2	Proposed Model1	92.09	90.89
3	Proposed Model2	99.11	92.33
4	Proposed Model3	92.29	92.52

The observed training and testing accuracies of the ResNet50 model (Table. 1) and its derivatives, the proposed Model 1 (Fig. 4), proposed Model 2 (Fig. 5), and proposed Model 3 (Fig. 6), reveal intriguing insights into their comparative performance. Notably, the baseline ResNet50 achieved a commendable training accuracy of 93% but exhibited a comparatively lower testing accuracy of 89.98%.

In contrast, the models derived from ResNet50, namely the proposed Models 1 (Fig. 4), 2 (Fig. 5) and 3 (Fig. 5) showcase improvements in both training and testing accuracies. The proposed Model 1 (Fig. 4), with training and testing accuracies of 92.09% and 90.98%, respectively, demonstrates a notable reduction in the overfitting gap observed in the baseline ResNet50. The proposed Model 2 (Fig. 5) exhibits remarkable performance, achieving a training accuracy of 99.11% and a testing accuracy of 92.33%. This result indicates that the additional layers introduced in proposed Model 2 (Fig. 5) contribute significantly to enhancing the model's generalization capabilities, leading to improved performance on unseen data. Similarly, the proposed Model 3 (Fig. 6), with a training accuracy of 92.52% and a testing accuracy of 92.29%, reinforces the trend of enhanced generalization across the derivative models. Overall, the consistent trend of improved testing accuracies in proposed Models 1 (Fig. 4), 2 (Fig. 5), and 3 (Fig. 6) compared to the baseline ResNet50 suggests that the introduced modifications contribute positively to the models' ability to generalize to new data. The investigation aimed at enhancing the performance of ResNet50 on the Intel dataset by adding supplementary layers resulted in the development of four distinct model architectures. These modifications are aimed to improve classification accuracy and robustness on a dataset comprising various natural scenes captured under diverse conditions.

Comparative analysis revealed insights into the performance of the baseline ResNet50 and its derivatives. While the baseline model achieved commendable training accuracy, it exhibited comparatively lower testing accuracy. However, the modified architectures showed improvements in both training and testing accuracies. Particularly, Proposed Model 2 (Fig. 5), which incorporated additional layers, demonstrated remarkable performance, indicating significant enhancement in generalization capabilities.

Furthermore, the study provided insights valuable for the research community. It demonstrated the practicality and effectiveness of transfer learning with ResNet50, offering guidance on architectural modifications for specific tasks. The reduction in overfitting and improved testing accuracies across the derived models highlighted the positive impact of introduced modifications on generalization to unseen data. The inclusion of a residual network architecture in Proposed Model 3 (Fig. 6) underscored the potential benefits of advanced architectural concepts, facilitating better gradient flow during training and resulting in enhanced generalization capabilities.

Overall, the research contributes to advancing deep learning for image classification by exploring novel architectural variations and evaluating their impact on model performance. These insights lay a foundation for further investigation and innovation in transfer learning and neural network architecture design.

5 Conclusion

This study explores transfer learning techniques and architectural modifications in deep learning models, focusing on image classification tasks. Demonstrated the practicality of employing transfer learning with the ResNet50 architecture, offering a robust means of adapting pre-trained models to specific tasks. Through systematic modifications to the base architecture, conducted a comparative analysis of different architectural variations, shedding light on their impact on model performance.

Additionally, the evaluation of training and testing accuracies reveals insights into the generalization capabilities of derived models compared to the baseline ResNet50. The observed reduction in overfitting and improved testing accuracies suggest the introduced modifications positively contribute to generalization. Furthermore, the detailed description of architectural variations provides a blueprint for designing tailored neural network architectures. The inclusion of a residual network architecture highlights the potential benefits of advanced architectural concepts, contributing to the advancement of deep learning for image classification.

6 Future Works

The study emphasizes the significance of exploring alternative transfer learning techniques and architectures beyond ResNet50 for image classification tasks. While ResNet50 is utilized in this research, future investigations could delve into diverse methodologies to ascertain their efficacy.

Moreover, deep learning models proficiency is intrinsically linked to the richness and diversity of training data. Thus, forthcoming endeavors could prioritize enlarging the Intel dataset or integrating supplementary datasets encompassing various scenes and environmental conditions. This augmentation would fortify the models capacity to generalize effectively across an extensive array of real-world scenarios.

Furthermore, there is potential for enhanced model performance through continued experimentation with architectural modifications, surpassing those delineated in the present study. These endeavors hold promise for refining the efficacy and versatility of deep learning models in image classification tasks.

References

[1] G. Cheng, X. Xie, J. Han, L. Guo and G. -S. Xia, "Remote Sensing Image Scene Classification Meets Deep Learning: Challenges, Methods, Benchmarks, and Opportunities," in IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 13, pp. 3735-3756, 2020, doi: 10.1109/JSTARS.2020.3005403.

[2] Delu Zeng, Minyu Liao, Mohammad Tavakolian, Yulan Guo Bolei Zhou, Dewen Hu, Matti Pietikainen, and Li Liu, "Deep Learning for Scene Classification: A Survey".

[3] ADITYA VAILAYA , ANIL JAIN , HONG JIANG ZHANG , "ON IMAGE CLASSIFICATION: CITY IMAGES VS. LANDSCAPES".

[4] Kaiming He; Xiangyu Zhang; Shaoqing Ren; Jian Sun, "Deep Residual Learning for Image Recognition,". .

[5] Jiazhi Liang, ”Image classification based on RESNET”

[6] Kanellopoulos & G. G. Wilkinson, ”Strategies and best practice for neural network image classification”.

[7] Robert M. Haralick; K. Shanmugam; Its’Hak Dinstein, ”Textural Features for Image Classification”.

[8] Leiyu Chen, Shaobo Li,Jing Yang, Sanlong Jiang and Yanming Miao, ”Review of Image Classification Algorithms Based on Convolutional Neural Networks”.

[9] Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, “Electron spectroscopy studies on magneto-optical media and plastic substrate interface,” IEEE Transl. J. Magn. Japan, vol. 2, pp. 740–741, August 1987 [Digests 9th Annual Conf. Magnetics Japan, p. 301, 1982].

[10] Fei Wang, Mengqing Jiang, Chen Qian, Shuo Yang, Cheng Li, Honggang Zhang, Xiaogang Wang, Xiaoou Tang; ”Residual Attention Network for Image Classification” Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 3156-3164

[11] Chen Jianfei; Zhu Changming, “Research on Image Recognition Based on Improved ResNet,”.

[12] Nair, Kasthuri AS, et al. ”Deep learning based approaches for accurate automated nuclei segmentation of Pap smear images.” 2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT). IEEE, 2023.

[13] Anand, Arundhathi, and Vinitha Panicker. ”Transformer-based Multiclass Classification of Cervicogram Images for Improved Cervical Cancer Detection.” 2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT). IEEE, 2023.

[14] Alisha, S., and Vinitha Panicker. ”Cervical Cell Nuclei Segmentation On Pap Smear Images Using Deep Learning Technique.” 2022 IEEE 3rd Global Conference for Advancement in Technology (GCAT). IEEE, 2022.

[15] Nayar, Sanjana, J. Vinitha Panicker, and Jyothisha J. Nair. ”Deep learning based model for multi-class classification of cervical cells using pap smear images.” 2022 IEEE 7th International conference for Convergence in Technology (I2CT). IEEE, 2022.

[16] Dhanya, S., and J. Vinitha Panicker. ”Detecting and rectifying adversarial images dealt by deep learning models.” 2021 5th International Conference on Electrical, Electronics, Communication, Computer Technologies and Optimization Techniques (ICEECCOT). IEEE, 2021.

[17] N. Aloysius and M. Geetha, ”A review on deep convolutional neural networks,” 2017 International Conference on Communication and Signal Processing (ICCSP), Chennai, India, 2017, pp. 0588-0592, doi: 10.1109/ICCSP.2017.8286426.