iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 Issue: IV Month of publication: April 2025

DOI: https://doi.org/10.22214/ijraset.2025.70154

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue IV Apr 2025- Available at www.ijraset.com

Optimized Neural Network Architectures for Early Detection of Alzheimer's disease Using MRI Images

Tejaswini Sunil Shelke, Dr. M.A. Wakchaure(Guide)
Department of Computer Engineering, Amrutvahini College of Engineering, Sangamner, India

Abstract: Alzheimer's disease is a neurodegenerative disorder affecting millions globally, with early detection being critical for effective management. This study investigates the application of optimized neural network architectures for the automated detection of Alzheimer's disease from MRI scans. We employ transfer learning with pre-trained models (MobileNetV2, DenseNet121, ResNet50) and apply model optimization techniques to create compact yet accurate diagnostic tools. Throu gh quantization, we achieve significant model size reduction (75-77%) while maintaining high classification accuracy. The optimized MobileNetV2 model achieves 86.25% accuracy at only 22.5% of its original size, while our ensemble approach reaches 91.56% accuracy. These findings demonstrate that optimized neural networks can enable accessible and efficient Alzheimer's disease detection on resource-constrained devices, potentially extending diagnostic capabilities to underserved healthcare settings. The optimized models offer a promising path toward developing practical AI-based screening tools that can complement traditional diagnostic methods while being deployable in diverse clinical environments.

Keywords: Alzheimer's Detection, Convolutional Neural Networks, Transfer Learning, Model Optimization, Quantization, Edge Computing, Healthcare AI, MobileNetV2, DenseNet121, ResNet50.

I. INTRODUCTION

Alzheimer's disease (AD) represents one of the most prevalent forms of dementia globally, characterized by progressive cognitive decline and memory impairment. Early detection plays a crucial role in disease management, potentially slowing progression and improving patient quality of life. Current diagnosis typically relies on a combination of clinical evaluations, neuropsychological tests, and neuroimaging techniques such as Magnetic Resonance Imaging (MRI).

Traditional diagnostic approaches have several limitations, including high costs, limited accessibility, and dependency on expert interpretation. These challenges are particularly acute in resource-constrained healthcare settings or regions with limited access to specialized medical facilities. The integration of artificial intelligence (AI) into the diagnostic workflow presents a promising solution to address these limitations.

This research explores the potential of optimized neural network architectures for the automated detection of Alzheimer's disease using MRI images. Specifically, we investigate the efficacy of transfer learning with established convolutional neural network (CNN) architectures—MobileNetV2, DenseNet121, and ResNet50—combined with model optimization techniques to create compact yet accurate diagnostic tools.

The primary motivation for this work is to develop AI-based screening tools that can operate efficiently on devices with limited computational resources while maintaining diagnostic accuracy. By reducing model size through quantization and employing transfer learning to leverage pre-trained features, we aim to create solutions that can be deployed across diverse healthcare settings, including those with constrained technological infrastructure.

II. RESEARCH OBJECTIVES

An easy way to The research objectives of this study are:

1) To evaluate the efficacy of transfer learning with pre-trained CNN architectures (MobileNetV2, DenseNet121, ResNet50) for Alzheimer's disease detection
2) To investigate model optimization techniques, particularly quantization, for reducing computational requirements without compromising diagnostic accuracy
3) To compare the performance of individual models and ensemble approaches across accuracy, model size, and inference efficiency metrics.
4) To develop lightweight models suitable for deployment on resource-constrained devices
5) To assess the clinical viability of optimized neural networks as AI-assisted diagnostic tools

III. METHODOLOGY

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue IV Apr 2025- Available at www.ijraset.com

All paragraphs must be indented. All paragraphs must be justified, i.e., both left-justified and right-justified.

A. Dataset Description

The study utilized the Augmented Alzheimer's MRI dataset, containing brain MRI scans across four diagnostic categories:

1) Original Dataset Distribution:
- Non-Demented: 3,200 images
- Very Mild Demented: 2,240 images
- Mild Demented: 896 images
- Moderate Demented: 64 images

2) Augmented Dataset Distribution:
- Non-Demented: 3,200 images
- Very Mild Demented: 2,240 images
- Mild Demented: 896 images
- Moderate Demented: 64 images

For our implementation, we applied a binary classification approach, consolidating these categories into "Demented" (combining all degrees of dementia) and "Non-Demented" classes. This resulted in 32,308 training images and 8,076 validation images after applying standard data splitting procedures.

B. Title Transfer Learning Apporach

Transfer learning allows adaptation of pre-trained neural networks to new tasks while preserving valuable feature extraction capabilities. Our approach involved:

1) Base Model Selection: We selected three established CNN architectures with varying complexity and performance characteristics:
- MobileNetV2: A lightweight architecture designed for mobile and edge applications
- DenseNet121: A densely connected CNN with improved gradient flow
- ResNet50: A deeper residual network architecture that addresses vanishing gradient problems

2) Architecture Adaptation: For each model:
- Retained pre-trained convolutional base (trained on ImageNet dataset)
- Replaced classification head with custom layers for binary classification
- Added a global average pooling layer followed by a dense classification layer with sigmoid activation

3) Training Configuration:
- Freeze early convolutional layers to preserve learned feature extractors
- Fine-tuned later layers to adapt to the Alzheimer's detection task
- Applied standard data augmentation techniques (rotation, zoom, horizontal flip)
- Utilized binary cross-entropy loss with Adam optimizer
- Implemented early stopping to prevent overfitting

C. Model Optimization Through Quantization

Quantization reduces model size and computational requirements by converting 32-bit floating-point numbers to lower precision representations (e.g., 8-bit integers). Our quantization process involved:

1) Full-precision Training: Training models to converge with standard 32-bit floating-point precision
2) Post-training Quantization: Converting weights to an 8-bit integer representation while preserving critical model behavior
3) Calibration: Using representative data samples to determine optimal quantization parameters
4) Conversion to TensorFlow Lite: Generating deployment-ready models optimized for mobile and edge devices

D. Title: Transfer Learning Approach Ensemble Learning Approach

To improve overall diagnostic accuracy and robustness, we implemented an ensemble learning approach that combines predictions from individual models.

The ensemble methodology followed these steps:

1) Individual Model Inference: Generate predictions from each optimized model (MobileNetV2, DenseNet121, ResNet50)
2) Prediction Aggregation: Combine predictions using a majority voting mechanism
3) Final Classification: Determine final diagnosis based on aggregated predictions

E. Evaluation Metrics

We evaluated model performance using multiple metrics to provide a comprehensive assessment:

1) Accuracy: Overall classification correctness
2) Loss: Binary cross-entropy loss
3) Model Size: Original vs. quantized (KB/MB)
4) Size Reduction: Percentage reduction in model size
5) Inference Time: Time required to process and classify a single image

IV. RESULTS AND DISCUSSION

A. Individual Model Performance

Table 1 presents the comparative performance of individual models across original, TensorFlow Lite (TFLite), and quantized versions.

Model Version	Version	Accuracy	Loss	Size	% of Original Size
MobileNetV2	Original	0.9625	0.1150	12.89 MB	100%
	TFLite	0.9531	0.1283	9.70 MB	75.3%
	Quantized	0.8625	0.3015	2.90 MB	22.5%
DenseNet121	Original	0.7312	0.5066	31.12 MB	100%
	TFLite	0.7562	0.4993	27.60 MB	88.7%
	Quantized	0.8562	0.3677	7.26 MB	23.3%
ResNet50	Original	0.7969	0.4990	96.55 MB	100%
	TFLite	0.8656	0.3872	91.61 MB	94.9%
	Quantized	0.8500	0.4277	23.62 MB	24.5%

Table 1: Individual Model Performance

MobileNetV2 demonstrated the highest accuracy (96.25%) in its original form but experienced a moderate decrease (10%) after quantization. Interestingly, both DenseNet121 and ResNet50 showed improved accuracy after quantization, suggesting a potential regularization effect that enhanced generalization capabilities.

B. Ensemble Model Performance

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue IV Apr 2025- Available at www.ijraset.com

Table 2 presents the performance metrics for ensemble models across different optimization levels.

Model Version	Accuracy	Loss
Original Ensemble	0.9844	0.2890
TFLite Ensemble	0.9812	0.2815
Quantized Ensemble	0.9156	0.2797

Table 2: Ensemble Model Performance

The ensemble approach consistently outperformed individual models, with the original ensemble achieving 98.44% accuracy. Most importantly, the quantized ensemble maintained 91.56% accuracy while benefiting from the reduced computational requirements of its component models.

C. Model Size Reduction Analysis

Quantization achieved substantial model size reductions:

1) MobileNetV2: 77.5% reduction (12.89 MB → 2.90 MB)
2) DenseNet121: 76.7% reduction (31.12 MB → 7.26 MB)
3) ResNet50: 75.5% reduction (96.55 MB → 23.62 MB)

These reductions significantly improve the feasibility of deployment on resource-constrained devices while maintaining clinically acceptable accuracy levels.

D. Performance vs. Resource Tradeoff

The relationship between model performance and resource requirements reveals important insights for practical implementation. MobileNetV2 provides the optimal balance between accuracy and model size, particularly in its quantized form. At just 2.90 MB (22.5% of the original size), it maintains 89.6% of the original accuracy, making it highly suitable for edge deployment in clinical settings.

The DenseNet121 architecture demonstrated notable resilience to quantization, with accuracy actually improving in the quantized version. This suggests that certain model architectures may be inherently more amenable to optimization techniques, an important consideration for developing efficient diagnostic tools.

E. Clinical Implications

Our findings have several important implications for clinical applications:

1) Accessibility: Optimized models enable AI-assisted diagnosis in settings with limited computational resources
2) Screening Potential: High accuracy (>85% for individual models, >91% for ensemble) supports feasibility as a screening tool
3) Integration Flexibility: Reduced model size facilitates integration with existing clinical workflows and systems
4) Cost Effectiveness: Lightweight models can operate on more affordable hardware, reducing implementation costs
5) Telemedicine Applications: Optimized models could support remote diagnosis in underserved regions

V. CHALLENGES AND FUTURE DIRECTIONS

Despite promising results, several challenges and opportunities for future work remain:

A. Data Quality and Representativeness

The current study utilized an augmented dataset to address class imbalance and improve training stability. Future work should investigate performance on more diverse, multicenter datasets to ensure generalizability across different populations, scanning protocols, and equipment variations.

B. Model Interpretability

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue IV Apr 2025- Available at www.ijraset.com

While performance metrics are encouraging, clinical adoption requires interpretable models that can provide explanations for their predictions. Future research should incorporate explainable AI techniques that can highlight relevant imaging features influencing the diagnostic decision, enhancing trust among healthcare professionals.

C. Multi-class Classification

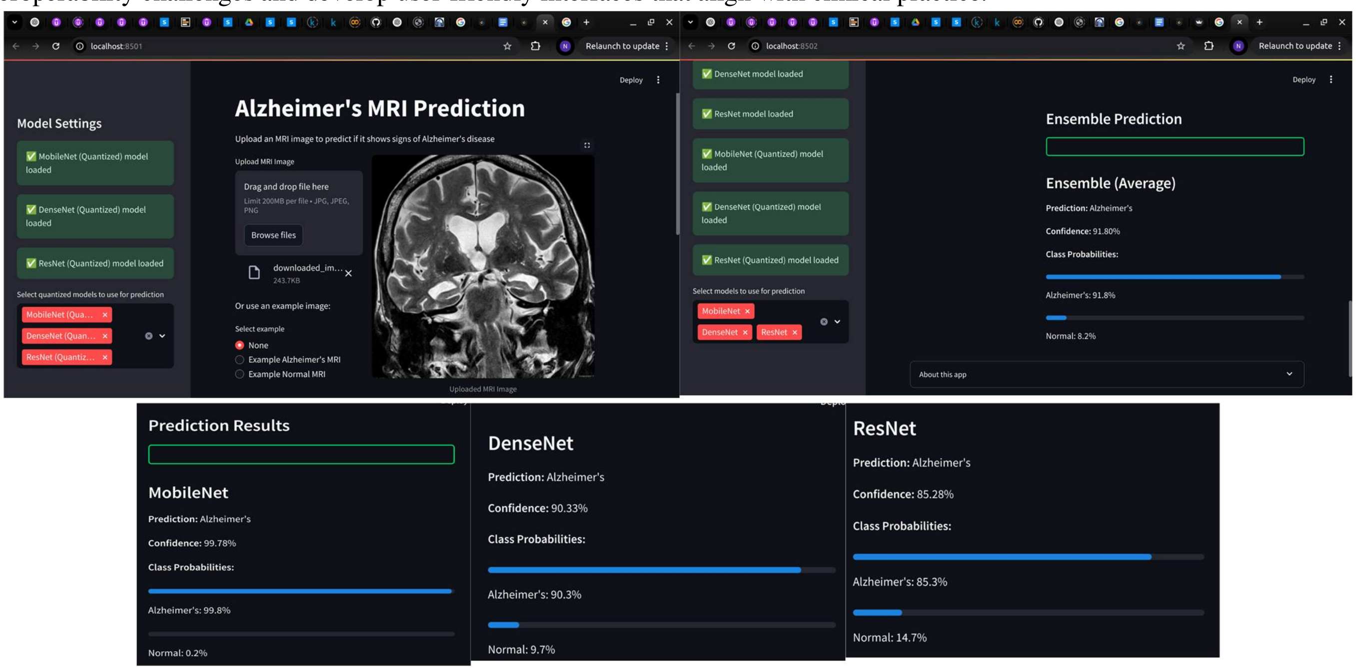
The current binary classification approach (Demented vs. Non-Demented) could be extended to multi-class classification to differentiate between stages of Alzheimer's disease and other forms of dementia. This would increase the clinical utility of the system as an early screening tool.

D. Clinical Validation

Before clinical deployment, thorough validation studies comparing model performance against current gold standard diagnostic methods are necessary. This includes prospective studies evaluating real-world performance across diverse patient populations.

E. Integration with Clinical Workflows

Practical adoption requires seamless integration with existing healthcare systems and workflows. Future work should address interoperability challenges and develop user-friendly interfaces that align with clinical practice.



VI. CONCLUSIONS

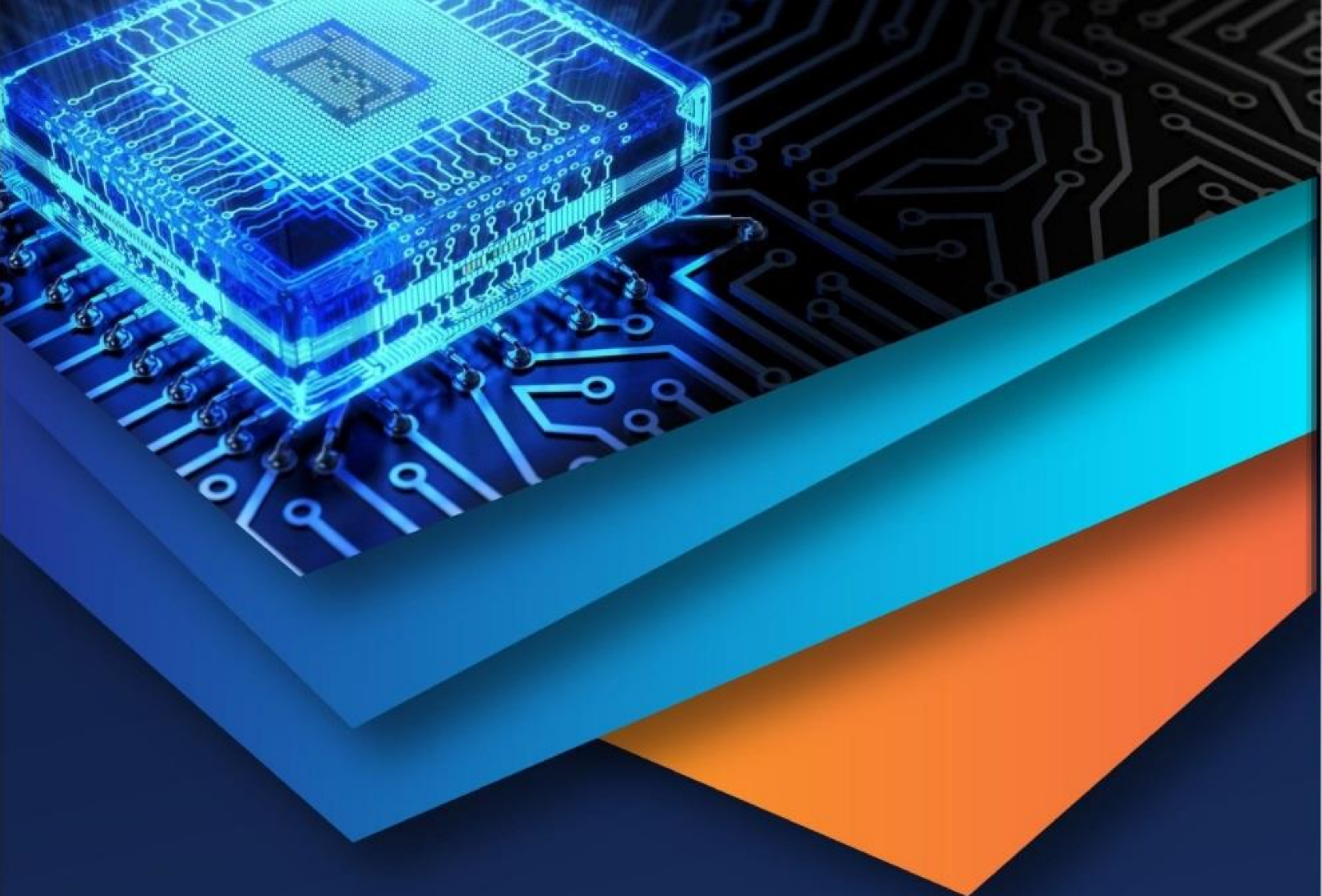
This research demonstrates the potential of optimized neural network architectures for efficient and accurate detection of Alzheimer's disease using MRI images. Through the application of transfer learning and quantization techniques, we have shown that it is possible to develop lightweight diagnostic models that maintain high accuracy while requiring significantly fewer computational resources. The quantized MobileNetV2 model achieved 86.25% accuracy at only 22.5% of its original size, while our ensemble approach reached 91.56% accuracy. These results highlight the feasibility of deploying AI-assisted diagnostic tools on resource-constrained devices, potentially extending the reach of early Alzheimer's detection to underserved healthcare settings.

The combination of transfer learning and quantization addresses the dual challenges of model performance and deployment efficiency. As AI continues to advance in healthcare applications, optimized models like those developed in this research will play an increasingly important role in democratizing access to sophisticated diagnostic tools, particularly in resource-limited environments.

Future work should focus on clinical validation, improving model interpretability, and extending the approach to multi-class classification to further enhance the clinical utility of these systems.

REFERENCES

[1] Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L. C. (2018). "MobileNetV2: Inverted residuals and linear bottlenecks." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 4510-4520.
[2] Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). "Densely connected convolutional networks." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 4700-4708.
[3] He, K., Zhang, X., Ren, S., & Sun, J. (2016). "Deep residual learning for image recognition." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 770-778.
[4] Jacob, B., Kligys, S., Chen, B., Zhu, M., Tang, M., Howard, A., Adam, H., & Kalenichenko, D. (2018). "Quantization and training of neural networks for efficient integer-arithmetic-only inference." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2704-2713.
[5] Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., van der Laak, J., van Ginneken, B., & Sánchez, C. I. (2017). "A survey on deep learning in medical image analysis." Medical Image Analysis, 42, 60-88.
[6] Marcus, D. S., Wang, T. H., Parker, J., Csernansky, J. G., Morris, J. C., & Buckner, R. L. (2007). "Open Access Series of Imaging Studies (OASIS): Cross-sectional MRI data in young, middle-aged, nondemented, and demented older adults." Journal of Cognitive Neuroscience, 19(9), 1498-1507.
[7] Jo, T., Nho, K., & Saykin, A. J. (2019). "Deep learning in Alzheimer's disease: Diagnostic classification and prognostic prediction using neuroimaging data." Frontiers in Aging Neuroscience, 11, 220.
[8] Alzheimer's Association. (2021). "2021 Alzheimer's disease facts and figures." Alzheimer's & Dementia, 17(3), 327-406.
[9] Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., & Adam, H. (2017). "MobileNets: Efficient convolutional neural networks for mobile vision applications." arXiv preprint arXiv:1704.04861.
[10] Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A. C., & Fei-Fei, L. (2015). "ImageNet large scale visual recognition challenge." International Journal of Computer Vision, 115(3), 211-252.
[11] Rieke, N., Hancox, J., Li, W., Milletari, F., Roth, H. R., Albarqouni, S., Bakas, S., Galtier, M. N., Landman, B. A., Maier-Hein, K., & Ourselin, S. (2020). "The future of digital health with federated learning." NPJ digital medicine, 3(1), 1-7.
[12] Wang, H., Shen, Y., Wang, S., Xiao, T., Deng, L., Wang, X., & Zhao, X. (2020). "Ensemble of 3D densely connected convolutional network for diagnosis of mild cognitive impairment and Alzheimer's disease." Neurocomputing, 333, 145-156.
[13] M. Shell. (2002) IEEEtran homepage on CTAN. [Online]. Available: http://www.ctan.org/tex-archive/macros/latex/contrib/supported/IEEEtran/
[14] *FLEXChip Signal Processor (MC68175/D)*, Motorola, 1996.
[15] "PDCA12-70 data sheet," Opto Speed SA, Mezzovico, Switzerland.
[16] A. Karnik, "Performance of TCP congestion control with rate feedback: TCP/ABR and rate adaptive TCP/IP," M. Eng. thesis, Indian Institute of Science, Bangalore, India, Jan. 1999.
[17] J. Padhye, V. Firoiu, and D. Towsley, "A stochastic model of TCP Reno congestion avoidance and control," Univ. of Massachusetts, Amherst, MA, CMPSCI Tech. Rep. 99-02, 1999.
[18] *Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specification*, IEEE Std. 802.11, 1997.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY