International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING

ISSN:2147-6799 www.ijisae.org Original Research Paper

Comparative Analysis of Deep learning Models for Various Optimizer Embedded with Gradient Centralization

[1]Vertika Agarwal, [2]Dr. M. C. Lohani, [3]Dr. Ankur Singh Bist

Submitted: 05/12/2023 **Revised**: 22/01/2024 **Accepted**: 29/01/2024

Abstract: Gradient Centralization (GC)emerges out an powerful optimization technique in area of Deep Convolutional neural network. It shows remarkable improvement in the execution time of deep learning models and opens up the scope of analyzing gradient vector to improve optimizer performance. It directly works upon the gradients and centralizes the gradient vector to have zero mean. One of the key factors which drives the attention of researchers is its embedding factor which allow its functionality to be explored with existing DNN optimizer. Our research works draws out individual and comparative analysis of GC embedded with RMS prop (Root Mean Square Propagation), Adam, Adagrad and Adadelta for three deep learning models: Mobile net, Nasnet and Densenet 201. Experiments are carried out with lung disease dataset. Highly motivating results are achieved through this embedding and accuracy of models has been enhanced up to 99%. Improved trends are also projected for Loss and Execution time.

Keywords: *Deep learning models, Densenet, Gradient Centralization, Mobile net, Nasnet*

1. Introduction

Optimizer plays an pertinent role in enhancing the performance of deep learning models. An optimizer modifies the weights of a neural network through various techniques. As a result, it aids in decreasing total loss and raising precision. A deep learning model typically has millions of parameters, making the task of selecting the proper weights for the model, a challenging task. Thus ,Selection of Optimizer algorithm for deep learning models is a crucial task which will affects the behaviour of model to a greater extent. Various deep learning optimizers like Adam, Rms prop,Adagrad, Adadelta with specific improvements to learning parameters are used over a period of time to accelerate the efficiency of models.

RMS prop (El Shamy et al.,2023)(Root Mean Square Propagation) is an optimization algorithm commonly used in machine learning and neural network training. It is an adaptive learning rate optimization algorithm designed to address some of the limitations of traditional gradient descent methods. RmsProp helps in training models more efficiently by adjusting the learning rate for each parameter based on the recent history of gradients.During each iteration, the current gradient of each parameter is divided by the square root of the exponential moving average of past squared gradients. This normalization helps prevent the learning rate from becoming too large or too small, which can result in slow convergence or divergence.RmsProp is particularly effective for optimizing models with sparse data or noisy gradients. Adam (Zhang et al.,2018)combines the benefits of both adaptive learning rate methods like RmsProp and momentum methods like stochastic gradient descent with momentum (SGD+Momentum).Adam's combination of adaptive learning rates and momentum makes it well-suited for a wide range of optimization problems. It can adaptively adjust the learning rates for different parameters based on their recent gradients while also incorporating momentum to help escape local minima and accelerate convergence. Adagrad (Okewu et al.,2019) is particularly effective in scenarios where some features have sparse gradients or require different learning rates for convergence. It adapts the learning rate for each parameter based on the historical information of gradients. Adagrad's main advantage is its ability to automatically adapt the learning rates for each parameter, which can be beneficial when dealing with features that have diverse scales or when some features require more or less aggressive updates. However, one limitation of Adagrad is that the learning rates tend to shrink over time due to the accumulating squared gradients, which can lead to very small updates and slow convergence.Adadelta(Okewu et al.,2019) is an optimization algorithm that addresses some of the limitations of the Adagrad optimizer, particularly the issue of diminishing learning rates over time. Adadelta adjusts the learning rates adaptively without explicitly accumulating all the past squared gradients, making it a more memory-efficient alternative to Adagrad. It also eliminates the need for a manual setting of the initial learning rate.

Gradient centralization technique which was introduced by (Yong et al.,2020) are providing benchmark results in improving the optimization technique .It is applied during the training of neural networks to improve convergence and enhance generalization performance. It focuses on the

[1,2,3] *Graphic era hill university Bhimtal*

Email id- vertika.agarwal21@gmail.com

gradients of the model's parameters, specifically by centering the gradients before using them to update the model's weights during the optimization process. This technique was introduced to mitigate the negative effects of large gradient magnitudes and speed up training.The goal of Gradient Centralization is to encourage the optimization process to focus on the direction of the gradients while reducing the impact of the magnitude of gradients. This can help with better convergence by reducing the chances of diverging due to large gradients and potentially lead to improved generalization on unseen data.

Our research work explores the integration of Gradient centralization with various optimizers on Lung disease dataset through three deep learning models Mobile net,Densenet 201 and Nasnet.Section 2 illustrates the related work . Section 3 describes the proposed integration of Optimization technique with various optimizers for deep learning models.Section 4 tabulates the experimental result of individual performance enhancement of deep learning models and and comparison of enhanced models. Section 5 winds the research paper with Future work and conclusion.

2. Related work

Researchers are continuously striving to improve the performance of Deep learning models by exploring its various dimensions related to its parameters, loss functions and optimization strategies.Efforts are made by Elshamy to improve the efficiency of RMS prop optimization algorithm(Elshamy et al.,2023) by adding a step that calculates the Nestrove for a next point, with respect to the average of the past squared gradients for the current point and called it as NRMSprop.Datasets like Fashion-MNIST, CIFAR-10 and Tiny-ImagNet datasets have been used and accuracy has been elevated to 97% from 86%.Z.Zhang proposed ND-ADAM (Z.Zhang et al.,2018)normalized direction-preserving Adam which enables more precise control of the direction and step size for updating weight vectors, and significantly improves generalization performance.Okewu performs the experimental evaluation of Adadelta, Adagrad,RMS prop and SGD over MNSIT dataset and concludes the accuracy of Adadelta as (0.9970) followed by Adam (0.9947), RMS Prop (0.9946), Adagrad (0.9938), and SGD (0.9772) and loss functions as Adadelta (0.0095) followed by Adam (0.0152), Adagrad (0.0220), RMS Prop (0.0223), and SGD (0.0736)(Okewu et al.,2019).Some of the researchers also proposes new optimization technique which significantly overcomes the drawback of traditional optimization technique .R.Dubey et al.,2020 proposes Diffgrad where the step size is adjusted for each parameter to have a larger step size for faster gradient changing parameters and a lower step size for lower gradient changing parameters. The convergence analysis is done using the regret bound approach of the online learning framework. Experiments are carried out over CIFAR 10 and CIFAR 100 using Resnet model and it outperforms Adagrad, Adadelta ,RMS prop and Adam.Comparative analysis for various optimizer has also been done by researchers for specific applications to analyse the best performing optimizer.Yaqub et al.,2020provides a comprehensive comparative analysis of popular optimizers of CNN namely Adaptive Gradient (Adagrad), Adaptive Delta (Ada Delta), Stochastic Gradient Descent (SGD), Adaptive Momentum (Adam), Cyclic Learning Rate (CLR), Adaptive Max Pooling (Ada max), Root Mean Square Propagation (RMS Prop), Nesterov Adaptive Momentum (Nadam), and Nesterov accelerated gradient (NAG) on the BraTS2015 data set.Adam optimizer achieved the highest accuracy of 99.2%.Taqiet carried out experimental analysis of multi optimizer like TF-CNN, Adagrad, Proximal Adagrad, Adam, and RMS Prop for Alzheimer disease (AD) classification(Taqiet al.,201) .The result demonstrates that the loss value of the Adam and RMS Prop optimizers was lower than the Adagrad and Proximal Adagrad optimizers. The classification accuracy using Adam optimizer is 95.8%, while it reaches 100% when using RMS Prop optimizer.Babu et al.,2020 illustrates the superiority of Whole swarm Optimization, meta-heuristic Algorithm over RMS prop for cardiac disease analysis. Area of optimization algorithm is continuously evolving with researchers coming up with novel concept based techniques which further can enhance the performance of deep learning models.Yong proposes Gradient centralization technique(Yong et al.,2020) which works on updating the gradients rather than on weights and centralizes the gradient vector to have zero mean.Effective results are observed for image classification, fine-grained image classification, detection and segmentation after using Gradient centralization technique. Fuhl explores the usage of weight centralization with gradient centralization and batch normalization for residual blocks(Fuhl et al.,2020) .Remarkable results has been achieved for cifar 10 and cifar 100 dataset in terms of generalization and accuracy.Yong proposed Gradient Centralization technique which centralizes gradient vector rather than weights , to have zero mean(Yong et al.,2020) .It tremendously boost the generalization performance of model and thus elevates its performance.Zang carries out Facial recognition with APnet (Asymmetrical Pyramidal network) and employs SGDGC(Stochastic gradient Descent Gradient centralization)(Zang et al.,2021).Model outperforms all the single model methods and has comparable performance with model fusion methods. Sadu proposes a moment centralization-based SGD optimizer for CNNs and uses Adam, Radam, and Ada belief on benchmark CIFAR10, CIFAR100, and Tiny Image Net datasets for image classification(Sadu et al.,2023).Encouraging results are achieved via this approach. Roy explores gradient angular information of previous iterations to control the step size and called it as Angular Grad .thus optimization step becomes smoother with past predictions and hence achieved desirable results(Roy et al.,2021). Lv proposes focal loss in multi task learning module along with Gradient centralization method to stabilize the training process. Highly competitive results are observed(Lv et al.).

3. Proposed work

Our proposed work carries out comprehensive analysis of integration of Gradient Centralization technique with optimizers like Adam, Rms prop, Adagrad and Adadelta for 30 epochs.Lung disease dataset have been employed which consists of six different types of lung disease like Bacterial Pneumonia,Viral pneumonia, Covid,Normal Lung opacity and Tuberculosis .Dataset have been prepared from various repositories like Kaggle,GitHub e.t.c.As clean dataset plays a crucial role in determining the efficacy of any deep learning model ,so our dataset have been preprocessed with one of the emerging image preprocessing technique Real Esrgan which took around 4hrs with PTesla100 GPU [15][16].

REALESRGAN: It is an image processing technique which creates training pairs with more realistic deterioration and thus restores common low-resolution images. By inculcating a second order degradation process, it leads to the degradation that occurs in the real world. Utilizing spectral normalization along with U-Net discriminator, it improves discriminator quality and stabilizes training dynamics. Real complex degradation is a synthesis of many degradation mechanisms like those found in camera imaging systems, Internet transmission, and picture manipulation [17][18].

Our research work delves deep into four optimizer Adam, Adagrad ,Rms prop and Adadelta which are frequently employed in many deep learning models .These optimizers are integrated with Gradient centralization and comparative analysis is drawn out [19][20].

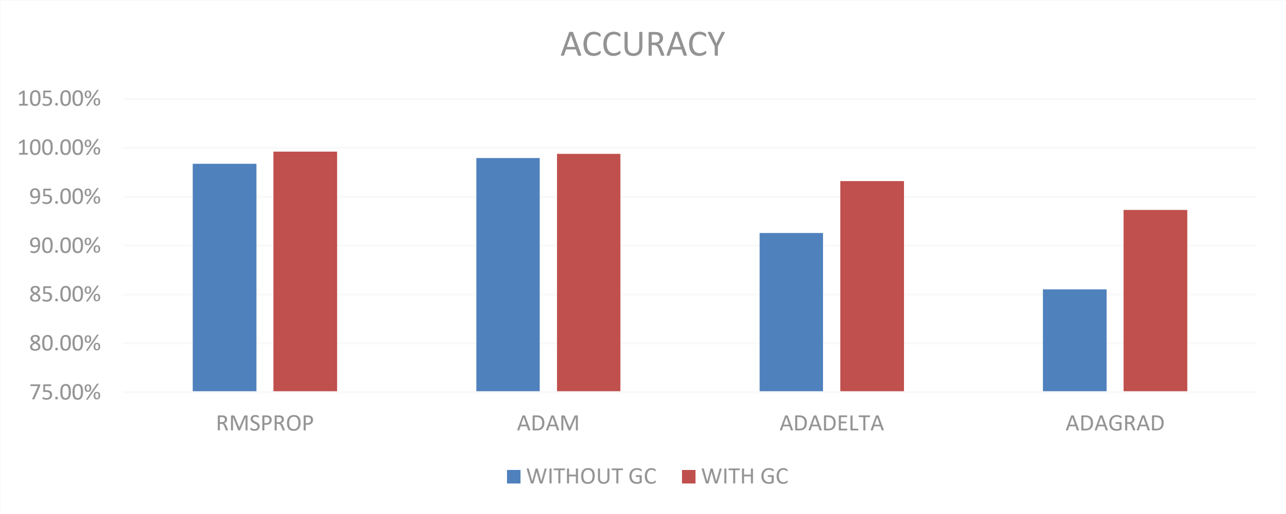
4. Experimental result and discussion

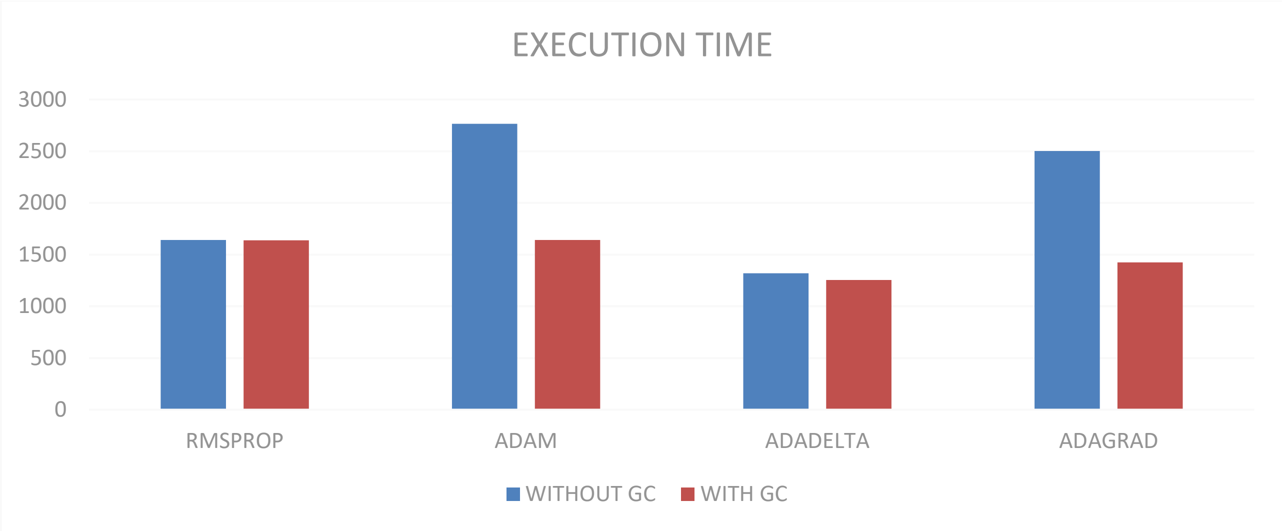
Mobile net,Densenet 201 and Nasnet models have been used to carry out the experimental analysis of proposed integration [21].

Mobile net : It has been observed that Adam, Adagrad ,RMS prop and Adadelta optimizer when embedded with Gradient centralization technique for Mobile net model ,performance get enhanced in terms of Accuracy, Loss and Execution time.Results are tabulated below in Table 1 and visualization is drawn in Figure 1 [22].

Table 1.

MOBILE NET		RMSPROP	ADAM	ADADELTA	ADAGRAD
WITHOUT GC	ACCURACY	98.376%	98.96%	91.29%	85.53%
	LOSS	.04750	.03752	0.22745	.3894
	EXECUTION TIME	1642.49	2764.024	1321.1050	2501.361
WITH GC	ACCURACY	99.60%	99.36%	96.60%	93.65%
	LOSS	.011211	.020320	.107188	0.1921
	EXECUTION TIME	1638.67	1640.8260	1256.095	1426.24





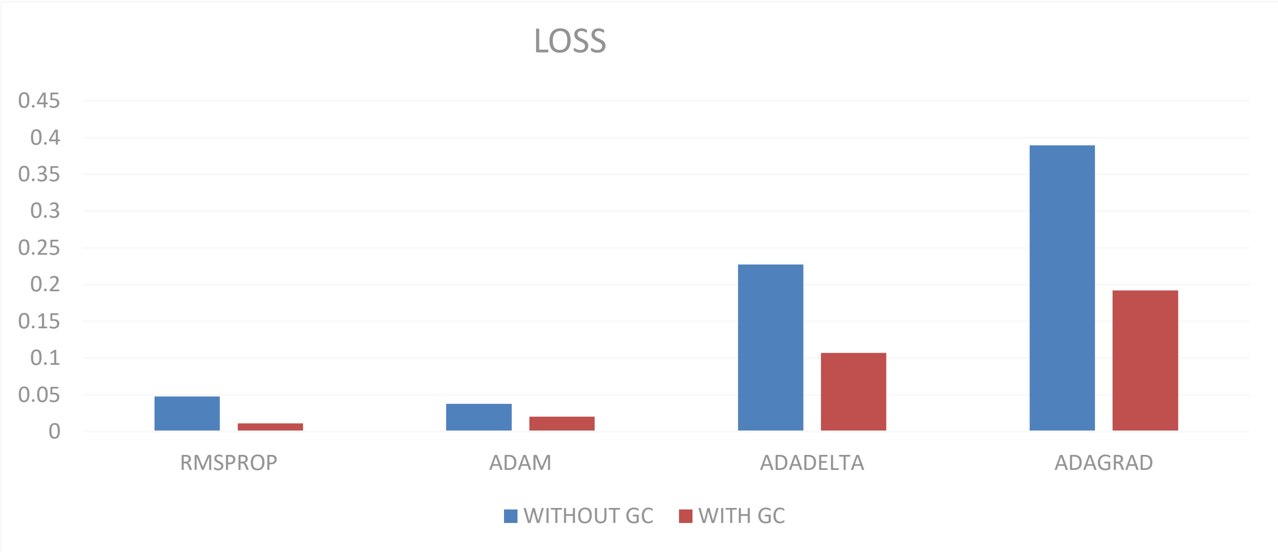


Fig 1. Visualization of Mobile net enhancement With GC integrated Optimizer

It has been observed that Mobile net model responds well for integration of GC with Adadelta and Adagrad as compare to RmsProp and Adam in terms of improving Accuracy while significant drop in execution time is recorded for Adam and Adagrad integration with GC.Losses are decreased for Adagrad and Rms prop GC integration. **So, Mobile net is exhibiting best results for Adagrad integration with GC.**

Nasnet: Results of Nasnet model for integration with GC with Adam,Rms prop,Adadelta and Adagrad are tabulated in Table 2 and visualization is exhibited in Fig 2.

Table 2.

NASNET		RMSPROP	ADAM	ADADELTA	ADAGRAD
WITHOUT GC	ACCURACY	94.39%	96.16%	25.036%	82.68%
	LOSS	0.1445	0.1148	1.9871	0.4931
	EXECUTION TIME	3382.466	3163.475	2462.688	1347.369
WITH GC	ACCURACY	97.21%	97.88%	36.05%	86.66%
	LOSS	0.0772	0.0558	1.649	0.3720
	EXECUTION TIME	2146.78	2204.06	1343.202	1413.257

Fig 2. Visualization of Nas net enhancement With GC integrated Optimizer

Experimental result for Nasnet brings out unusual result with Adadelta optimizer as model shows accuracy of 25% which is quite low .Though integration with GC enhances it but use of Adadelta optimizer with Nas net is showing degraded result.It has been observed that Rms prop is showing promising result with substantial increase in accuracy from 94% to 97% and significant drop in execution time and loss.**Hence Nasnet model exhibit remarkable performance with Rms prop integration with GC.**

Densenet:Results of Densenet model for integration with GC with Adam,Rmsprop ,Adadelta and Adagrad are tabulated in Table 3 and visualization is depicted in Fig 3.

DENSENET 201		RMSPROP	ADAM	ADADELTA	ADAGRAD
WITHOUT GC	ACCURACY	97.09%	98.62%	29.36%	88.14%
	LOSS	.0936	.05278	1.743	.36168
	EXECUTION TIME	2320.0614	1359.374	1380.199	1518.33
WITH GC	ACCURACY	99.016%	99.11%	45.05%	92.12%
	LOSS	.03119	.0311	1.433	.2484
	EXECUTION TIME	1387.515	1333.382	1369.206	1477.572

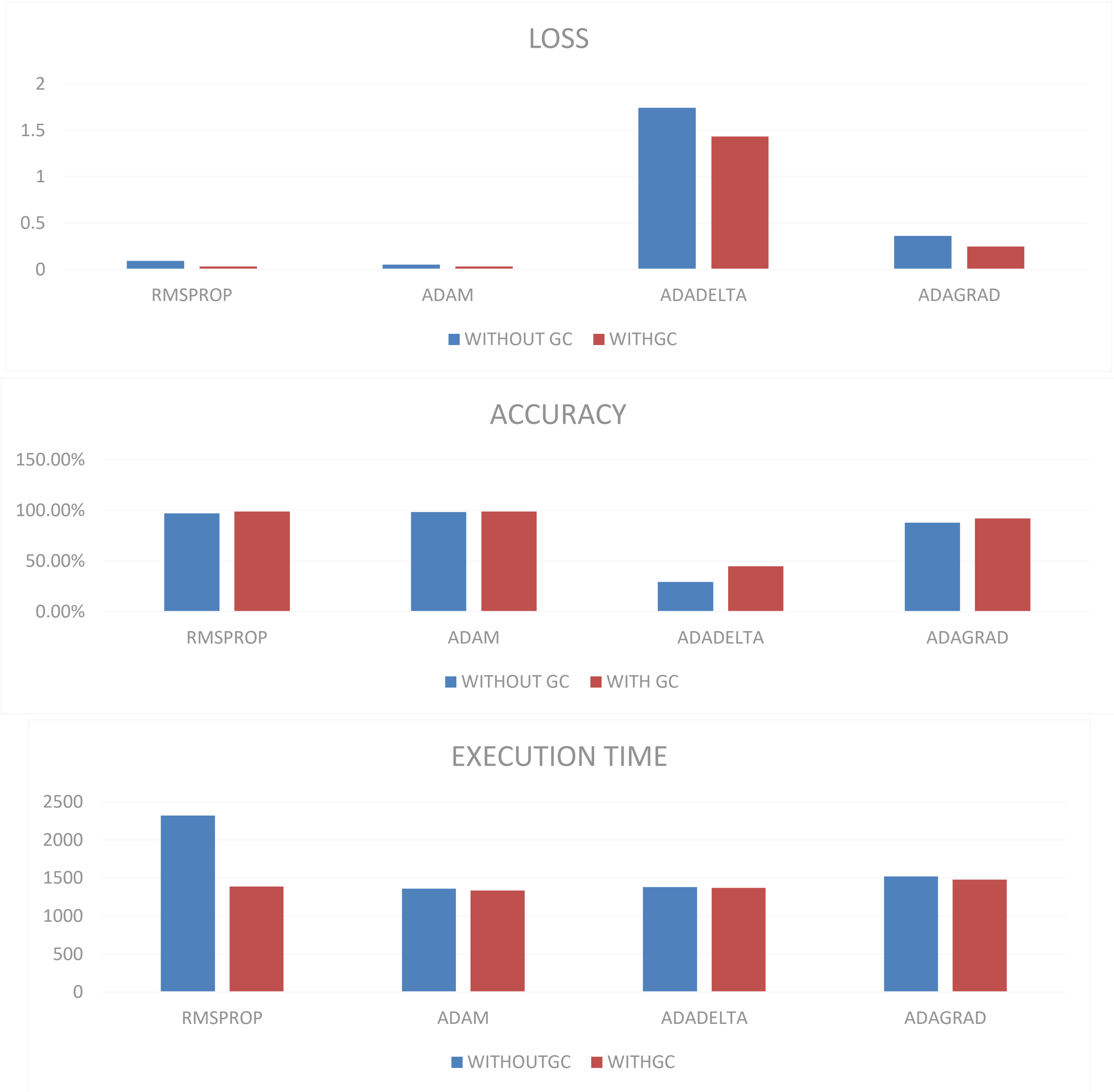


Fig 3. Visualization of Densenet enhancement With GC integrated Optimizer

Densenet model too presents inappropriate result for Adadelta optimizer. Significant Performance enhancement has been observed for Rms prop optimizer when embedded with GC.But Adam performs well with accuracy reached to nearly 99.11% and losses and execution time is also better as compared to other optimizer integration.

Comparison Of Models Based On Integration Of Gc With Various Optimizer

Our research work has taken three models Mobile net ,Densenet 201 and Nasnet for exploring the effect of GC integrated optimizer.Individual models shows considerable improvement with this approach. Our work also carries out the comparative Analysis of Models as which model is responding best for this integrated frame work. Models are compared for three factors: Accuracy,Loss and Execution time and their results are tabulated in Table 4,5 and 6 respectively and their visualization are shown in Fig 4,5 and 6 respectively.

Table 4: Accuracy Factor

MODEL	ADAM	ADADELTA	ADAGRAD
MOBILENET	99.360%	**96.606%**	93.65%
NASNET	97.88%	36.05%	86.66%
DENSENET	99.11%	45.05%	92.12%

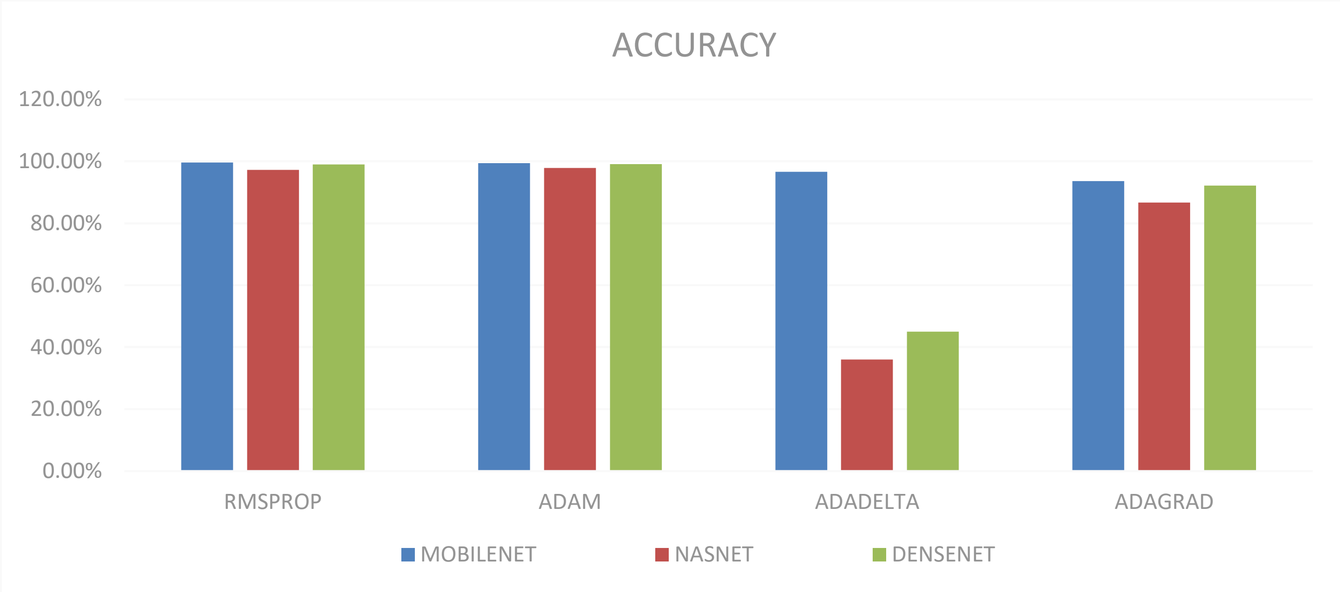


Fig 4. Comparative analysis of Models for Accuracy

From perspective of Accuracy ,Mobile net performs fairly well for all integrated optimizer and highest accuracy achieved is 99.606%.Adadelta which is not performing for other two models ,works quite well for Mobile net and exhibit accuracy of 96.60%.

Table 5:LOSS Factor

MODEL	RMSPROP	ADAM	ADADELTA	ADAGRAD
MOBILENET	**.01121**	.020320	.107188	.1921
NASNET	.0772	.0558	1.6499	.3720
DENSENET	.03119	.0311	1.433	.2484

Minimum loss is incurred with Rmsprop integration with GC for Mobile net which is .01121. Densenet too show as quite low losses for Rmsprop .Losses for Adadelta for Nasnet and Densenet is quite high comparatively .Adagrad show minimum losses for Mobile net.

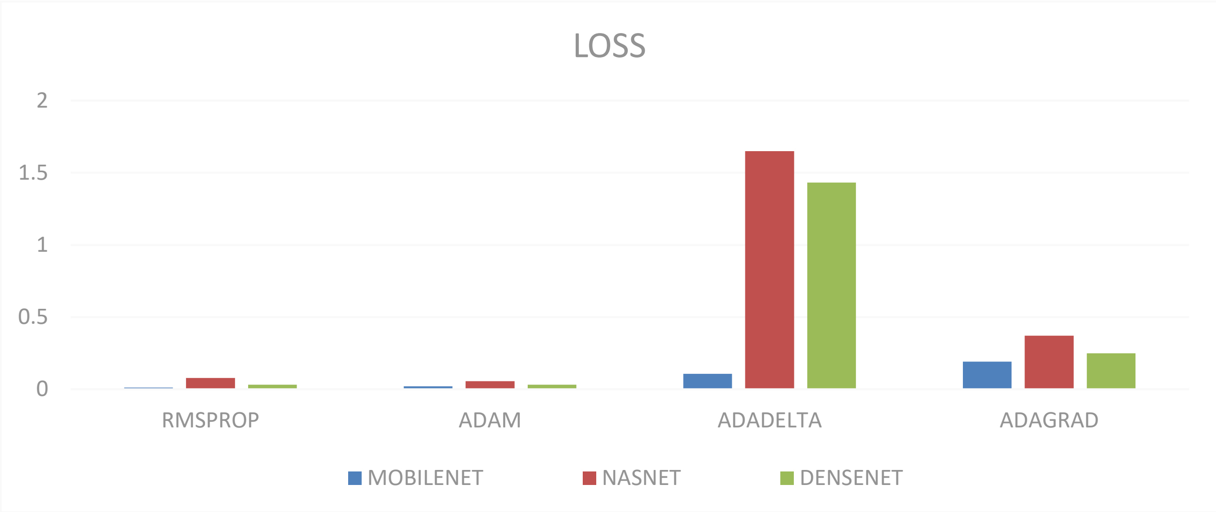


Fig 5. Comparative analysis of Models for Loss

Table 6: EXECUTION FACTOR

MODEL	RMSPROP	ADAM	ADADELTA	ADAGRAD
MOBILENET	1638.67	1640.8260	**1256.095**	1426.24
NASNET	2146.780	2204.0611	1343.202	1413.2571
DENSENET	1387.515	1333.382	1369.286	1477.572

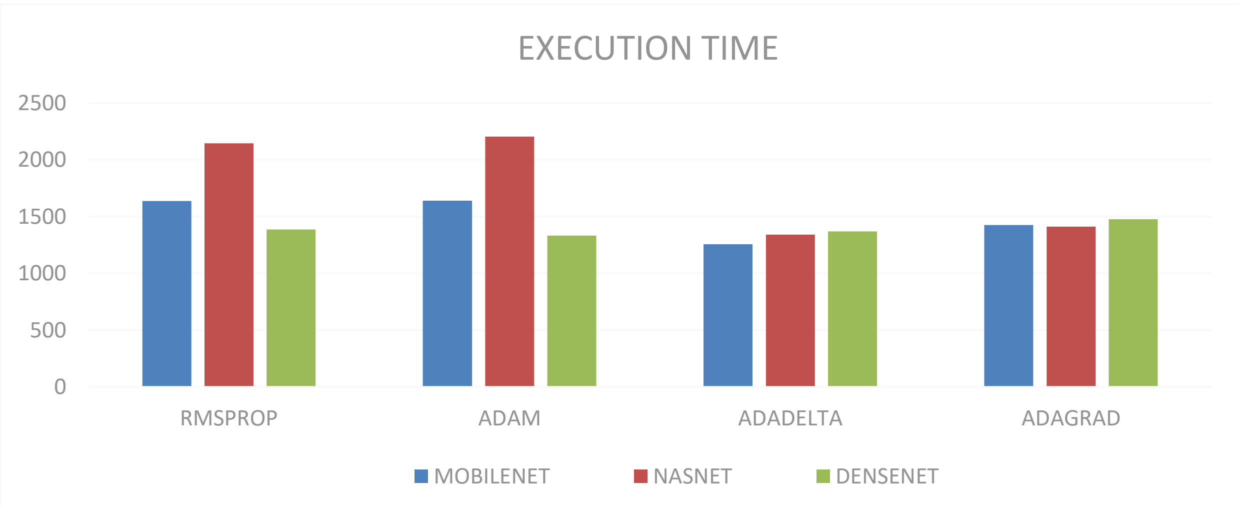


Fig 6. Comparative analysis of Models for Execution Time

Execution time is crucial factor for measuring the performance of any deep learning model.Minimum execution time has been observed for Adadelta integration for Mobile net which is 1256.095 sec.Nasnet shows dismal performance for Rms prop and Adam integration as compared to other optimizer.Adagrad show average performance for integration.

5. Conclusion and future work

Optimizer plays a very crucial role in converging any Deep learning algorithm as it will converges the model towards attaining an optimizing error and thus improves its performance. Our research explores the embedding of Gradient centralization technique ,an emerging Optimization technique with traditional optimizer Adam, Rmsprop,Adadelta and Adagrad for three deep learning models :Mobile net,Nasnet and Densenet 201and monitors its effect on Accuracy, Loss and Execution time. Experimental analysis of embedding clearly brings out an optimistic elevation for these three factors .Though there are variations in their response but almost each model show improving trends. Exceptions occurs for Densenet 201 model and Nasnet for Adadelta optimizer which exhibit very low accuracy and high losses.Evaluation of Individual enhancement of model is followed by comparative analysis of three models for the integrated frame work . Accuracy and Losses incurred achieved with Mobile net when GC is embedded with Rms prop is best which is 99.60%,.01121 respectively .Best execution time is shown by GC integrated Adadelta for Mobile net which is 1256.095sec.Future research can be carried out with other optimizer integration like SGD ,Gradient descent e.t.c with GC. We have employed Lung disease dataset in out research work .Other dataset like Retina, Skin and brain dataset can also be explored with this framework. Variation in learning rate can also be inculcated in further research to bring out the best parameter settings for a particular dataset.

Conflict of Interest

The author declares no conflict of interest.

References

[1] Elshamy, R., Abu-Elnasr, O., Elhoseny, M., & Elmougy, S. (2023). Improving the efficiency of RMSProp optimizer by utilizing Nestrove in deep learning. Scientific Reports, 13(1), 8814.

[2] Z. Zhang, "Improved Adam Optimizer for Deep Neural Networks," 2018 IEEE/ACM 26th International Symposium on Quality of Service (IWQoS), Banff, AB, Canada, 2018, pp. 1-2, doi: 10.1109/IWQoS.2018.8624183.

[3] Okewu, E., Adewole, P., Sennaike, O. (2019). Experimental Comparison of Stochastic Optimizers in Deep Learning. In: Misra, S., et al. Computational Science and Its Applications – ICCSA 2019. ICCSA 2019. Lecture Notes in Computer Science(), vol 11623. Springer, Cham. https://doi.org/10.1007/978-3-030-24308-1_55

[4] S. R. Dubey, S. Chakraborty, S. K. Roy, S. Mukherjee, S. K. Singh and B. B. Chaudhuri, "diffGrad: An Optimization Method for Convolutional Neural Networks," in IEEE Transactions on Neural Networks and Learning Systems, vol. 31, no. 11, pp. 4500-4511, Nov. 2020, doi: 10.1109/TNNLS.2019.2955777.

[5] Yaqub, M., Feng, J., Zia, M. S., Arshid, K., Jia, K., Rehman, Z. U., & Mehmood, A. (2020). State-of-the-art CNN optimizer for brain tumor segmentation in magnetic resonance images. Brain Sciences, 10(7), 427.

[6] A. M. Taqi, A. Awad, F. Al-Azzo and M. Milanova, "The Impact of Multi-Optimizers and Data Augmentation on TensorFlow Convolutional Neural Network Performance," 2018 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR), Miami, FL, USA, 2018, pp. 140-145, doi: 10.1109/MIPR.2018.00032.

[7] Babu, D. V., Karthikeyan, C., & Kumar, A. (2020, December). Performance analysis of cost and accuracy for whale swarm and RMSprop optimizer. In IOP Conference Series: Materials Science and Engineering (Vol. 993, No. 1, p. 012080). IOP Publishing.

[8] Yong, H., Huang, J., Hua, X., & Zhang, L. (2020). Gradient centralization: A new optimization technique for deep neural networks. In Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part I 16 (pp. 635-652). Springer International Publishing.

[9] Fuhl, W., & Kasneci, E. (2020). Weight and gradient centralization in deep neural networks. arXiv preprint arXiv:2010.00866

[10] Yong, H., Huang, J., Hua, X., & Zhang, L. (2020). Gradient centralization: A new optimization technique for deep neural networks. In Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part I 16 (pp. 635-652). Springer International Publishing.

[11] Zang, H., Foo, S. Y., Bernadin, S., & Meyer-Baese, A. (2021). Facial emotion recognition using asymmetric pyramidal networks with gradient centralization. IEEE Access, 9, 64487-64498.

[12] Sadu, S., Dubey, S. R., & Sreeja, S. R. (2023). Moment Centralization-Based Gradient Descent Optimizers for Convolutional Neural Networks. In Computer Vision and Machine Intelligence: Proceedings of CVMI 2022 (pp. 51-63). Singapore: Springer Nature Singapore.

[13] Roy S. K., Paoletti, M. E., Haut, J. M., Dubey, S. R., Kar, P., Plaza, A., & Chaudhuri, B. B. (2021). Angulargrad: A new optimization technique for angular convergence of convolutional neural networks. arXiv preprint arXiv:2105.10190.

[14] Lv, N., Xiang, X., Wang, X., Yang, J., & Abdein, R. (2022). Efficient person search via learning-to-normalize deep representation. Neurocomputing, 495, 169-177.

[15] Narayan, Vipul, et al. "7 Extracting business methodology: using artificial intelligence-based method." Semantic Intelligent Computing and Applications 16 (2023): 123

[16] Narayan, Vipul, et al. "A Comprehensive Review of Various Approach for Medical Image Segmentation and Disease Prediction." Wireless Personal Communications 132.3 (2023): 1819-1848.

[17] Mall, Pawan Kumar, et al. "Rank Based Two Stage Semi-Supervised Deep Learning Model for X-Ray Images Classification: AN APPROACH TOWARD TAGGING UNLABELED MEDICAL DATASET." Journal of Scientific & Industrial Research (JSIR) 82.08 (2023): 818-830.
[18] Narayan, Vipul, et al. "Severity of Lumpy Disease detection based on Deep Learning Technique." 2023 International Conference on Disruptive Technologies (ICDT). IEEE, 2023.
[19] Saxena, Aditya, et al. "Comparative Analysis Of AI Regression And Classification Models For Predicting House Damages İn Nepal: Proposed Architectures And Techniques." Journal of Pharmaceutical Negative Results (2022): 6203-6215.
[20] Kumar, Vaibhav, et al. "A Machine Learning Approach For Predicting Onset And Progression"“Towards Early Detection Of Chronic Diseases “." Journal of Pharmaceutical Negative Results (2022): 6195-6202.
[21] Chaturvedi, Pooja, Ajai Kumar Daniel, and Vipul Narayan. "Coverage Prediction for Target Coverage in WSN Using Machine Learning Approaches." (2021).
[22] Chaturvedi, Pooja, A. K. Daniel, and Vipul Narayan. "A Novel Heuristic for Maximizing Lifetime of Target Coverage in Wireless Sensor Networks." Advanced Wireless Communication and Sensor Networks. Chapman and Hall/CRC 227-242.