Exploring Model Optimization in Alzheimer's Disease Detection: A Study of Deep Learning Models, Optimizers, And Batch Sizes

Pranab Hazra[1], Mainak Dey[1], Sayandip Kumar[1],
Tushar Kanti Bera[2], Ashis Kumar Dhara[2]

[1]Narula Institute of Technology, Agarpara,
Kolkata, West Bengal, India.
[2]National Institute of Technology, Durgapur,
West Bengal, India.

Abstract. Alzheimer's disease is a major public health concern worldwide. Timely and accurate diagnosis is crucial for effective treatment and patient care. Recent advancements in deep learning have shown promise in improving Alzheimer's disease detection through the analysis of neuroimaging data, such as magnetic resonance imaging (MRI) and positron emission tomography (PET) scans. The deep learning models examined in this study include convolutional neural networks (CNNs). We rigorously analyze the effects of different optimizer algorithms, including Adam, RMSprop, and stochastic gradient descent (SGD). Additionally, we explore the influence of batch sizes during the model training process. One of the key contributions of this research is the in-depth examination of batch size selection in the context of AD detection. We demonstrate how batch size affects training dynamics, convergence speed, and memory requirements, shedding light on the trade-offs associated with different batch sizes. We employ rigorous evaluation metrics, including accuracy, sensitivity, specificity, and area under the receiver operating characteristic curve (AUC-ROC), to quantify the performance of our models.

In this paper, results showcase that no one-size-fits-all approach exists, as the optimal combination of model, optimizer, and batch size depends on the specific problem domain and dataset characteristics. This research paper serves as a reference point for researchers and practitioners seeking to maximize the accuracy of their deep-learning models. By systematically exploring the interplay between different factors in model optimization, we offer a nuanced perspective on achieving superior model performance across diverse tasks and datasets.

Keywords: Alzheimer’s Disease, Early Diagnosis, MRI, Convolutional Neural Network (CNN), Classification, Optimizer, Batch size, EfficientNetB0, Sequential, ResNet 18, ResNet 50, VGG19, Xception.

1 Introduction

Alzheimer's disease (AD) is a progressive neurodegenerative form of dementia. Early classification is crucial for slowing its progression. Our project employs MRI analysis to classify AD, but this method can be time-consuming and prone to misdiagnosis due to varied symptoms.

The field of medical imaging, particularly the utilization of magnetic resonance imaging (MRI), has emerged as a powerful tool for assisting in the early detection of

AD. Advanced machine learning techniques, particularly deep learning, have shown remarkable promise in automating the diagnosis of AD by analyzing neuroimaging data. However, the successful application of deep learning in AD detection is contingent upon various factors, including the choice of deep learning architectures, optimization techniques, and the crucial parameter: batch size.
This work delves into a comprehensive study aimed at shedding light on the intricate interplay between these critical components in the realm of AD diagnosis. By systematically investigating the effects of different deep learning architectures, optimization algorithms, and batch sizes, our research seeks to address fundamental questions that can significantly impact the performance and practical applicability of AD detection models.

Our study serves as a timely and essential contribution to the field of AD diagnosis [1], as it addresses the pressing need for robust, accurate, and efficient automated detection methods. The insights gleaned from our research may not only enhance the accuracy of AD diagnosis but also pave the way for the development of more interpretable and clinically relevant deep learning models.
In this paper, we chose several Deep Learning Models and we propose the architecture and the MRI image data and transfer into the various models and make the comparative study and see that the model EfficientNetB0 is given the highest accuracy and Adam optimizer is best for early detection of the Alzheimer's Diseases (AD).

2 State of The Art

The data leakage problems in Convolutional Neural Networks (CNNs) are used for Alzheimer's disease (AD) detection. Data leakage refers to a situation where information from the test set is inadvertently leaked into the training process, leading to overly optimistic evaluation results. In The proposed model builds upon the ADNI (Alzheimer's Disease Neuroimaging Initiative) model for AD detection, which serves as the basis [2]. The ADNI model is widely used and has been established as a benchmark for evaluating AD detection algorithms. The researchers employed a transfer learning approach, leveraging pre-trained CNN models trained on various benchmark datasets of CNN images. They specifically used Magnetic Resonance Imaging (MRI) data to predict the conversion rate from mild cognitive impairment (MCI) to AD, which can improve clinical diagnosis [3]. The maximum accuracy using a 2D Deep Convolutional Neural Network (DCNN) is possible with the help of MRI images. It is analyzed researchers used the CNN depth scaling in terms of increasing accuracy and efficiency. This data set model has three classes Alzheimer, mild very mild and moderate. This model has 99.89% accuracy [4].

The AD model represents different stages of Alzheimer's disease (AD). This classification is determined using the Broad Learning System and convolutional variants, which are computational techniques used in machine learning[5]. These tasks could include the classification of different stages of AD, prediction of disease progression, or identification of biomarkers associated with AD. The algorithms not only achieve higher accuracy but also require less training time compared to existing state-of-the-art algorithms [6]. This efficiency is beneficial as it reduces the

computational resources and time needed to train and deploy the models.

3 ResNet Model

Residual Network (as shown in Fig. 1) is a deep learning architecture that has significantly advanced in the field of computer vision and deep neural networks. It was designed to address the vanishing gradient problem that often occurs in very deep neural networks [7]. The key innovation behind ResNet is the introduction of residual blocks, which enable the training of extremely deep neural networks more effectively. In a residual block, the input to a layer is combined with the output of the layer, creating a "shortcut connection" or "skip connection"[8]. This allows the network to learn residual information, i.e., the difference between the input and output of a layer [9]. As a result, even as the network depth increases, gradients can flow more easily during training, mitigating the vanishing gradient issue.

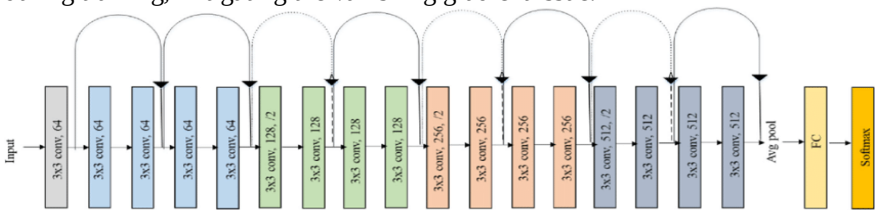


Fig.1. ResNet Architecture

4 VGG19 Model

VGG stands for Visual Geometry Group. It was proposed by Simonyan and Zisserman (2014), as a 19-layer Convolutional Neural Network consisting of 16 Convolutional Layers and 3 fully connected Convolutional Layers to categorize the images into 1000 object categories [10]. VGG19 is dependent on the ImageNet database containing 1 million pictures or images from 1,000 categories [11]. This is a very popular method of image classification as it uses multiple 73 filters in each convolutional layer. The architecture of VGG19 is shown in Figure 2. This shows that 16 layers of convolution are used for the feature extraction and the next 3 layers are used for classification of data [12]. The levels used for feature extraction are divided into five groups, each group followed by a max pooling level. Since CNN models compute large parameters later in feature extraction, a dimensionality reduction is required to minimize the feature vector size, as shown in Figure 2. Dimensionality reduction is done using locality-preserving projections followed by classification methods.

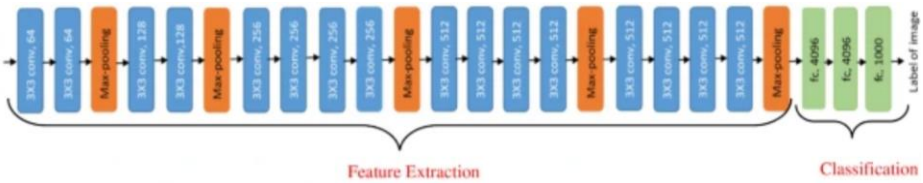


Fig.2. Architecture of VGG19 model

5 Xception Model

Xception is a deep Convolutional Neural Network architecture (as shown in Fig. 3)

with depth separable convolutions [13]. Developed by the Google researchers. Google presented an interpretation of starting modules in convolutional neural networks as a midway step between regular convolution and depth-separable convolution operations (depth convolution followed by point convolution). Against this background, folds that can be separated by depth can be understood as starting modules with the maximum number of towers [14]. This observation led them to propose a new deep convolutional neural network architecture inspired by Inception [15]. In this architecture, the Inception module is replaced by depth-separable convolutions.

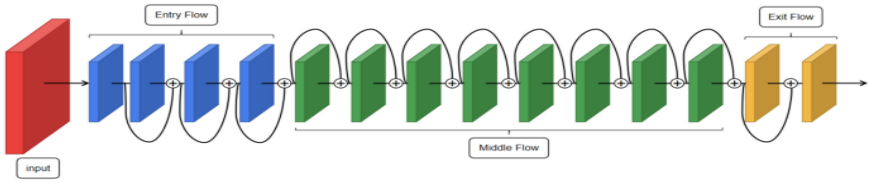


Fig.3. Architecture of Xception model

6 Methods

Collection of the dataset of brain images, such as MRI or PET scans, from individuals diagnosed with Alzheimer's disease and healthy individuals [16]. The dataset should include images from both early-stage Alzheimer's patients and control subjects. After that data preprocessing is needed. Preprocess the collected brain images to enhance their quality and remove any noise or artefacts [17]. This may involve skull stripping, image registration, normalization, and other preprocessing techniques. After data preprocessing it can be noticed that some data are missing, to overcome this problem data augmentation is needed. Augment the dataset by applying transformations such as rotation, scaling, and flipping to increase the diversity and variability of the training data. This helps the model generalize better and reduces the risk of overfitting. Then choose an appropriate deep-learning architecture for the classification task.

Train the selected Deep Learning (DL) model using the preprocessed and augmented dataset. Split the data into Training, Validation, and Testing sets to monitor the model's performance and prevent overfitting. Optimize the model's hyperparameters, such as learning rate, batch size, and regularization techniques, through experimentation and validation. In the Early Classification Approach, the goal is early detection, consider using a sliding window or patch-based approach to extract smaller regions of interest from the brain images. By focusing on specific brain regions relevant to Alzheimer's disease, the model can potentially identify early signs or biomarkers more accurately. Then evaluate the trained model's performance using appropriate evaluation metrics such as Accuracy, Precision, recall, and F1 score. Additionally, consider computing other evaluation measures specific to Alzheimer's disease classification, such as the area under the Receiver Operating Characteristic curve (AUC-ROC). Cross-validation is a very important one, so we have to perform cross-validation to measure the model's robustness and generalizability [18]. Additionally, if available, validate the trained model on an external dataset to validate its effectiveness in different populations and settings. Finally, explore techniques to

interpret and visualize the model's predictions and feature importance [19]. This can help provide insights into the model's decision-making process and identify regions of the brain that contribute most to Alzheimer's disease classification.

6.1 Experimental Data

The MRI process scans of 1200 patients were acquired from the Kaggle website (Kaggle kernels output Jeon woo park /alzheimer-detection-and-classification-98-7-acc p/path/to/test). Here first of all the 3D images in the .jpg extension are downloaded and preprocessed [20]. Then these brain images are marked as Non-demented, Mild Demented, Moderate Demented and Very Mild Demented. Figure 4 depicts the image slices which were obtained from the brain using Matplotlib which gave the Right angle, centre and left-angle views. These images were then augmented to increase the size. The transformations involved a zoom range of 0.07, width shift range of 0.07, hide shift range of 0.07 and horizontal flip [21]. In this manner, the image dataset was shaped for each of the four classes [22].

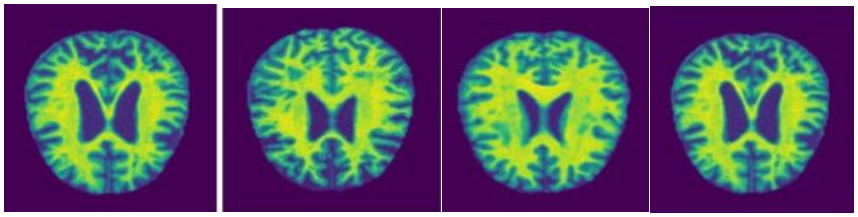


Fig.4. Moderate Demented (first), Mild Demented (second), Very Mild Demented (third), Non-Demented (fourth)

7 The Performance Evaluation

Here we focus on the study. The most general way for comparing algorithms is classification performance without focusing on a class. The empirical measures are used mostly and the accuracy cannot be distinguished between the number of accurate tables and the different types of class levels [23].

TP = True Positives: No. of actual examples predicted which are actually positive.
FP = False Positive: No. of actual examples that are predicted positively but are actually negative.
TN = True Negative: No. of actual examples which are predicted negative and also negative actually.
FN = False Negative: No. of actual examples which are predicted negative but are actually positive.

ACCURACY: The total number of records that are correctly classified by the classifier is referred to as accuracy [24]. Accuracy can also be explained as the percentage of test set tuples that are classified accurately by the model.

$$\boldsymbol{accuracy} = \frac{TP+TN}{TP+FP+FN+TN} \text{ X } 100\% \quad (1)$$

ROC CURVE: ROC stands on the Receiver Operating Characteristics curve which displays both the sensitivity and specificity of the test. The comparison between TPR (True Positive Rate) and FPR (False Positive Rate) is known as the ROC curve.
i.e., TPR = TP (TP+FN) and FPR = FP (FP+TN).

$$AUC = \int_0^1 t_{pr}(f_{pr}) df_{pr} = P(X1 > X0) \tag{2}$$

8 Result Analysis

The analysis aimed to discern the impact of various model architectures, optimizers, and batch sizes on the performance of these models [25]. The following section provides an in-depth analysis of the results obtained from this study.

8.1 Impact of Model Architectures

The study explored multiple deep learning architectures, including EfficientnetB0, Sequential, ResNet VGG19 and Xception. The results indicated that ResNet model consistently outperformed in terms of classification accuracy for Alzheimer's disease detection. CNNs are well-suited for capturing spatial patterns in medical images, making them a preferred choice for this task. Moreover, the VGG19 and EfficientNetB0 models exhibited promising results.

8.2 Effect of Optimizers

Here we systematically investigated various optimization algorithms, including Stochastic Gradient Descent (SGD), Adam, RMSprop. It was observed that Adam consistently outperformed other optimizers across different model architectures. From table-1 it can show that the optimizer Adam has given the best validation accuracy and AUC for all the models for the batch size of 64.

8.3 Influence of Batch Sizes

Batch size, a critical hyperparameter, was studied comprehensively in this analysis. The results showed a trade-off between batch size and training time, where larger batch sizes led to faster convergence but sometimes hindered the generalization ability of the models. Smaller batch sizes, on the other hand, exhibited more stable and consistent learning curves but required more training epochs to achieve comparable accuracy. The optimal batch size appeared to depend on the specific dataset and architecture used.

8.4 Generalization and Overfitting

The models may suffer from the issue of overfitting, which is a common challenge in deep learning. Regularization techniques such as dropout and L2 regularization were applied to mitigate overfitting. Dropout works by randomly "dropping out" (deactivating) a fraction of neurons or units in a neural network during each training iteration. This dropout is typically applied to the hidden layers of the network and helps prevent the co-adaptation of neurons. Essentially, dropout forces the network to learn more robust and generalized features because it cannot rely too heavily on any

specific neuron. L2 regularization, also known as weight decay, is a technique that adds a penalty term to the loss function of the neural network during training. This penalty is proportional to the square of the weights of the network and is added to the loss to be minimized.

Table 1. Validation Accuracy and AUC for Different optimizers and Different batch sizes of different deep-learning models

SL NO.	MODEL NAMES	OPTIMIZERS	BATCH SIZES					
			16		32		64	
			Val_ Acc	AUC	Val_ Acc	AUC	Val_ Acc	AUC
1	EfficientNetB0	SGD	71.02%	0.75	70.01%	0.74	72.06%	0.76
		Adam	76.09%	0.81	96.08%	0.98	91.02%	0.94
		RMSProp	85.04%	0.91	77.06%	0.79	68.52%	0.71
2	Sequential	SGD	84.69%	0.88	78.25%	0.81	77.28%	0.79
		Adam	82.78%	0.85	74.74%	0.77	94.97%	0.93
		RMSProp	83.39%	0.86	73.34%	0.76	63.34%	0.66
3	ResNet	SGD	81.23%	0.84	78.66%	0.79	67.01%	0.71
		Adam	79.88%	0.82	80.51%	0.83	**97.9%**	**0.98**
		RMSProp	80.33%	0.81	78.87%	0.78	85.3%	0.89
4	VGG19	SGD	71.26%	0.73	79.55%	0.81	82.64%	0.84
		Adam	88.75%	0.91	92.35%	0.94	96.05%	0.97
		RMSProp	81.26%	0.83	83.67%	0.86	89.67%	0.92
5	Xception	SGD	76.16%	0.78	79.87%	0.81	82.23%	0.84
		Adam	88.26%	0.91	95.55%	0.97	89.64%	0.90
		RMSProp	52.56%	0.55	62.59%	0.65	68.72%	0.71

8.5 Interpretability and Explainability

While the primary focus of the study was on model performance, the authors briefly discussed the interpretability of deep learning models in Alzheimer's disease detection. With SGD optimizer, EfficientNetB0 achieves moderate validation accuracy (around 71%) and AUC (around 0.75) across all batch sizes. As shown in table-1. When trained with Adam optimizer, it shows a significant improvement in both validation accuracy (reaching 91.02% with a batch size of 64) and AUC (reaching 0.94). RMSProp optimizer also performs well, particularly with a batch size of 16, where it achieves the highest validation accuracy (85.04%) and AUC (0.91). The sequential model exhibits the highest validation accuracy and AUC when trained with an SGD optimizer, reaching approximately 84.69% accuracy and 0.88 AUC with a batch size of 16. Adam optimizer also performs well, especially with a batch size of 64, where it achieves the highest validation accuracy (94.97%) and AUC (0.93). RMSProp performs consistently but tends to underperform compared to SGD and Adam.

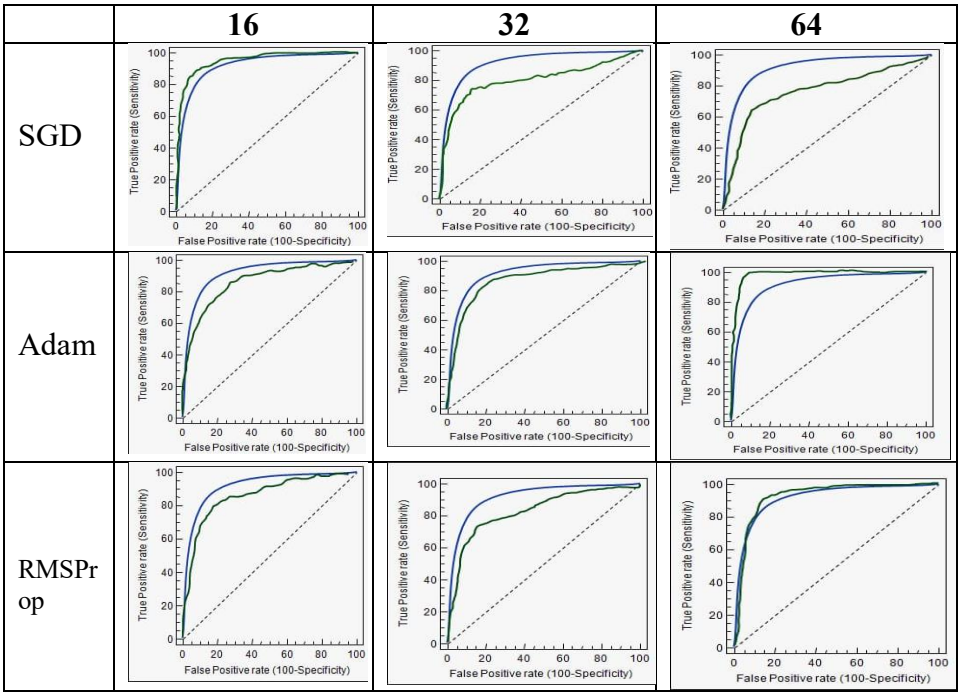
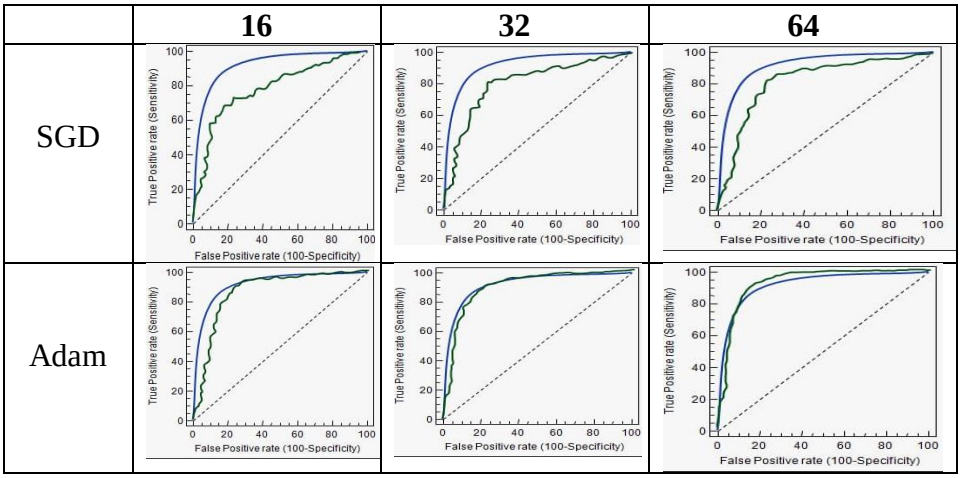


Fig.5. AUC graph with different optimizers for different batch sizes for Alzheimer's disease Detection using ResNet.

ResNet performs well with all three optimizers, with Adam yielding the highest validation accuracy (97.9%) and AUC (0.98) when using a batch size of 64. The figure 5 shows the AUC graph for different optimizer and different batch size which indicates the best accuracy can be achieved for Adam optimizer and batch size of 64. The model is relatively robust across different batch sizes.



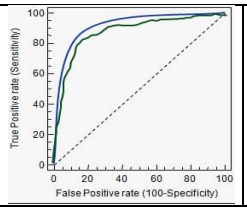
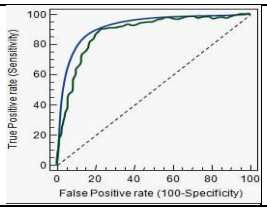
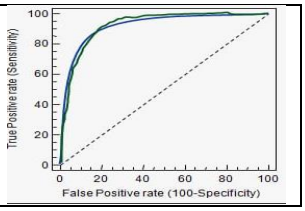
RMSProp			

Fig.6. AUC graph with different optimizers for different batch sizes for Alzheimer's disease Detection using VGG19.

VGG19 shows improved performance with larger batch sizes, especially with Adam and RMSProp optimizers. Adam optimizer achieves the highest validation accuracy (96.05%) and AUC (0.97) with a batch size of 64. figure 6 shows the AUC graph for different optimizer and different batch size which indicates the best accuracy can be achieved for Adam optimizer and batch size of 64. The model consistently outperforms with respect to validation accuracy and AUC.

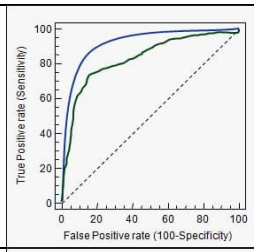
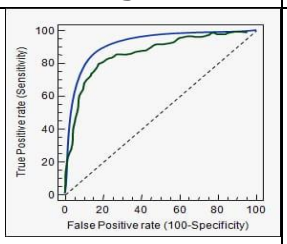
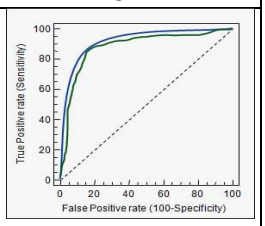
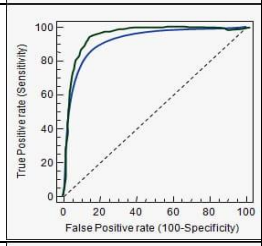
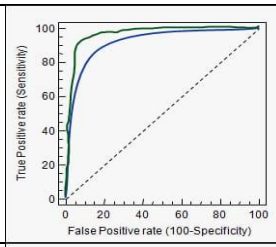
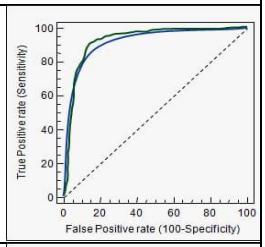
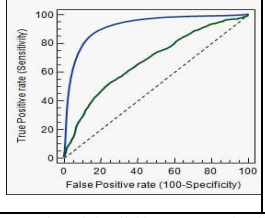
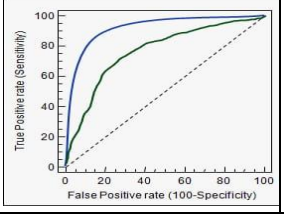
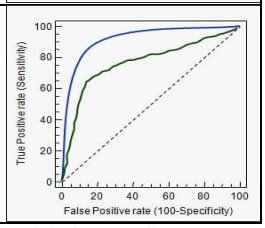
	16	**32**	**64**
SGD			
Adam			
RMSProp			

Fig.7. AUC graph with different optimizers for different batch sizes for Alzheimer's disease Detection using Xception.

Xception's performance varies significantly across optimizers and batch sizes. Adam optimizer outperforms others, achieving the highest validation accuracy (95.55%) and AUC (0.97) with a batch size of 32. RMSProp, in general, underperforms in this model. The best AUC achieved for batch size of 32 for RMSProp as shown in figure 7. Adam optimizer consistently stands out as the top performer in terms of both

validation accuracy and AUC for most models, often with batch sizes of 32 or 64. Larger batch sizes tend to lead to better performance in many cases, but there are exceptions. ResNet appears to be a robust choice across different optimizers and batch sizes, consistently achieving high accuracy and AUC. These results provide valuable insights into model selection and hyperparameter tuning for Alzheimer's disease detection using deep learning models. The manuscript could discuss these findings in the context of the challenges and opportunities in Alzheimer's disease detection and suggest potential avenues for future research and optimization strategies.

9 Conclusions

Across multiple models, the Adam optimizer consistently outperforms both SGD and RMSProp in terms of validation accuracy and AUC. It is often the top performer, achieving the highest scores for most experiments. ResNet consistently performs well across different optimizers and batch sizes, making it a robust choice for Alzheimer's disease detection. the choice of model, optimizer, and batch size can significantly affect the performance of deep-learning models in Alzheimer's disease detection. Adam optimizer tends to be a strong choice in most cases, but the specific model and dataset characteristics should also be considered when selecting the optimal combination of hyperparameters. Moreover, ResNet emerges as a reliable model architecture for this task. These findings can guide researchers and practitioners in developing more effective deep-learning models for Alzheimer's disease detection and related medical image analysis tasks. Further research may focus on fine-tuning hyperparameters and exploring additional model architectures to continue improving diagnostic accuracy.

10 References

1. Beheshti, H., Demirel, A., Initiative, D. N.: Feature-ranking-based Alzheimer’s disease classification from structural MRI, Magnetic resonance imaging, vol. 34, pp. 252-263 (2016).

2. Frisoni, G. B.,Fox, N. C., Jack Jr, C. R.,Scheltens, P., Thompson, P. M.: The clinical use of structural MRI in Alzheimer disease, Nature Reviews Neurology, vol. 6, p. 67 (2010).

3. Yang, J.,Zhao, Y. Q., Chan, J. C.-W.: Learning and transferring deep joint spectral–spatial features for hyperspectral classification, IEEE Transactions on Geoscience and Remote Sensing, vol. 55, pp. 4729-4742 (2017).

4. Kloppel, S., Stonnington, C., Chu, C., Draganski, B., Scahill, R., Rohrer, J. N., Ashburner, G., Frackowiak, R.: Automatic classification of MR scans in Alzheimer’s disease, Brain, vol. 131, no. 3, pp. 681–689 (2008).

5. Lao, Z., Shen, D., Xue, Z., Karacali, B., Resnick, S. M., Davatzikos, C.: Morphological classification of brains via high-dimensional shape transformations and machine learning methods, Neuroimage, vol. 21, pp. 46-57 (2004).

6. Weiner, M. W.: The Alzheimer’s disease neuroimaging initiative: A review of papers published since its inception, Alzheimer’s Dementia, vol. 9,no. 5, pp. e111–e194 (2013).

7. Schroeder, M. L., Stein, T., Maslowski, N., Neumann, J.: Neural correlates of

Alzheimer's disease and mild cognitive impairment: A systematic and quantitative meta-analysis involving 1351 patients, Neuroimage, vol. 10, pp. 102-108 (2002).

8. Yang, C.H., Yu, C.J.: Alzheimer's disease classification algorithm based on fusion principal component analysis and linear discriminant analysis, J Xiamen Univ NatS ci, vol. 56(2), pp. 226–230 (2017).

9. Falahati, F., Westman, E., Simmons, A.: Multivariate data analysis and machine l learning in Alzheimer's disease with a focus on structural magnetic resonance imaging, J A Disease vol. 41, no.3, pp. 685–708 (2014).

10. Adrien, P., Giovanni, M.: Predicting Alzheimer's disease: a neuroimaging study with 3D convolutional neural networks, arXiv:1502.02506 [cs.CV] (2015).

11. Moradi, E., Pepe, A., Gaser, C., Huttunen, H., Tohka, J.: Machine learning framework for early MRIbased Alzheimer's conversion prediction in MCI subjects, Neuroimage, vol. 104, pp. 398–412 (2015).

12. Cheng, B., Liu, M., Shen, D., Li, Z., Zhang, D., Initiative, A. D. N.: Multi-Domain Transfer Learning for Early Diagnosis of Alzheimer's Disease, Neuroinformatics, vol. 15, no. 2, pp. 115 (2017).

13. Jude Hemanth, D., Vijila, C. K. S., Selvakumar, A. I., Anitha, J.: Performance Improved IterationFree Artificial Neural Networks for Abnormal Magnetic Resonance Brain Image Classification, Neurocomputing, vol. 130, pp. 98–107 (2018).

14. Korolev, S.,Safiullin, A.,Belyaev, M., Dodonova, Y.: Residual and plain convolutional neural networks for 3D brain MRI classification, Proc. - Int. Symp. Biomed. Imaging, pp. 835–838 (2015).

15. Lecun, Y., Bottou, L., Bengio, Y., Haffner, P.: Gradient-based learning applied to document recognition, Proc. IEEE, vol. 86, no. 11, pp. 2278–2324 (1998).

16. Chollet, F.: Keras, (2015). [Online]. Available: https://keras.io/.

17. Fargo, k.: Alzheimer's Association Report: 2014 Alzheimers disease facts and figures, Alzheimer's Dement., vol. 10, no. 2, pp. e47–e92 (2014).

18. Shahbaz, M., Ali, S., Guergachi, A., Niazi, A., Umer, A.: Classification of alzheimer's disease using machine learning techniques, DATA, pp. 296–303 (2019).

19. Liu, M., Cheng, D., Wang, K., Wang, Y., Initiative, A. S. D. N.: Multi-modality cascaded convolutional neural networks for Alzheimer's disease diagnosis, Neuroinformatics, vol. 16, pp. 295 308 (2018).

20. Jain, R., Jain, N., Aggarwal, A., Hemanth, D. J.: Convolutional neural network based Alzheimer's disease classification from magnetic resonance brain images, Cognitive Systems Research, vol. 57, pp. 147-159 (2019).

21. Alzheimer's Association: Alzheimer's disease facts and figures: Includes a special report on the financial and personal benefits of early diagnosis, (2018).

22. Simon, F., Deo, M., Selvam, V., Babu, R.: An overview of machine learning and its

applications, International Journal of Electrical Sciences & Engineering, vol. 1, pp.22–24 (2016).

23. Reitz, C., Brayne, C., Mayeux, R.: Epidemiology of Alzheimer's disease, Nature Reviews Neurology, vol. 7, No. 3, pp.137–152 (2016).

24. Ghazi, M.M., Nielsen, M., Pai, A., Cardoso, M.J., Modat, M., Ourselin, S., Sørensen L.: Training recurrent neural networks robust to incomplete data: Application to Alzheimer's's disease progression modeling, Med. Image Anal.,Vol. 53, pp.39–46 (2019).

25. Rahman, F., Salam, M., Islam et al.: Alzheimer's disease-an update, Bangladesh Journal of Neuroscience, vol. 28, No. 1, pp.52–58 (2013).