



Classifying Alzheimer's Disease Using Hybrid Model: Xception and Machine Learning

Cheryl Angelica^{1*}, Derwin Suhartono²

¹ Computer Science Department, BINUS Graduate Program-Master of Computer Science, Bina Nusantara University, Jakarta 11480, Indonesia

² Computer Science Department, School of Computer Science, Bina Nusantara University, Jakarta 11480, Indonesia

Corresponding Author Email: cheryl.angelica@binus.ac.id

Copyright: ©2024 The authors. This article is published by IETA and is licensed under the CC BY 4.0 license (<http://creativecommons.org/licenses/by/4.0/>).

<https://doi.org/10.18280/ria.380223>

ABSTRACT

Received: 25 October 2023
Revised: 1 January 2024
Accepted: 2 February 2024
Available online: 24 April 2024

Keywords:

Alzheimer's disease, hybrid model, depthwise separable convolution, extreme gradient boost, support vector machine, random forest, xception, health care

This research introduces a novel approach to Alzheimer's disease detection by combining Xception's efficiency with machine learning classifiers, notably XGBoost. The hybrid model strategically uses Xception for feature extraction and integrates machine learning algorithms to enhance early detection accuracy, leveraging depthwise separable convolution for reduced computational complexity. Addressing imbalanced data, the study incorporates SMOTE, showcasing the hybrid model's effectiveness. Before SMOTE, the model achieved 72.89% accuracy and a 74.35% F1 score, outperforming the non-hybrid Xception model. Post-SMOTE, accuracy increases to 86.75%, and the F1 score to 86.84%, demonstrating substantial improvement without excessive computational demands. In comparison, the non-hybrid Xception model exhibits 78.71% accuracy and a 78.27% F1 score after SMOTE, emphasizing the pronounced enhancement achieved by the hybrid model. The Kaggle-derived dataset, totaling 6400 images, undergoes meticulous preprocessing, acknowledging dataset-specific constraints on generalizability. Emphasizing the importance of addressing data imbalance for robust classification, the hybrid model offers a promising solution for accurate and efficient Alzheimer's disease detection. This study contributes valuable insights to the field, showcasing the potential of innovative hybrid models to address complex healthcare challenges while balancing accuracy and computational efficiency.

1. INTRODUCTION

In the realm of human memory, three fundamental components play pivotal roles: working memory, short-term memory, and long-term memory, each serving distinct functions in the cognitive processes. Working memory facilitates attention and concentration during the intake of data and information, while short-term memory temporarily stores information for immediate use. In contrast, long-term memory serves as the repository for a lifetime of experiences [1]. However, this intricate system of memory is susceptible to various maladies, with Alzheimer's disease representing a significant and devastating affliction. As a progressive neurodegenerative disorder, Alzheimer's erodes memory, cognitive abilities, and even basic daily functioning. It constitutes the predominant form of dementia, contributing to a substantial percentage of dementia cases globally, and its prevalence is poised to rise exponentially. The World Alzheimer's Report of 2015 underscores a grave global concern, with over 50 million individuals grappling with dementia across the world, a number that is poised to double every two decades, as visually represented in Figure 1. This alarming trend is not isolated to the global stage; Indonesia, for instance, presented recent statistics in 2022 revealing a staggering 1.2 million of its citizens contending with the challenges of Alzheimer's disease [2]. Regrettably, a definitive

cure remains elusive, as the disease continues to ravage brain cells [3]. Nonetheless, early detection holds promise in enabling medical professionals, particularly doctors, to explore interventions that may temporarily ameliorate symptoms, slow disease progression, and mitigate neural damage.

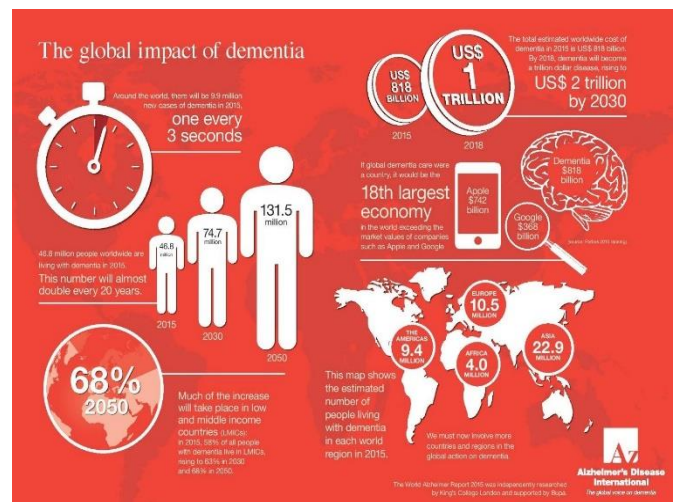


Figure 1. Statistics of dementia over the world

Deep learning (DL) holds the transformative potential to revolutionize medical diagnostics, and within this domain, convolutional neural networks (CNNs), a subset of deep learning algorithms, have exhibited promise in the direct diagnosis of Alzheimer's disease using medical imaging data [4]. The LeNet-5 design was used in an earlier study by Sarraf and Tofghi [5], which also highlighted the necessity for more convolutional neural layers to enhance the accuracy of Alzheimer's disease diagnosis using MRI scans. However, traditional CNN models face limitations, including significant computational demands and a high number of training parameters, necessitating expensive computing resources.

To address these challenges, this work proposes a hybrid model that combines Xception, a cutting-edge deep learning architecture introduced by Chollet in 2017, with machine learning strategies for Alzheimer's disease classification [6]. Notably, prior studies have shown the efficacy of various models but often encounter difficulties related to imbalanced data and expensive computing resources. The suggested hybrid approach aims to leverage the strengths of both Xception and machine learning, providing a more robust and efficient solution for Alzheimer's classification. This research significantly contributes by explicitly addressing the drawbacks of previous methodologies, potentially advancing Alzheimer's disease detection and improving patient care.

Chollet's introduction of Xception in 2017 is noteworthy, combining residual connections and Depthwise Separable Convolution (DSC) for increased accuracy and reduced computing complexity [6]. Compared to conventional convolutional layers, DSC employs fewer parameters and computational calculations while maintaining an equivalent level of performance [7]. The central research question guiding this study is: How can a hybrid model, combining Xception and machine learning techniques, improve the efficiency of Alzheimer's disease detection compared to traditional CNN models?

To address this question, the research objectives include evaluating the effectiveness of the proposed hybrid models compared to non-hybrid models, specifically Xception alone. The hybrid models, incorporating Xception and machine learning techniques, will be systematically compared with the non-hybrid model to comprehensively evaluate their effectiveness. By combining the advantages of both methods, the proposed model aims to overcome the limitations of traditional CNN models, offering improved F1-Scores and reduced computing complexity.

Additionally, the challenge of imbalanced data is addressed through the application of the Synthetic Minority Over-sampling Technique (SMOTE), notably enhancing the performance of the selected models. This comprehensive approach ensures valuable insights into the most suitable method for Alzheimer's detection, furthering the understanding and advancement of diagnostic methodologies in the field.

2. RELATED WORKS

In one notable study conducted by Shahbaz et al. [8], a comprehensive investigation was undertaken to compare various ML models, including K-Nearest Neighbor (KNN), Decision Tree (DT), Rule Induction, Naïve Bayes, Generalized Linear Model (GLM), and deep learning models. They utilized the TADPOLE Alzheimer's Disease

Neuroimaging Initiative (ADNI) dataset for their analysis. Remarkably, the GLM model emerged as the top performer, achieving an impressive accuracy rate of 88.24%. Similarly, Murugan et al. [9] conducted research that focused on the modification of a CNN architecture called DEMentia NETwork (DEMNET) for Alzheimer's disease detection. They leveraged a publicly available dataset from Kaggle and compared various data pre-processing techniques, including the use of Synthetic Minority Over-sampling Technique (SMOTE) to address dataset imbalance. Their results revealed that SMOTE had a substantial impact, significantly enhancing accuracy by a substantial 10% margin.

Building on this, Prakash et al. [10] employed the ADNI dataset and introduced pre-trained CNN models and transfer learning strategies to enhance model accuracy. They conducted a comparative analysis of different CNN architectures, including ResNet-101, ResNet-50, and ResNet-18, with ResNet-101 emerging as the top performer, achieving an accuracy rate of 98.37%. Likewise, Khan et al. [11] in their research using the ADNI dataset, compared VGG-16 and VGG-19 and found that VGG-19 slightly outperformed ResNet-101, achieving an accuracy rate of 98.47%.

A different approach was taken by Chui et al. [12], who utilized the Open Access Sequence of Image Studies (OASIS) dataset. They proposed a novel method that combined a Generative Adversarial Network (GAN), CNN, and transfer learning. The GAN module was used to generate additional training data for the minority class, effectively addressing class imbalance. Their method displayed improvements in detection model accuracy across various evaluation scenarios, with improvements ranging from 2.85% to 40.1%. Meanwhile, Ganesh et al. [13] employed pre-trained CNN models such as VGG-16, InceptionV3, and Xception. Among these, VGG-16 achieved the highest accuracy, reaching 75%.

In a recent study Zena et al. [14] focusing on Alzheimer's disease diagnosis and classification, the authors emphasized the critical importance of early diagnosis for effective treatment and management of the disease. They explored the use of deep learning methods, including popular architectures like MobileNetV2, ResNet-101, DenseNet-121, and a modified convolutional neural network (CNN) model inspired by VGG16, for classifying normal brain scans and various stages of Alzheimer's disease using magnetic resonance imaging (MRI) data obtained from Kaggle. Their investigation centered on evaluating the classification performance of these deep learning architectures, with a specific focus on accuracy, precision, recall, and F1-score as performance metrics. Remarkably, the results of their study demonstrated the superiority of the proposed modified CNN model, which achieved an outstanding accuracy of 97.625%, along with a recall, precision, and F1-score of 98%. This suggests that their approach outperforms the other deep learning models considered.

Beyond these studies, another promising avenue for Alzheimer's disease detection involves the development of hybrid models that integrate CNNs as feature extractors with ML models as classifiers. For instance, Tuan et al. [15] combined CNN with XGBoost and SVM, achieving an accuracy of 89% with their CNN+XGBoost hybrid model. Eroglu et al. [16] utilized Darknet 53, InceptionV3, and ResNet101 as pre-trained CNN models combined with SVM and KNN, reaching an impressive accuracy of 96.1% with the ResNet101+KNN and Darknet 53 +KNN hybrid models. Sharma et al. [17] employed DenseNet-121 and DenseNet-201

in combination with SVM, Gaussian Naïve Bayes, and XGBoost, achieving a top accuracy of 91.75% with the DenseNet201+Gaussian Naïve Bayes hybrid model.

In addition, the work [7] stands out for its focus on creating a more compact and accurate model for Alzheimer's disease detection. They conducted a comparative analysis of convolutional CNN, Deep Separable Convolution (DSC), and transfer learning algorithms. DSC exhibited a significant reduction in complexity and computational cost while maintaining high accuracy. Transfer learning models, such as AlexNet and GoogLeNet, achieved accuracy rates of 91.40% and 93.02%, respectively, albeit at a higher computational cost. The proposed DSC module offers a potential solution to reduce model complexity, offering a balance between accuracy and efficiency.

Given this extensive literature review, the current research introduces a hybrid model, combining Xception and machine learning techniques. Unlike existing single-modality approaches, such as ResNet-101, VGG-16, and InceptionV3, our hybrid approach synergistically utilizes the accuracy and computational efficiency of Xception for feature extraction, coupled with machine learning strategies as classifiers including Gaussian Naïve Bayes, XGBoost, Random Forest, and SVM, for early detection of Alzheimer's disease. These selected models have demonstrated promising accuracy in prior research and the hybrid model is expected to strike a balance between compactness and accuracy, presenting a robust solution for effective Alzheimer's disease detection.

This integration overcomes challenges seen in traditional models, addressing issues like imbalanced data and reducing computational demands. Our model stands out by uniquely

employing DSC in Xception, significantly reducing computing complexity. This innovation offers a balanced solution, enhancing F1 Score while reducing computational demands, making it suitable for resource-constrained environments.

In comparison to traditional CNN + XGBoost/SVM models [15], our hybrid approach showcases efficiency by utilizing DSC in the CNN architecture, requiring fewer parameters and computations while maintaining performance levels. This distinctive combination presents an innovative solution, advancing Alzheimer's disease detection by overcoming limitations faced by existing models.

3. METHODOLOGY

The methodology follows a well-structured framework comprising several crucial stages which are shown in Figure 2. It initiates with data preparation, involving the acquisition of relevant datasets and meticulous pre-processing to enhance data quality. The pre-processing phase incorporates techniques such as image augmentation, oversampling, and data splitting. At its core, the approach centers on constructing a hybrid model that utilizes Xception for feature extraction, allowing the identification and extraction of key patterns from the pre-processed data to enhance model performance. Subsequently, a machine learning model serves as the classifier. The final step entails model evaluation, which rigorously assesses the performance of the models and the overall success of the approach.

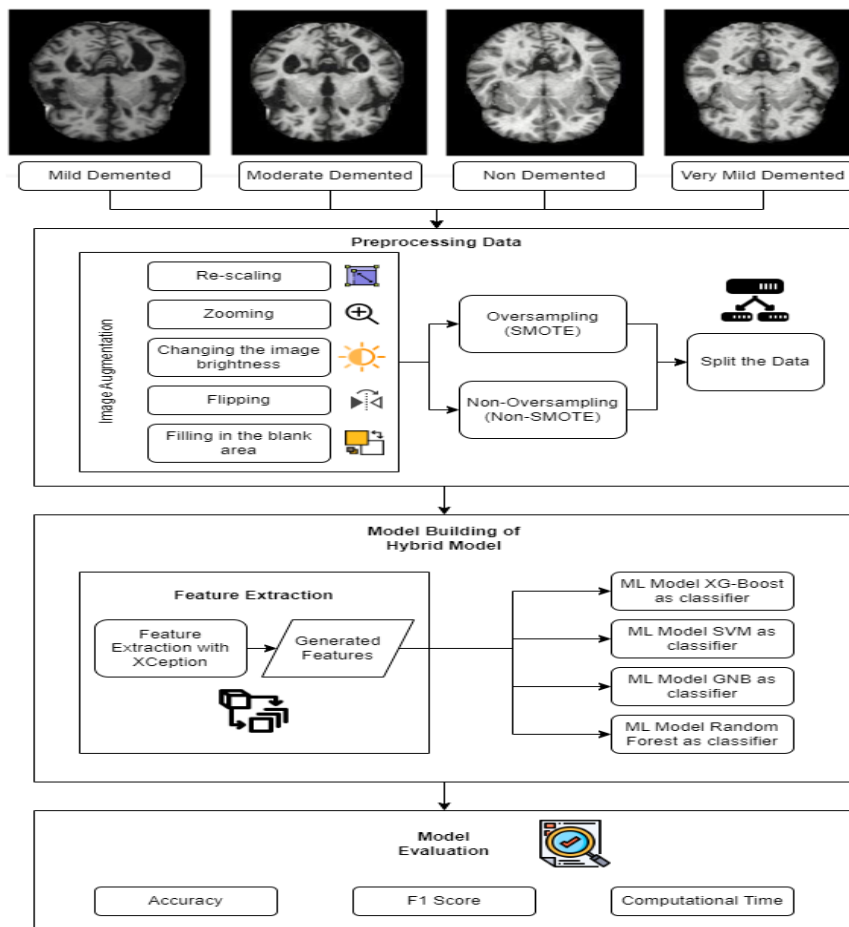


Figure 2. Proposed methodology

3.1 Data preparation

In this research, the dataset was obtained from various websites with each and every label verified. This dataset is available on Kaggle [18] and comprises 6,400 images depicting four phases of Alzheimer's disease: mild dementia (MD), moderate dementia (Mod. D), non-demented (ND), and very mild dementia (VMD). Importantly, it's worth noting that each of these phases of Alzheimer's disease in the dataset is based on different individuals. The distribution of MRI pictures for each class, as well as the number of images in the training and testing sets, are shown in Table 1. The MD class is composed of 896 images, while the Mod. D class consists of 64 images. In total, the ND class comprises 3,200 photos. Lastly, the VMD class encompasses 2,249 images. The examples of the images have been shown in Figure 3.

Table 1. Number of Alzheimer's MRI Image Datasets

Class	Total
MildDemented (MD)	896
ModerateDemented (Mod. D)	64
NonDemented (ND)	3,200
VeryMildDemented (VMD)	2,249
Total Set	6,400

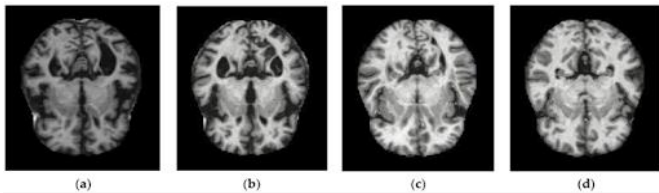


Figure 3. Alzheimer MRI dataset: (a) MD, (b) Mod.D, (c) ND, and (d) VMD

3.2 Preprocessing data

Preprocessing data is the first step in building a model that involves collecting raw data and converting it into a format that can be processed by the model. In this research, data preprocessing involves several stages, namely image augmentation, oversampling with SMOTE, and splitting the dataset.

3.2.1 Image augmentation

The process begins with defining image classes and sizes to facilitate image augmentation for improved image classification. The image dimensions were set to 299x299 pixels with RGB color channels, in alignment with the Xception model's requirements. Image augmentation is then employed to diversify the dataset by applying transformations such as horizontal flipping, zooming, filling, and brightness adjustments to create new data samples. Subsequently, the dataset is loaded using the 'flow_from_directory' method from the 'ImageDataGenerator' class, structured to match the custom directory layout. This data generator streamlines data loading and seamlessly integrates predefined image augmentations, simplifying the model training process.

3.2.2 SMOTE

The Synthetic Minority Oversampling Technique (SMOTE) is a method for resolving imbalanced data that oversamples the minority class using "synthetic" instances. However, it is

essential to acknowledge the potential risk of overfitting associated with the generation of synthetic data. Instead than concentrating on all data points, SMOTE is carried out based on the value and features of the data relationships by employing synthetic examples in "feature space" as opposed to "data space." SMOTE works by oversampling each minority class and injecting synthetic cases along the axes that connect any or all of the k-nearest neighbors of each minority class. In accordance with the required level of oversampling, neighbors from the k-nearest neighbors are chosen at random [19]. To mitigate the risk of overfitting, careful consideration and validation of the synthetic instances' impact on the model's generalization should be undertaken, possibly through cross-validation or other relevant techniques. This ensures that the benefits of addressing class imbalance with SMOTE are achieved without compromising the model's ability to generalize to new, unseen data.

3.2.3 Splitting the dataset

For the hybrid model, the dataset is initially divided into two subsets: an 80% training set and a 20% testing set. In the case of the experiment in a non-hybrid model, the training set is further subdivided into two distinct subsets: an 80% training set and a 20% validation set. This additional step ensures the non-hybrid model's performance can be effectively evaluated during the training process.

3.3 Model building

In the model building stage, a novel approach is employed by constructing a hybrid model that merges the capabilities of the Xception architecture as a feature extractor with the proficiency of a machine learning model serving as the classifier. This integration leverages the strengths of both components to create a comprehensive and powerful model.

3.3.1 Feature extraction using xception

In the context of feature extraction using Xception, a pre-trained Xception model is employed in its original state as the foundation for this process. The decision to use the pre-trained Xception model without fine-tuning on the dataset is rooted in the model's proficiency in image recognition tasks and its ability to generalize well to diverse datasets. Xception, often referred to as "extreme inception," stands out for its convolutional neural network architecture designed to significantly enhance the efficiency and effectiveness of image recognition tasks, distinguishing it from the Inception architecture.

The key to Xception's success lies in its utilization of depthwise separable convolution, a two-step convolutional approach. It initiates with depthwise convolution, where each filter independently processes a single channel of the input image, effectively reducing computational load while preserving crucial spatial information. Following this, pointwise convolution takes over, employing 1x1 filters to amalgamate and transform information from individual channels. What truly distinguishes Xception is the sequence in which it applies these convolutions; it commences with 1x1 pointwise convolution and subsequently proceeds with channel-wise spatial convolution which shown in Figure 4. This unique order optimizes feature extraction and concurrently minimizes computational overhead, rendering Xception an exceptionally efficient and effective architecture for image recognition tasks [20].

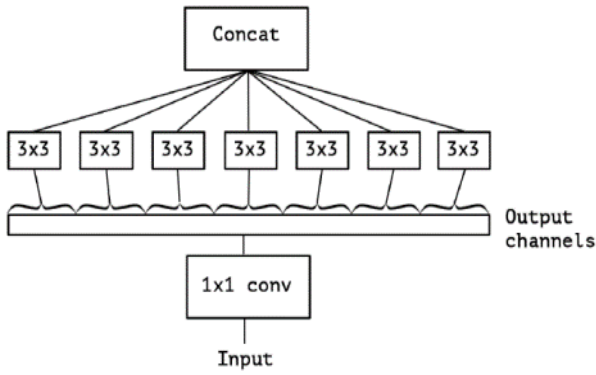


Figure 4. Xception module

This Xception model is then configured to exclude its top classification layers and is tailored to accept images with dimensions of 299 by 299 pixels and three color channels (RGB). A Global Average Pooling 2D layer is added to the Xception model, reducing the output dimension. This modified model now serves as a feature extractor, taking input data and producing feature representations for each input image. The feature extraction is applied to both the training and test datasets. The resulting feature arrays are then reshaped into a format suitable for further processing. Additionally, the original labels are transformed to a one-hot encoded format, where each label is represented as a binary vector. This meticulous feature extraction process is crucial in preparing the data for subsequent machine learning or classification tasks, ensuring optimal utilization of Xception's capabilities without fine-tuning for the specific dataset.

3.3.2 Classification using machine learning model

The extracted features from Xception serve as the foundation for Alzheimer's stage classification, with a range of machine learning models selected for their distinct strengths in addressing this task. Gaussian Naive Bayes (GNB) is embraced for its simplicity and efficiency. It operates under the assumption of independence among predictors, making it well-suited for datasets with a multitude of features. GNB's foundation in probabilistic principles, using Bayes' theorem, enables it to accurately model the conditional probabilities of features given a class label [21].

Support Vector Machine (SVM), another vital choice, excels in high-dimensional feature spaces. SVM's robustness is attributed to its capacity to identify an optimal hyperplane, thereby maximizing the margin between different classes. It thrives in scenarios with complex decision boundaries, offering the versatility to address both linear and non-linear data separations. This adaptability proves advantageous for Alzheimer's stage classification [22].

XGBoost (Extreme Gradient Boosting) is a prominent ensemble method renowned for its high predictive accuracy and scalability. It builds upon gradient boosting and is particularly adept at capturing intricate relationships within data. By combining weak learners into a robust ensemble model, XGBoost is well-suited for tasks like Alzheimer's stage classification, where detecting subtle data patterns is crucial [23].

Random Forest (RF) is a versatile ensemble learning approach. It assembles multiple decision trees, each constructed from random training data subsets. The strength of RF lies in its ability to aggregate predictions from these trees

through averaging, which mitigates overfitting and enhances model accuracy. RF excels at managing high-dimensional data and intricate feature relationships, making it a valuable asset in Alzheimer's stage classification. Additionally, RF is known for its capacity to provide robust predictions, even in the presence of outliers and noise [24].

The selection of these models is deliberate, aiming to harness their individual strengths, ranging from simplicity and efficiency to predictive accuracy and robustness in handling complex data distributions. Notably, in this experiment, parameter tuning was not performed for these models. The best model was determined by comparing the F1 Score and computation time, with a focus on achieving an optimal trade-off between classification performance and computational efficiency. This strategic combination ensures the achievement of accurate and versatile Alzheimer's stage classification, contributing to the ongoing progress in the field.

3.4 Evaluation

In the model evaluation stage, this study employs a range of metrics to assess performance, encompassing Accuracy and F1-Score. Additionally, this study utilizes Computation Time as a critical metric to gauge the model's efficiency.

3.4.1 Accuracy

Accuracy measures the proportion of correctly classified instances out of the total number of examples in the dataset. Accuracy provides a convenient way to assess how well the model predicts the correct class label. In other words, accuracy is the ratio of the number of correct positives and negatives (correctly classified instances) to the total number of instances in the dataset. To find the accuracy value, the formula used can be seen in Eq. (1).

$$Accuracy = \frac{\text{Number of correct predictions}}{\text{Total number of predictions}} \quad (1)$$

3.4.2 F1 score

A classifier's performance is assessed using the F1 score, which combines precision and recall. Precision measures the accuracy of positive predictions made by the classifier, indicating how many of the predicted positive instances were actually correct. On the other hand, recall, also known as sensitivity, measures the classifier's ability to correctly identify all relevant instances in the dataset, highlighting the proportion of true positive instances that were successfully detected.

The F1 score merges these two essential measures into a single statistic [25], providing a balanced assessment of a classifier's performance. It is frequently used to evaluate how well various classifiers perform. To calculate the F1 score, the formula used can be seen in Eq. (2). This single metric encapsulates both precision and recall, offering a comprehensive view of the classifier's ability to make accurate positive predictions while capturing all relevant instances.

$$F1 \text{ score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (2)$$

3.4.3 Computation time

Computation time serves as a pivotal metric in the evaluation process, allowing for a comprehensive comparison between the hybrid model and the non-hybrid model. The

objective of this analysis is to underscore the notable advantage of employing DSC-based CNN models, such as Xception, within the hybrid model. By meticulously measuring and comparing the computational time required for both models, it is aimed to demonstrate that the DSC implementation in the hybrid model significantly reduces the overall computational load.

In this experiment, a V100 GPU on Google Colab Pro for computation was utilized. This detail is essential for reproducibility and ensuring a fair comparison. The efficiency demonstrated by the reduced computational calculations is pivotal not only in terms of model training and inference speed but also in resource utilization. By highlighting this reduction in computational load and providing information about the hardware environment, the study sheds light on the potential benefits and optimizations that can be harnessed through the integration of Xception within the hybrid model, ultimately enhancing the model's overall performance and practical utility.

4. RESULT AND DISCUSSIONS

The preprocessing results in image resizing to 299x299 pixels and dataset splitting into training and testing sets. The image augmentation techniques, including horizontal flipping, zooming, filling, and brightness adjustments, diversify the dataset. SMOTE resolves imbalanced data by oversampling the minority class with synthetic instances. The data is then divided into subsets: 80% training and 20% testing for the hybrid model and, in the non-hybrid model, further split into 80% training and 20% validation subsets. This meticulous data preparation underpins the subsequent model building and evaluation steps, ensuring robust and meaningful results.

This research provides compelling results that substantiate the efficacy of the proposed hybrid model in accurately classifying the stages of Alzheimer's disease, as demonstrated in Table 2. Before the application of SMOTE, a thorough examination of the model's performance reveals the efficacy of the hybrid model, Xception+XGBoost, in classifying Alzheimer's stages. This hybrid model achieved a noteworthy accuracy of 72.89%, and a commendable F1 score of 74.35%, all while operating efficiently within a brief computational time of 218 seconds. This striking performance exemplifies the hybrid model's capacity to effectively classify the four stages of Alzheimer. Importantly, in direct comparison to the non-hybrid Xception model, which achieved an accuracy of 66.17% and F1 score of 65%, the hybrid model exhibits

superior classification performance. The comparison with non-hybrid deep learning models further underscores the incremental benefit of the hybrid approach over deep learning alone. Additionally, it does so with significantly reduced computational demands, as the non-hybrid Xception model necessitates a substantially longer computational time of 3,556 seconds, underscoring the compelling advantage of the hybrid approach in terms of both accuracy and efficiency. The integration of traditional machine learning models, like XGBoost, with deep learning architectures, such as Xception, showcases a synergistic effect, improving overall classification performance.

Table 2 presents a further comparison between different hybrid models based on accuracy, F1 Score, and computational time. The Xception+XGBoost hybrid model outperforms other hybrid models in terms of accuracy and F1 score, indicating its superior classification capability. Followed by the Xception+SVM hybrid model, which achieved an accuracy of 71.64% and an F1 score of 71.05%, and notably, it demonstrates a more efficient computational time of 27 seconds compared to the Xception+XGBoost hybrid model. The results from this comparison emphasize the effectiveness of hybrid models in addressing the complexities of Alzheimer's disease classification. The synergy between deep learning, represented by the Xception architecture, and traditional machine learning, represented by XGBoost and SVM, results in models that excel both in accuracy and computational efficiency. These findings highlight the potential for enhancing diagnostic processes and clinical decision-making in Alzheimer's disease detection.

The implementation of SMOTE significantly enhances the robustness of the models, improving accuracy and F1 scores across various machine learning models, as demonstrated in Table 3. All models' accuracy and F1 score has improved. Notably, the same hybrid model which achieved the highest accuracy and F1 score before SMOTE got an accuracy of 86.76%, and F1 score of 86.85% after SMOTE, all while maintaining a reasonable computational time of 439 seconds. This underscores the efficacy of SMOTE in enhancing the hybrid model's classification capabilities without imposing excessive computational overhead, reaffirming its role in addressing data imbalance and bolstering model robustness. However, it's crucial to note that SMOTE introduces potential drawbacks, such as the generation of synthetic instances that may introduce artificial noise, impacting model generalization. Future work should carefully consider this trade-off between addressing data imbalance and maintaining model generalizability.

Table 2. Experimental results without SMOTE

Model	Accuracy	Precision	Recall	F1 Score	Computational Time (Seconds)
Xception+XGBoost	72.89%	77.92%	72.89%	74.35%	218s
Xception+Gaussian Naïve Bayes	45.78%	64.19%	45.78%	48.60%	76s
Xception+SVM	71.64%	71.40%	71.64%	71.05%	191s
Xception+Random Forest	65.15%	76.32%	65.15%	69.17%	91s
Xception	66.17%	70.23%	60.62%	65.06%	3556s

Table 3. Experimental results with SMOTE

Model	Accuracy	Precision	Recall	F1 Score	Computational Time (Seconds)
Xception+XGBoost	86.75%	86.99%	86.75%	86.84%	439s
Xception+Gaussian Naïve Bayes	48.78%	67.81%	48.78%	53.08%	141s
Xception+SVM	83.82%	83.79%	83.82%	83.80%	413s
Xception+Random Forest	82.50%	83.36%	82.50%	82.75%	174s
Xception	78.71%	82.53%	74.57%	78.27%	7,127s

The diverse performance across models may be attributed to the inherent complexity of Alzheimer's disease classification. Different models exhibit varying degrees of sensitivity to nuanced patterns within the data, leading to differences in accuracy and F1 scores. For instance, Gaussian Naive Bayes (GNB) operates under the assumption of feature independence, potentially limiting its effectiveness in capturing intricate relationships present in Alzheimer's disease data. Similarly, the non-hybrid Xception model, while a powerful deep learning architecture, may struggle with certain aspects of feature representation crucial for Alzheimer's classification. The hybrid models, on the other hand, showcase the benefits of combining the strengths of Xception's feature extraction capabilities with the discriminative power of machine learning classifiers. These nuanced insights into model performance shed light on the challenges inherent in Alzheimer's disease classification and underscore the importance of a hybrid approach that leverages the strengths of different models for enhanced accuracy and efficiency. Further exploration of model interpretability and feature importance analysis could provide deeper insights into the specific characteristics of Alzheimer's disease data that contribute to varied model performance.

The research findings hold significant clinical implications, with the enhanced accuracy and computational efficiency of the hybrid model offering the potential for more precise and timely diagnoses of Alzheimer's disease. This advancement not only promises improved patient outcomes but also carries the prospect of reducing the strain on healthcare systems, particularly in the context of an aging population and rising Alzheimer's disease cases.

Looking ahead, there are promising avenues for future research. Exploring various deep learning architectures and the integration of alternative machine learning algorithms can further refine and extend the findings of this study. Additionally, expanding the dataset and conducting clinical validation and deployment in real-world healthcare settings will be crucial to solidify the practical applicability of the proposed approach.

Nonetheless, it is essential to acknowledge the study's limitations. These encompass the use of a specific dataset from Kaggle [18] which may introduce constraints on the generalizability of the results. The Kaggle dataset, while comprehensive, may not fully encapsulate the diverse range of imaging characteristics and patient demographics encountered in real-world clinical settings. Variations in imaging protocols, equipment, and patient populations across different healthcare institutions can significantly impact the performance of machine learning models. Therefore, the reliance on a single dataset, albeit rich in Alzheimer's disease images, could potentially limit the external validity of the proposed hybrid model.

In addition to dataset-specific limitations, there are inherent concerns related to overfitting, particularly given the use of SMOTE. The introduction of synthetic instances through SMOTE aims to address imbalanced data but raises the possibility of overfitting to the augmented data. To mitigate this concern, the study adopts a cautious approach by refraining from parameter tuning.

Furthermore, the ethical deployment of machine learning models for Alzheimer's disease diagnosis demands a comprehensive approach to safeguard patient privacy, ensure data security, and enhance model transparency. Adherence to privacy regulations, robust data anonymization, and

encryption protocols are essential to protect sensitive medical data. Transparent communication and obtaining informed consent from patients regarding data usage are critical ethical practices. Ensuring fairness, monitoring and addressing biases, and conducting rigorous clinical validation contribute to responsible model deployment. Additionally, efforts to enhance model interpretability and decision-making transparency build trust among healthcare professionals and patients. Ethical considerations underscore the need for a balanced integration of technological advancements with patient-centric principles, emphasizing the responsible and transparent use of machine learning in healthcare settings.

5. CONCLUSION

In conclusion, this research significantly contributes to the field of Alzheimer's disease detection. The presented hybrid model, which combines DSC-based CNN models like Xception with machine learning classifiers, demonstrates its effectiveness in classifying Alzheimer's stages with computational efficiency. Particularly, before SMOTE, the hybrid model, Xception + XGBoost, achieves higher accuracy than the non-hybrid model, all within a computational time of 218 seconds. The incorporation of SMOTE further enhances the model's robustness, improving accuracy and F1 scores by approximately 14%, with a computational time of 439 seconds. Post-SMOTE application, the hybrid model sustains its classification prowess while significantly reducing computational demands. This study underscores the hybrid model's unique ability to harmonize accuracy, compactness, and computational efficiency, presenting a promising avenue for advancing Alzheimer's disease detection. Furthermore, it accentuates the critical role of addressing data imbalance for robust and reliable classification.

Looking ahead, the practical implications of these findings extend to the potential integration of the hybrid model into clinical workflows, facilitating early Alzheimer's diagnosis and intervention. While the model showcases substantial promise, its seamless adoption into clinical practice necessitates a thorough understanding of potential barriers. Future research endeavors could delve into exploring the translational aspects, considering factors such as model interpretability, validation across diverse patient populations, and the incorporation of the model into existing diagnostic workflows.

In reflecting on the broader impact of this research, it not only provides valuable insights to the medical field but also lays the foundation for redefining current diagnostic procedures for Alzheimer's disease. The hybrid model's capacity to navigate the intricate landscape of Alzheimer's classification opens avenues for more precise and timely diagnoses. As we move forward, the research invites a collaborative effort between the realms of artificial intelligence and clinical practice, fostering a dialogue that paves the way for transformative advancements in Alzheimer's disease diagnosis and patient care.

REFERENCES

- [1] Mahale, S.S. (2022). Classification of Alzheimer's disease using convolutional neural networks. *International Research Journal of Modernization in*

- Engineering Technology and Science. <https://doi.org/10.56726/irjmets30750>
- [2] Guerchet, M., Prince, M., Prina, M. (2020). Numbers of people with dementia worldwide: An update to the estimates in the World Alzheimer Report 2015.
- [3] Breijyeh, Z., Karaman, R. (2020). Comprehensive review on Alzheimer's disease: Causes and treatment. *Molecules*, 25(24): 5789. <https://doi.org/10.3390/molecules25245789>
- [4] Aggarwal, R., Sounderajah, V., Martin, G., Ting, D.S., Karthikesalingam, A., King, D., Ashrafian, H., Darzi, A. (2021). Diagnostic accuracy of deep learning in medical imaging: A systematic review and meta-analysis. *NPJ Digital Medicine*, 4(1): 65. <https://doi.org/10.1038/s41746-021-00438-z>
- [5] Sarraf, S., Tofghi, G. (2016). Classification of Alzheimer's disease using fMRI data and deep learning convolutional neural networks. *arXiv Preprint arXiv: 1603.08631*. <https://doi.org/10.48550/arXiv.1603.08631>
- [6] Chollet, F. (2017). Xception: Deep learning with depthwise separable convolutions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Honolulu, HI, USA, pp. 1251-1258. <https://doi.org/10.1109/cvpr.2017.195>
- [7] Liu, J., Li, M., Luo, Y., Yang, S., Li, W., Bi, Y. (2021). Alzheimer's disease detection using depthwise separable convolutional neural networks. *Computer Methods and Programs in Biomedicine*, 203: 106032. <https://doi.org/10.1016/j.cmpb.2021.106032>
- [8] Shahbaz, M., Ali, S., Guergachi, A., Niazi, A., Umer, A. (2019). Classification of Alzheimer's disease using machine learning techniques. In *Data*, pp. 296-303. <https://doi.org/10.5220/0007949902960303>
- [9] Murugan, S., Venkatesan, C., Sumithra, M.G., Gao, X.Z., Elakkiya, B., Akila, M., Manoharan, S. (2021). DEMNET: A deep learning model for early diagnosis of Alzheimer diseases and dementia from MR images. *IEEE Access*, 9: 90319-90329. <https://doi.org/10.1109/access.2021.3090474>
- [10] Prakash, D., Madusanka, N., Bhattacharjee, S., Kim, C.H., Park, H.G., Choi, H.K. (2021). Diagnosing Alzheimer's disease based on multiclass MRI scans using transfer learning techniques. *Current Medical Imaging Formerly Current Medical Imaging Reviews*, 17(12): 1460-1472. <https://doi.org/10.2174/1573405617666210127161812>
- [11] Khan, R., Akbar, S., Mehmood, A., Shahid, F., Munir, K., Ilyas, N., Asif, M., Zheng, Z. (2023). A transfer learning approach for multiclass classification of Alzheimer's disease using MRI images. *Frontiers in Neuroscience*, 16. <https://doi.org/10.3389/fnins.2022.1050777>
- [12] Chui, K.T., Gupta, B.B., Alhalabi, W., Alzahrani, F.S. (2022). An MRI scans-based Alzheimer's disease detection via convolutional neural network and transfer learning. *Diagnostics*, 12(7): 1531. <https://doi.org/10.3390/diagnostics12071531>
- [13] Ganesh, C.H.R., Nithin, G.S., Akshay, S., Rao, T.V.N. (2022). Multi class Alzheimer disease detection using deep learning techniques. In *2022 International Conference on Decision Aid Sciences and Applications (DASA)*, Chiangrai, Thailand, pp. 470-474. <https://doi.org/10.1109/dasa54658.2022.9765267>
- [14] Zena, J.I., Lucky, E., Ellaine, C.G., Edbert, I.S., Suhartono, D. (2022). Deep learning approach based classification of Alzheimer's disease using brain MRI. In *2022 5th International Seminar on Research of Information Technology and Intelligent Systems (ISRITI)*, Yogyakarta, Indonesia, IEEE, pp. 397-402. <https://doi.org/10.1109/isriti56927.2022.10053030>
- [15] Tuan, T.A., Pham, T.B., Kim, J.Y., Tavares, J.M.R.S. (2020). Alzheimer's diagnosis using deep learning in segmenting and classifying 3D brain MR images. *International Journal of Neuroscience*, 132(7): 689-698. <https://doi.org/10.1080/00207454.2020.1835900>
- [16] Eroglu, Y., Yildirim, M., Cinar, A. (2021). mRMR-based hybrid convolutional neural network model for classification of Alzheimer's disease on brain magnetic resonance images. *International Journal of Imaging Systems and Technology*, 32(2): 517-527. <https://doi.org/10.1002/ima.22632>
- [17] Sharma, S., Gupta, S., Gupta, D., Altameem, A., Saudagar, A.K.J., Poonia, R.C., Nayak, S.R. (2022). HTLML: Hybrid AI based model for detection of Alzheimer's disease. *Diagnostics*, 12(8): 1833. <https://doi.org/10.3390/diagnostics12081833>
- [18] Alzheimer's Dataset (4 class of Images). Kaggle. <https://www.kaggle.com/datasets/tourist55/alzheimers-dataset-4-class-of-images>, accessed on Dec. 26, 2019.
- [19] Elreedy, D., Atiya, A.F. (2019). A comprehensive analysis of synthetic minority oversampling technique (SMOTE) for handling class imbalance. *Information Sciences*, 505: 32-64. <https://doi.org/10.1016/j.ins.2019.07.070>
- [20] Polat, Ö.Z.L.E.M. (2021). Detection of covid-19 from chest CT images using Xception architecture: A deep transfer learning based approach. *Sakarya University Journal of Science*, 25(3): 800-810. <https://doi.org/10.16984/saufenbilder.903886>
- [21] Mahesh, B. (2020). Machine learning algorithms-a review. *International Journal of Science and Research*, 9(1): 381-386. <https://doi.org/10.21275/ART20203995>
- [22] Cortes, C., Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 20(3): 273-297. <https://doi.org/10.1007/bf00994018>
- [23] Chen, T., Guestrin, C. (2016). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd Acm Sigkdd International Conference on Knowledge Discovery and Data Mining*, pp. 785-794. <https://doi.org/10.1145/2939672.2939785>
- [24] Speiser, J.L., Miller, M.E., Tooze, J., Ip, E. (2019). A comparison of random forest variable selection methods for classification prediction modeling. *Expert Systems with Applications*, 134: 93-101. <https://doi.org/10.1016/j.eswa.2019.05.028>
- [25] Dalianis, H., Dalianis, H. (2018). Evaluation metrics and evaluation. *Clinical Text Mining: Secondary Use of Electronic Patient Records*, 45-53. https://doi.org/10.1007/978-3-319-78503-5_6