

Received October 22, 2021, accepted November 6, 2021, date of publication November 10, 2021, date of current version November 23, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3127394

A Novel AI-Based System for Detection and Severity Prediction of Dementia Using MRI

VARUN JAIN¹, OM NANKAR¹, DARYL JACOB JERRISH¹, SHILPA GITE², SHRUTI PATIL², AND KETAN KOTECHA²

¹Symbiosis Centre for Applied Artificial Intelligence (SCAAI), Symbiosis International (Deemed) University, Pune 412115, India

²Computer Science and Information Technology Department, Symbiosis Institute of Technology, Symbiosis International (Deemed) University, Pune 412115, India

Corresponding authors: Shilpa Gite (shilpa.gite@sitpune.edu.in) and Ketan Kotecha (head@scaai.siu.edu.in)

This work was supported by the Symbiosis International (Deemed University), Pune, India.

ABSTRACT Dementia is a symptom of Alzheimer's Disease (A.D.) that affects many people around the globe each year. There is no effective cure to treat this disease, and it can prove to be deadly to the patient if left untreated or undetected. In this paper, the authors propose a novel DCGAN-based Augmentation and Classification (D-BAC) model approach to identify and classify dementia into various categories depending upon its prominence and severity in the available MRI scans. The proposed detection of early onset of dementia, also referred to as Mild Cognitive Impairment (MCI), is also studied with the help of a novel GAN-augmented dataset. The proposed model can predict MCI with an accuracy of 74% and can classify dementia into four categories depending upon its prominence in the MRI scan. The authors have also utilized Visual Explainable A.I. (XAI) and have used GradCAM to visually represent the internal working of the model. This novel approach helps verify the differentiating features of the MRI scans learned by the CNN model during training. Three different datasets, namely the original dataset, geometrically transformed images, and a GAN-augmented dataset, have been used for performance analysis. A comparison of the performance of the CNN model has been made on these datasets, and the superiority of results using the novel GAN-augmented dataset has been studied and discussed. Moreover, progressive resizing has also been applied on this GAN-dataset, and different CNN architectures have also been used to achieve better performance scores. The model proposed in the end has a training accuracy of 97% and a testing accuracy of 82% when tested using a conventional CNN architecture and has a testing accuracy of 84% and 87% when tested using VGG-16 and VGG-19 architecture, respectively.

INDEX TERMS Augmentation, dementia, generative adversarial networks, magnetic resonance imaging, mild cognitive impairment.

I. INTRODUCTION

Advancements in the field of medicine have led to better life expectancy around the world. However, a higher number of elderly citizens now have dementia. The symptoms involved in this disease include memory loss, improper brain functioning, impairment in Activities of Daily Living (ADL), and difficulties in communication and expression [1]. Early dementia starts in Mild Cognitive Impairment (MCI) in elderly patients and slowly develops to a harsher and much severe stage called Serious Dementia. Serious Dementia is a life-threatening disease, and there is no effective cure for

this disease to date. The only known method to detect and prevent this disease is in the early MCI stage itself. But existing State-Of-The-Art (SOTA) models and systems fail to predict MCI accurately, leading to Serious Dementia in a substantial number of elderly patients globally [2]. This paper discusses dementia and its early detection to prevent it from developing to later stages for which there is no cure. Estimates show that 32 million patients had been affected with dementia in a span of 15 years between 2000-2015 [2]. And it is estimated that there will be more than 135 million patients affected with dementia globally by 2050. In addition to this, the financial cost involved is far from cheap and unrealistically burdensome for both patients and hospitals. Research shows that timely intervention in the early stages of

The associate editor coordinating the review of this manuscript and approving it for publication was Jeon Gwanggil.

this disease can save about 7.9 trillion dollars globally [2]. Therefore, it is imperative to detect MCI and stop this disease from affecting individuals.

Artificial Intelligence (A.I.) has rapidly advanced and is being used in medical domains for different purposes ranging from tumor detection and classification to robot-based precision surgeries [3]. Deep Neural Networks (DNN) like Convolutional Neural Networks (CNN) [4]–[21], Long-Short Term Memory (LSTM) networks [28], [29], Recurrent Neural Networks (RNN) [30], [31] are being deployed for the detection and classification of dementia due to their ability to automatically extract information in the form of features and process on high-resolution MRI scans with perfect accuracy and with minimum intervention required from the researcher [7]. These models are beneficial over other conventional forms of A.I. algorithms like Support Vector Machines (SVM) [32], Hand-crafted feature learning machines [33], Ensemble Learning models [34], and Random Forest models (RF) [35] as these require manual feature extraction which is very difficult and time-consuming work. Despite all the meticulous work involved, these models [32]–[35] do not guarantee SOTA results [13]. This is why it has become increasingly necessary to opt for A.I. algorithms like DNN and CNN when working on Computer Vision (CV) and image-processing tasks [14]. A similar CNN model is proposed in this paper that takes Magnetic Resonance Image (MRI) scans and performs various operations like convolution, pooling, and dropout to extract features from these MRI scans without any manual intervention required from the researchers. These extracted features are processed to map them to their corresponding class label, and then MCI and severity of dementia are detected.

Researchers in the medical domain face a significant problem: the unavailability of good quality and class-balanced open-source datasets [15]. The datasets are not readily available for implementation in research, or the classes are highly imbalanced as they do not have the same number of samples for each class [15]. This poses an underfitting and biasing problem [15]. Researchers must manually collect MRI scans from medical institutions and design their dataset to implement a dementia classification model [16]. Medical institutions are hesitant to disclose sensitive patient data required for balancing the dataset, due to which the class imbalance problem persists. Researchers have to resort to conventional image augmentation techniques like flipping, rotating, and scaling images to balance the dataset and convert it to a machine-learnable [4]–[21], [23]–[30]. Although conventional image augmentation techniques work very well on other images to balance the dataset, they are not recommended when working on medical images. Medical images need to maintain their spatial conformity, but it gets altered when cropped, rotated, or flipped. These augmented images can be used for binary classification, i.e., when the model only distinguishes between A.D. patients and non-AD patients. Still, it does not work well when more than two classes are involved. A multi-label CNN model differentiates between MRI scans based on

the tumor size present in the scan. This tumor gets improperly cropped or rescaled when conventional augmentation is applied to it, resulting in mislabeled class samples.

A. D-BAC MODEL

The D-BAC (DCGAN-based Augmentation and Classification) model proposed in this paper tries to address these issues with the help of a GAN-based augmentation approach wherein the spatial conformity of the augmented images remains unchanged and unaltered. This helps train the CNN model smoothly and reduces misclassification of image samples, which helps achieve better performance metric scores. The classified images are visually presented using XAI to analyze the progression of dementia in a patient. Progressive resizing is also applied to the dataset to improve the performance metrics and scalability using different CNN models.

This paper proposes the following contributions to the field of AI-assisted dementia research:

1. The paper addresses the problem of improper augmentation of medical images using conventional augmentation techniques. A multi-label class-balanced novel dataset consisting of MRI images augmented using a custom-DCGAN is proposed.
2. Early-onset of dementia (MCI) is predicted, which goes unforeseen in the existing literature. This is done using models trained on all 3 datasets, and a comparison is made among them based on various metrics.
3. A multi-class CNN-based classifier is proposed to classify dementia depending upon the severity of the disease into four discrete classes. The existing models mostly revolve around binary classifications, which fail to predict the stage and severity of dementia.
4. The CNN models are visualized for their features using Grad-Cam, which helps to give a visual perspective of the tumor progression in MRI scans of patients with dementia.
5. The samples of the novel dataset are progressively resized, and different open-source architectures are applied to them to help improve the accuracy and metric scores of the model proposed in this paper.

The rest of the manuscript is divided into 4 sections –

Section II talks about existing literature and advancements in the field of dementia detection. Section III looks at the methodology and workflow proposed in this paper. Section IV discusses the results and their implications, followed by Section V which concludes the paper and discusses the future scope of this research. The last section acknowledges the reference works used in this research.

II. LITERATURE SURVEY

Dementia is a disease caused in elderly citizens due to a lack of exercise, inconsistent routine, and social interaction. It directly affects the brain of a patient. It can have adverse symptoms like memory loss, difficulty concentrating on a particular task, confusion leading to forgetfulness while performing daily tasks such as brushing teeth, bathing,

TABLE 1. A comprehensive review of existing literature.

Reference	Task	Method of Implementation	Dataset used	Limitations
[4]	Classification	CNN	ADNI	Difficult to integrate 2 different models
[5]	Classification	CNN	ADNI, MILAN	Failed to predict early-onset and advancement of Dementia
[6]	Classification	SCNN	OASIS	Only binary classification
[7]	Segmentation, Classification	U-net like CNN	ADNI	Lack of training data, No augmentation
[8]	Regression, Classification	CNN	ADNI	Limited regression accuracy and sparse input samples
[9]	Classification	DCA	ADNI, OASIS	Poor testing accuracy
[10]	Classification	2D + 3D CNN	N.G. Hospital, China	Inadequate data provided by the hospital
[11]	Classification, Augmentation	TF-CNN	ADNI	Only binary classification failed to predict early-onset
[12]	Classification	ResNet, GoogLeNet, DenseNet	ADNI	Unbalanced dataset, Failed to predict severity of Dementia
[13]	Classification	1D Temporal CNN	ADNI	Hand-crafted features, Time-consuming, No guarantee of good accuracy
[14]	Prediction	Deep InfoMax (DIM)	ADNI	Low accuracy scores
[15]	Literature review	Neural Networks	A.D. Articles	Inadequate results
[16]	Super-resolution images	GAN	Lung MRIs	Resource Intensive, Doesn't work well with limited hardware
[17]	Classification	CNN	MRI scans, FDG-PET scans	Inadequate results as a sparse dataset were used
[18]	Classification	GAN	PET images, MRI scans	Inadequate results
[19]	Classification	CNN	Neuroimaging data	Only binary classification, Failed to predict early-onset, Severity of Dementia is not predicted
[20]	Classification	CNN	CT scans	Lack of adequate data
[21]	Classification	Enhanced Inception (V3)	MRI images	Lack of data, No augmentation

cleaning, etc., and frequent mood swings. The only known method to detect this disease is through MRI scans of the patient's brain. There is no cure for this disease, and the only possible way to stop it from affecting patients is by detecting it in its early developmental stages. This early stage is also known as Mild Cognitive Impairment (MCI). Dementia progresses to more severe and harsher forms from the MCI stage onwards. There have been various methods and models used in the past for early detection of dementia and for preventing dementia altogether. These approaches are based on the MRI scans of patients and other symptoms leading to dementia, such as gait analysis, handwriting analysis, and consistency in performing ADLs (Activities of Daily Living).

A. MEDICAL DATA-BASED MODELS

Davis-Owusu *et al.* [1] proposed a novel approach to preventing dementia in elderly citizens by building a bidirectional activity-based system. This system helped in coordinating the actions and routines of elderly citizens and their caregivers.

This system helped create a sense of belongingness and attachment in the elderly citizens and averted feelings of detachment and loneliness, which helped effectively prevent dementia. Pundane *et al.* [3] proposed another novel approach in which they studied various Ambient Assisted Living (AAL) systems. These systems are meant to assist elderly citizens in their Activities of Daily Living (ADLs) so that they do not have to deviate from their routine and perform their daily activities with ease. They studied various mobility-based systems that help give a sense of freedom to elderlies so that they do not have to depend on others for moving around and performing exercises. They also studied social interaction-based systems, which help elderlies stay socially and emotionally connected to their relatives and other elderly citizens. As proposed in [2], such systems help give a sense of freedom and liberty to elderly citizens, which helps to avert dementia and other related diseases in elderly patients.

Sabo *et al.* [36] proposed a novel gait-based analysis approach for predicting dementia in the citizens of a

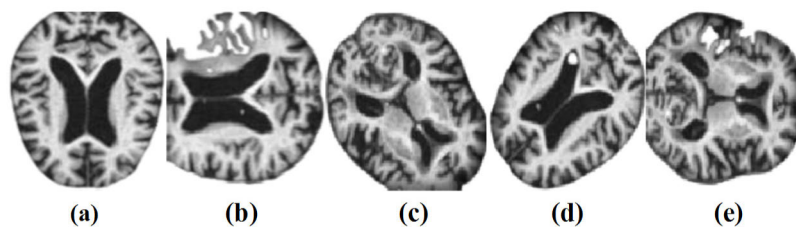


FIGURE 1. Images augmented using geometric transformations. (a) Original (b) Cropped (c) Scaled (d) Rotated (e) Cropped.

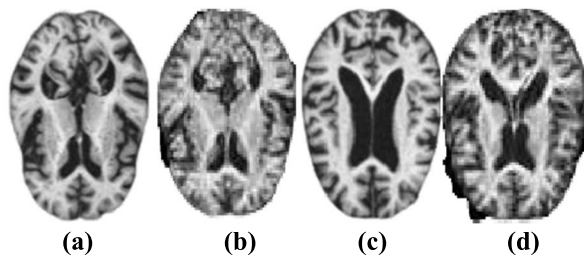


FIGURE 2. GAN-augmented images (a) MCI real (b) MCI fake (c) SD real (d) SD fake.

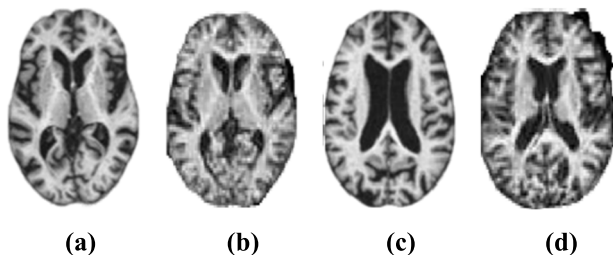


FIGURE 3. GAN-augmented images rotated 180° in anterior-posterior view (a) MCI real (b) MCI fake (c) SD real (d) SD fake.

residential society. They used a two-stage pipeline procedure consisting of Graph CNN called ST-GCN. This approach helped detect dementia in citizens by analyzing abnormalities in their gait patterns compared to the gait patterns of normal subjects. Various pose-estimation-based libraries are used for this purpose, but this approach was a bit complex and challenging to implement in a real-life setting. Ma *et al.* [37] proposed a novel finger-tapping and gait-based combinatorial approach for dementia analysis and prevention. Finger-tapping pattern data like time intervals, frequency, speed, etc., were collected and combined with the gait pattern data of the patients. This data was fed to an ML model to improve fall-risk predictions in dementia patients. This approach required extensive data collection and thorough cleaning of data before it could be implemented. Luz *et al.* [38] proposed a speech-based approach to detect and analyze the rate of cognitive decline. A combination of acoustic and linguistic recordings is used to monitor patients’ cognitive status. This required extensive cleaning of audio samples and routine and timely collection of audio data from patients, which were time-consuming and laborious. Rohanian *et al.* [39] presented another audio-based approach that combined audio recordings with text data to predict cognitive impairment. Both regression and classification models were used to monitor

the rate of cognitive decline. Creating the dataset through data collection of diverse patients and then cleaning the data was a very laborious task. Cohen and Pakhomov [40] proposed a linguistic model trained on the Dementia Bank dataset. The model was a mixture of healthy controls and dementia-based linguistic samples and could predict dementia in patients effectively. This model used a readily available and pre-cleaned dataset (DementiaBank), making its implementation simpler. But creating the dataset for such a model could be a tedious task, and its implementation could be even more tedious.

B. MRI IMAGE-BASED MODELS

After reviewing and analyzing the drawbacks of the above literature, the authors came across another well-renowned method of dementia prediction through MRI brain scans of dementia-affected patients. This method involved various classification algorithms like CNN, LSTM, RNN, etc., applied on MRI scans that helped differentiate and classify different levels of dementia. The review of these papers and their drawbacks is presented below.

Li *et al.* [4] proposed a novel approach to classify dementia into two categories – Normal Cognition (N.C.) and Alzheimer’s Disease (A.D.) using CNN. They applied feature maps of CNN to build a time-to-event model with additional features related to A.D. to classify dementia [4]. The dataset used was a combination of ADNI and AIBL. Model training and learning were done with the help of a CNN [4]. A regression model was also built using the features learned by the CNN model. However, determining results from two different models has 2 significant drawbacks - It is time-consuming to train different models, and it gets difficult for untrained individuals to integrate these models. Basaia *et al.* [5] proposed a deep-learning model to predict A.D. and MCI. Simple data augmentation techniques (scaling, rotation, translation, etc.) were applied to address the overfitting problem. They were able to achieve exceptional results on popular datasets like ADNI and MILAN [5]. However, these results applied only to a certain number of unseen datasets, and they failed to consider the advancement of dementia to later stages. Amin-Naji *et al.* [6] built a Siamese Convolutional Neural Network (SCNN) to detect dementia using features maps generated from a Res-Net model. They used an unsupervised approach to classify dementia into two categories which is helpful when unlabelled datasets are involved in the research

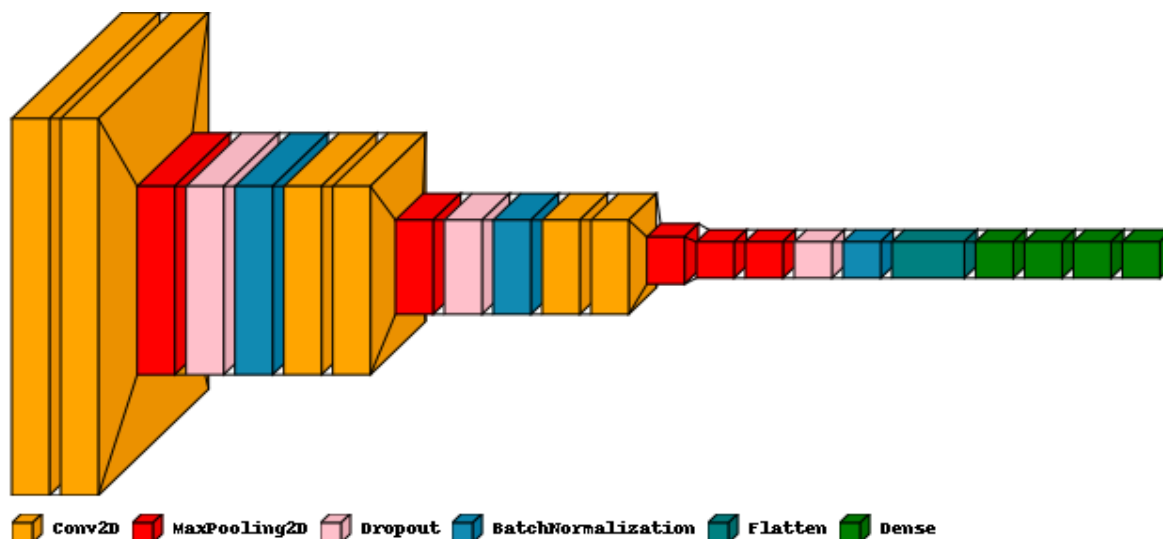


FIGURE 4. CNN architecture used in this study.

and annotating labels to each sample is difficult [6]. The drawback of this approach was that SCNN could only classify dementia into two classes, and its functionality was limited. SCNN also failed to predict early dementia or MCI.

Kavitha *et al.* [7] proposed a U-net-like Convolutional Neural Network (CNN). This model was used to segment images and classify dementia into four classes. Remarkable results were obtained on the ADNI dataset [7]. The only limitation to this approach was that the training data was small and limited [7], and no augmentation techniques were applied to increase the number of image samples. This model cannot generalize well to future unseen images, leading to misclassification of dementia and a poor accuracy score. Li and Liu *et al.* [8] proposed a combination of classification and a regression model to predict the diagnosis of dementia. Feature maps from CNN were used with the clinical scores obtained from the regression model to classify dementia into four classes [8]. ADNI dataset was used. However, the regression accuracy was limited and saturated after a certain point, and input samples were sparse, posing an overfitting problem as no augmentation was applied [8].

Martinez-Murcia *et al.* [9] proposed a method to determine A.D. using Deep Convolutional Autoencoder (DCA). DCA was composed of an encoder and decoder, which were used to down-sample and up-sample the MRI scans to learn the underlying features in the images [9]. The limitation was that it failed to learn the statistical patterns and map the input MRI images in different categories, leading to inferior accuracy scores. Gao *et al.* [10] proposed a model in which a 2D-CNN model and a 3D-CNN model were used. The final layer of this combined model-averaged the results of feature maps obtained from both models to predict the diagnosis of A.D. [10]. However, this work was hugely dependent on a readily available balanced dataset and many MRI images that are not simply available in the medical domain. No augmentation techniques were used

to address the class imbalance problem posed by medical datasets.

El-Sappagh *et al.* [41] proposed a Machine learning model using random forest classification. The experiments were conducted using the ADNI dataset. The results obtained were as follows; the first layer had a cross-validation accuracy of 93.95% and an F1-score of 93.94%, and the second layer had a cross-validation accuracy of 87.08% and an F1-Score of 87.09%. The model consisted of 2 layers; for the first layer, features and predictors were discovered, MMSE being the essential feature for the Alzheimer's disease class, and Clinical dementia rating sum of boxes (CDRSB) as the rating scale or predictor for C.N. and MCI classes. The Functional Assessment Questionnaire was the most important feature for the second layer's sMCI (Stable) and pMCI (progressive MCI) classes. One of the limitations was the use of R.F. classification, which can be slow for real-time predictions that raise questions about the model's practicality; also, the classification was only for Mild cognitive impairment and did not cover other stages based on the intensity of the disease.

Venugopalan *et al.* [42] proposed using deep learning models; predictions by multi-modality data are better than single-modality data. The three setups implemented and discussed on the ADNI dataset are: electronic health records (EHR) + single nucleotide polymorphisms (SNP), electronic health records+Imaging+single nucleotide polymorphisms, and electronic health records+Imaging. In the external test performance: The EHR produced Accuracy: 0.76 Precision: 0.76 Recall: 0.77 F1 Scores: 0.76, imaging Accuracy: 0.84 Precision: 0.83 Recall: 0.83 F1 Scores: 0.83, snp Accuracy: 0.66 Precision: 0.66 Recall: 0.57 F1 Scores: 0.53, EHR + SNP + Imaging Accuracy: 0.78 Precision: 0.77 Recall: 0.78 F1 Scores: 0.78, EHR+SNP Accuracy: 0.78 Precision: 0.78 Recall: 0.79 F1 Scores: 0.78, EHR+imaging Accuracy: 0.77 Precision: 0.76 Recall: 0.77 F1 Scores: 0.77, snp+imaging Accuracy: 0.63 Precision: 0.62 Recall: 0.57 F1 Scores: 0.56.

TABLE 2. Parameters and layers involved in the proposed CNN model.

Sr. No.	Layer	Output shape	Kernel size	Stride size	Pool size	Activation function	No. of parameters
1	<i>Conv2d</i>	176x208x32	3	1	-	ReLU	320
2	<i>Conv2d</i>	176x208x32	3	1	-	ReLU	9248
3	<i>Max Pooling</i>	88x104x32	-	-	2	-	0
4	<i>Dropout</i>	88x104x32	-	-	-	-	0
5	<i>Batch Normalization</i>	88x104x32	-	-	-	-	128
6	<i>Conv2d</i>	88x104x32	3	1	-	ReLU	18496
7	<i>Conv2d</i>	88x104x32	3	1	-	ReLU	36928
8	<i>Max Pooling</i>	44x52x64	-	-	2	-	0
9	<i>Dropout</i>	44x52x64	-	-	-	-	0
10	<i>Batch Normalization</i>	44x52x128	-	-	-	-	256
11	<i>Conv2d</i>	44x52x128	3	1	-	-	73856
12	<i>Conv2d</i>	44x52x128	3	1	-	-	147584
13	<i>Max Pooling</i>	22x26x128	-	-	2	-	0
14	<i>Max Pooling</i>	11x13x128	-	-	2	-	0
15	<i>Max Pooling</i>	5x6x128	-	-	2	-	0
16	<i>Dropout</i>	5x6x128	-	-	-	-	0
17	<i>Batch Normalization</i>	5x6x128	-	-	-	-	512
18	<i>Flatten</i>	3840	-	-	-	-	0
19	<i>Dense</i>	512	-	-	-	ReLU	1966592
20	<i>Dense</i>	256	-	-	-	ReLU	131328
21	<i>Dense</i>	128	-	-	-	ReLU	32896
22	<i>Dense</i>	4	-	-	-	Softmax	516
	<i>Total</i>	-	-	-	-	-	2,418,660

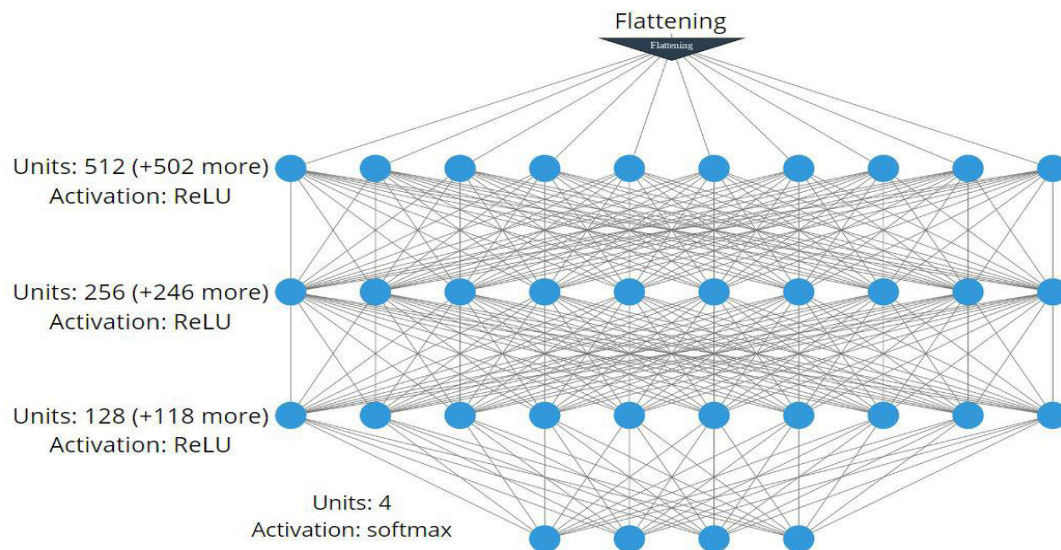


FIGURE 5. The final fully connected and dense layer of the model culminating in 4 output nodes. Each output node represents a particular class of dementia.

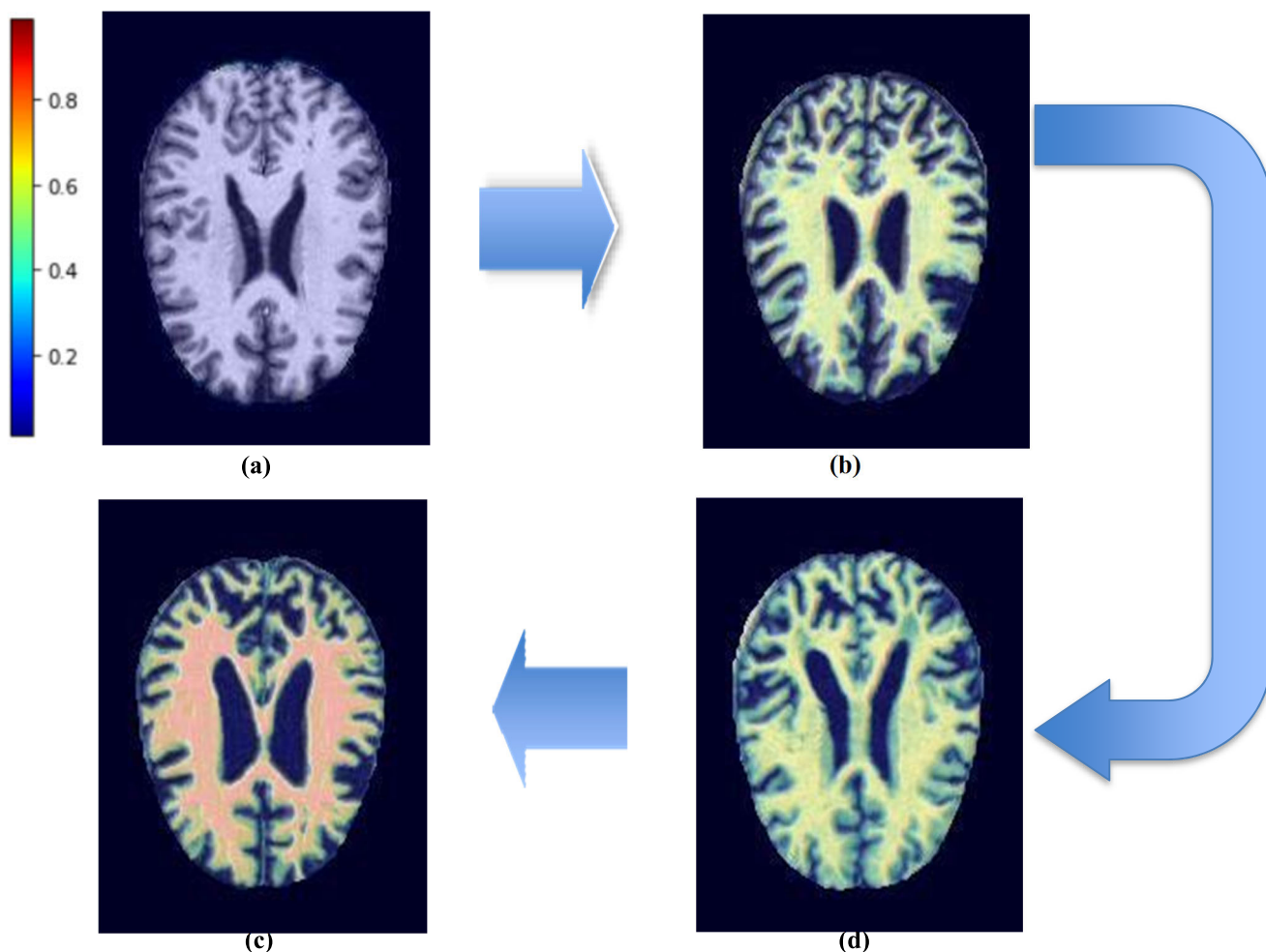


FIGURE 6. Grad-Cam heatmap of MRI scans visually representing the progression of dementia in patients with Color-Bar for GradCAM heatmap showing intensity range for different dementia levels (a) ND (IV = 0.2) (b) MCI (IV = 0.4) (c) MD (IV = 0.5) (d) SD (IV = 0.7).

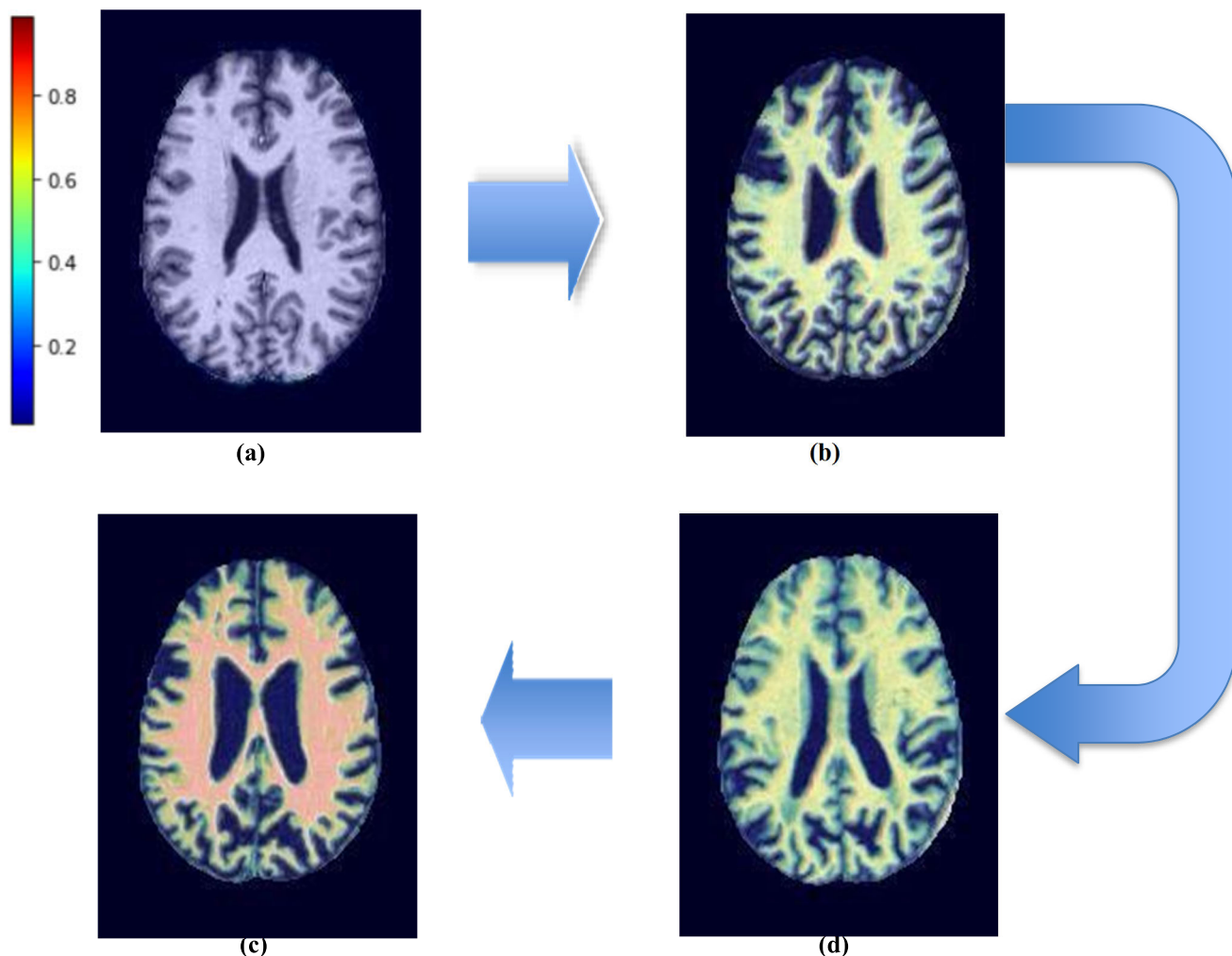


FIGURE 7. Grad-Cam heatmap of MRI scans rotated in anterior-posterior view (a) ND (IV = 0.2) (b) MCI (IV = 0.4) (c) MD (IV = 0.5) (d) SD (IV = 0.7).

As observed, the results are not good enough, although the bottleneck is less data which has been mentioned in the paper. The lack of visualization can also be pointed out. Odusami *et al.* [43] suggested binary classification on ResNet-18 network trained with and without dropout. The paper proposed a method to predict MCI, early MCI (EMCI), late MCI (LMCI), and A.D. The trained ResNet18 network achieved a classification accuracy of 99.99% on EMCI vs. A.D., 99.95% on LMCI vs. A.D., and 99.95% on MCI vs. EMCI classification scenarios, respectively. The limitation of the model is the lack of visualization, which has been mentioned in the future scope. The data was taken from the ADNI dataset but no augmentation was applied.

Han *et al.* [44] proposed an unsupervised medical anomaly detection GAN (MADGAN) to detect brain anomalies at different stages on multi-sequence structural MRI. This method helped reliably predict the successive three MRI slices from the previous three images for unseen healthy images and detect subtle anatomic anomalies and lesions. The drawbacks of this method to the dataset chosen are that this method requires a sizeable balanced dataset and

does not account for the detection of the different stages of dementia. Spasov *et al.* [45] proposed a parametric-efficient method combining MRI, demographic, neuropsychological, and APOe4 genotyping data to predict MCI to A.D. conversion simultaneously and A.D. vs. healthy classification. This method effectively predicted conversion from mild cognitive impairment to Alzheimer’s disease within three years. Still, it could not predict the exact stage at which a patient’s dementia would be without the presence of a sizeable balanced dataset.

C. AUGMENTATION AND OPTIMIZATION-BASED MODELS

Additional literature focuses on conventional data augmentation methods like rotation, cropping, scaling and resizing, and optimization techniques that produce the best results on medical images.

Taqi *et al.* [11] studied data augmentation and various optimization techniques that could be used for medical images. They found that RMSProp and Adam optimizer gave the best results compared to other optimizers when working on medical images [11]. Senanayake *et al.* [12] worked on A.D.

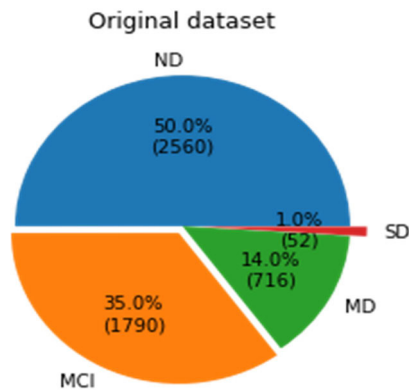


FIGURE 8. Distribution of original dataset (imbalanced).

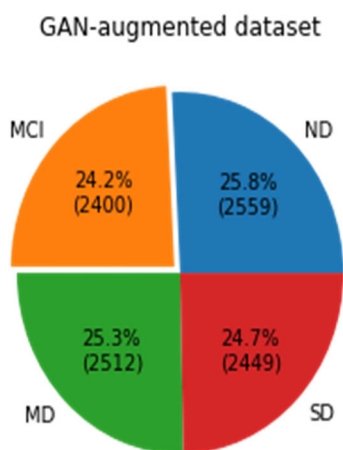


FIGURE 9. Distribution of GAN-augmented dataset (balanced).

diagnosis by fusing different residual networks like ResNet, GoogLeNet, and DenseNet to reduce the complexity of the datasets and enable effective fusion. They achieved good results using this method for binary classification, but the classification of dementia based on severity and unbalanced datasets weren't studied by them [12]. The binary classification was done between different categories of dementia, such as N.D. v/s MCI and N.D. v/s S.D., and the results were compared with SOTA models. The results obtained using different CNN models were significantly better than other conventional models, but the proposed approach doesn't work well in real-world scenarios [12]. The complexity keeps increasing as the number of binary classifiers increases, and it becomes difficult for untrained medical officials to decide which classifier to use. This model also lacked in predicting early dementia. Jarrett *et al.* [13] posited a 1D temporal convolutional network to classify dementia using hand-engineered features. They achieved good accuracy but extracting features manually is time-consuming and laborious and doesn't always guarantee good accuracy scores. Early dementia wasn't studied by [13].

Fedorov *et al.* [14] proposed an unsupervised model called Deep InfoMax (DIM) to predict the progression of A.D. This model was used to explore the brain structure

TABLE 3. Comparison of time taken by the same model to train and test on different datasets.

Time/Dataset	Training (mins)	Testing (mins)
Original	50	15
Geometrically transformed	90	30
GAN-augmented	102	38

in a flexible non-linear manner. It performed well. However, this was a novel implementation methodology, and it didn't perform well compared to other existing models [14]. Valverde *et al.* [15] performed a systematic literature review for articles that applied transfer learning to MRI brain scans, categorized them, and extracted relevant information. Patterns were recognized, which showed the models that worked best with transfer learning on MRI data [15]. Alam *et al.* [16] proposed a GAN model that generates new data with the same statistical patterns as the training set using a generator as the network to improve the resolution and discriminator to train the generator better. This helped in creating super-resolution MRI images [16]. However, this model was very resource-intensive and could not provide possible results with limited hardware capabilities. Lu *et al.* [17] proposed the early diagnosis of A.D. by patch-wise feature extraction from structured M.R. and FDG-PET images. They found that the performance of classifiers built using a combination of FDG-PET and structural MRI images was better than those built using structural MRI or FDG-PET scans alone [17]. However, this model could not perform well with sparse datasets and only with extensive datasets rarely available in the medical field. Early-onset wasn't studied. Shin *et al.* [18] proposed a model that uses a GAN with discriminator-adaptive loss fine-tuning using PET scans and MRI scans for A.D. diagnosis. Unlike other GANs, this model uses A.D. diagnosis in the GAN training process to better A.D. classification performance [18]. However, other deep CNN and GANs provided better results on the testing data than the model proposed in [18]. Jo *et al.* [19] proposed a deep learning model to classify A.D. by using a combination of fluid biomarkers and multimodal neuroimaging data. It was used to detect structural and functional biomarkers for A.D. in the MRI scans. It produced good results but was limited to a binary classification of dementia, reducing its functionality [19]. Early dementia wasn't studied by [19].

Cui *et al.* [20] proposed a 3D-like approach by using deep learning to extract information using 2D slices and 3D blocks of C.T. scans of the brain. An advanced CNN was used, which integrated both 2D and 3D CNN networks [20]. It produced high-quality results when predicting A.D. but was limited due to the lack of readily available C.T. images. Goodfellow *et al.* [21] proposed an enhanced version of the Inception (V3) neural network on MRI scans. They used a method called "Multi-Attention Combination" to learn the tumor size efficiently [21]. This was applied to the existing

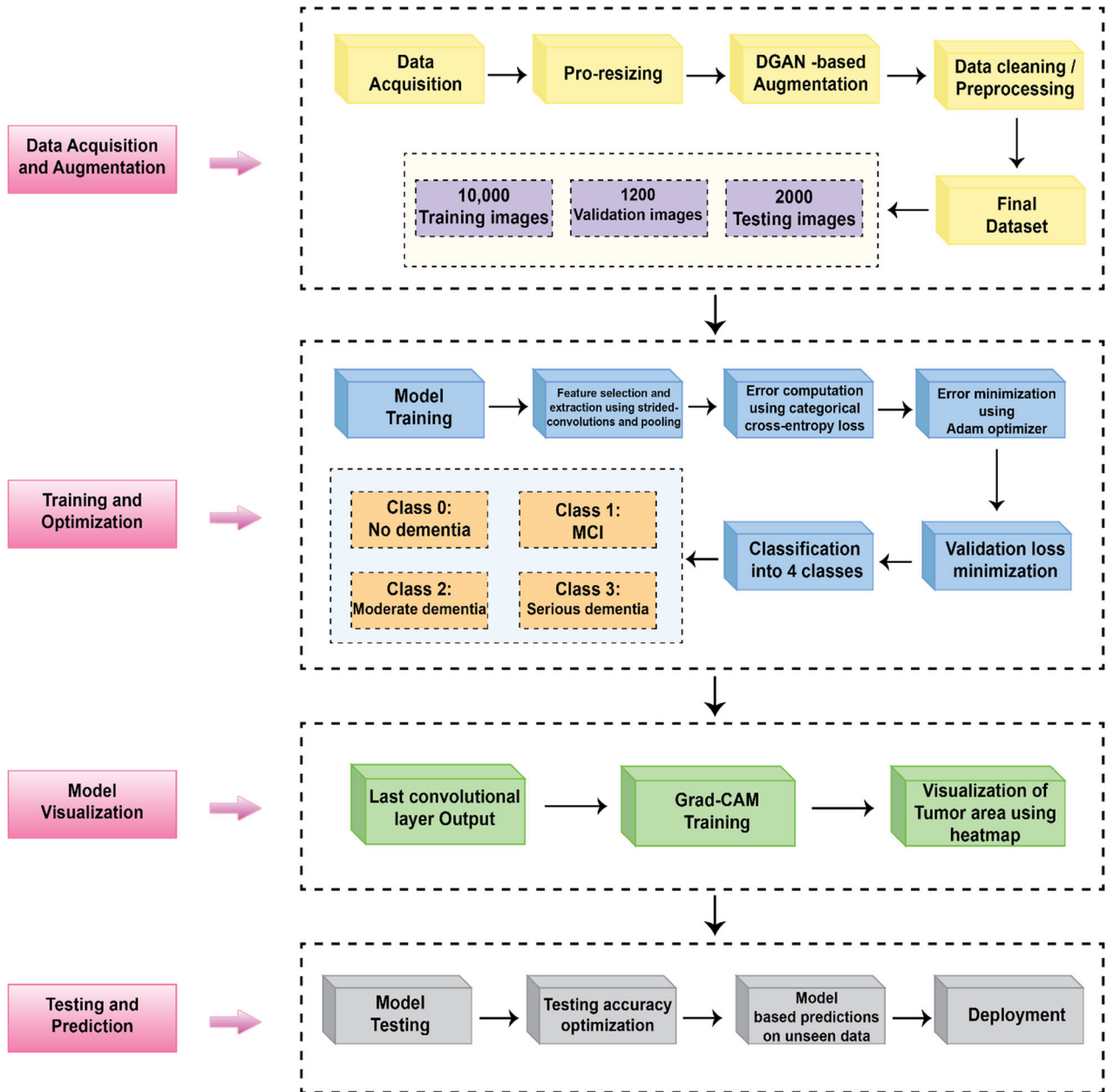


FIGURE 10. System workflow representing the course of action at each stage while building an ML model.

Inception model to enhance its performance. However, its accuracy was limited due to the lack of available data, and the authors did not use any augmentation techniques to increase the dataset size [21]. Early-onset of Dementia wasn't studied in [21]. Spasov *et al.* [45] proposed a Progressively growing GAN (PGGAN) augment brain MRI, which used a multi-staged generative training method to infinitely generate 256×256 sized M.R. images for tumor detection. Although this method was very effective in generating brain MRI, it required a large balanced dataset to train, without which the images would be subpar. Han *et al.* [46] proposed a method that uses a Super-Resolution GAN to learn 3D shape variations in adult brains and a pix2pix GAN to upgrade images slices with realistic local contrast patterns to synthesize 3D

brain MRI images. This method effectively generated realistic 3D MRI images with high accuracy, with its only limitations being feature representation between slices.

III. METHODOLOGY

This section explores the implementation methodology followed in this research.

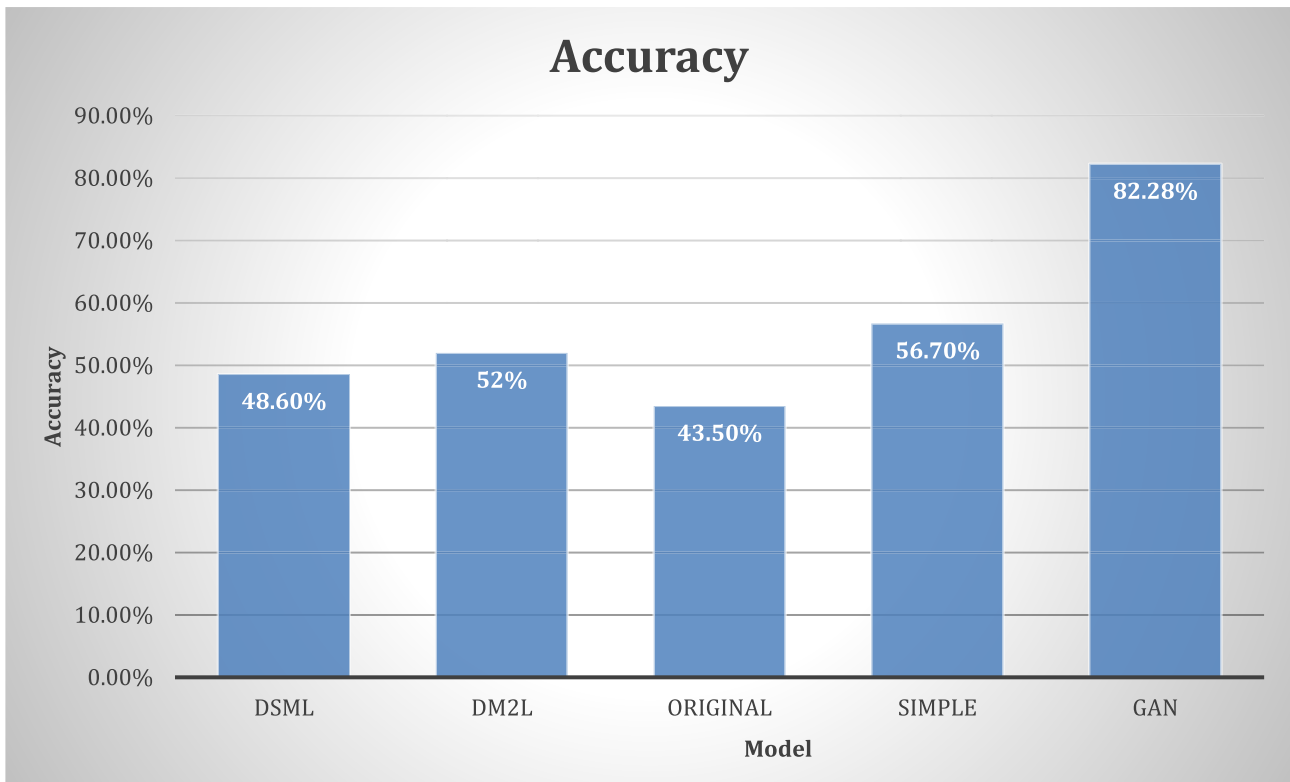
A. AUGMENTATION USING DCGAN TO GENERATE A NOVEL DATASET

Generative Adversarial Network (GAN) is a data augmentation model used to balance datasets, generate new data samples, and learn features [22]. A generator (G) generates

TABLE 4. Accuracy metric comparison of proposed D-BAC system with existing SOTA models.

Model	DSML [8]	DM2L [8]	Original dataset	Dataset augmented using geometric transformations	GAN-augmented dataset
No. of features	-	-	2,418,660	2,418,660	2,418,660
Accuracy	48.6%	52%	43.5% ($\pm 2\%$)	56.7% ($\pm 2\%$)	82.28% ($\pm 2\%$)

Note: “-” represents an unknown value; “ \pm ” represents standard deviation.

**FIGURE 11.** Graphical comparison of the accuracy of D-BAC with existing SOTA model.**TABLE 5.** Comparison of existing SOTA binary classifiers based upon accuracy.

Reference	Dataset	Accuracy
CNN [5]	ADNI	99.2%
SCNN [6]	OASIS	98.72%
3D CNN [10]	ADNI	87.62%
Inception [21]	ADNI	85.7%
DenseNet-121 [36]	ADNI	94.97%

new images identical to the standard data distribution of the existing data samples, and a Discriminator (D) classifies those images as real or fake [23]. The Generator and the Discriminator compete closely in a min-max game to improve their accuracy on their respective tasks [24].

The GAN model used in this paper is called the Deep Convolutional GAN (DCGAN) [25]. Various types of GAN models are available depending upon the type and number of layers, activation functions used, and the complexity of

the input dataset [23]. DCGAN uses a CNN-type architecture to build the Generator and the Discriminator [25]. It replaces the pooling layers with additional convolutional layers so that the Generator and Discriminator can learn their up-sampling and down-sampling [25]. It also prevents important features from getting dropped. This model is preferred in this research because feature dropping is prevented, which can cause essential features like the brain tumor and severity of dementia to be dropped altogether. This causes

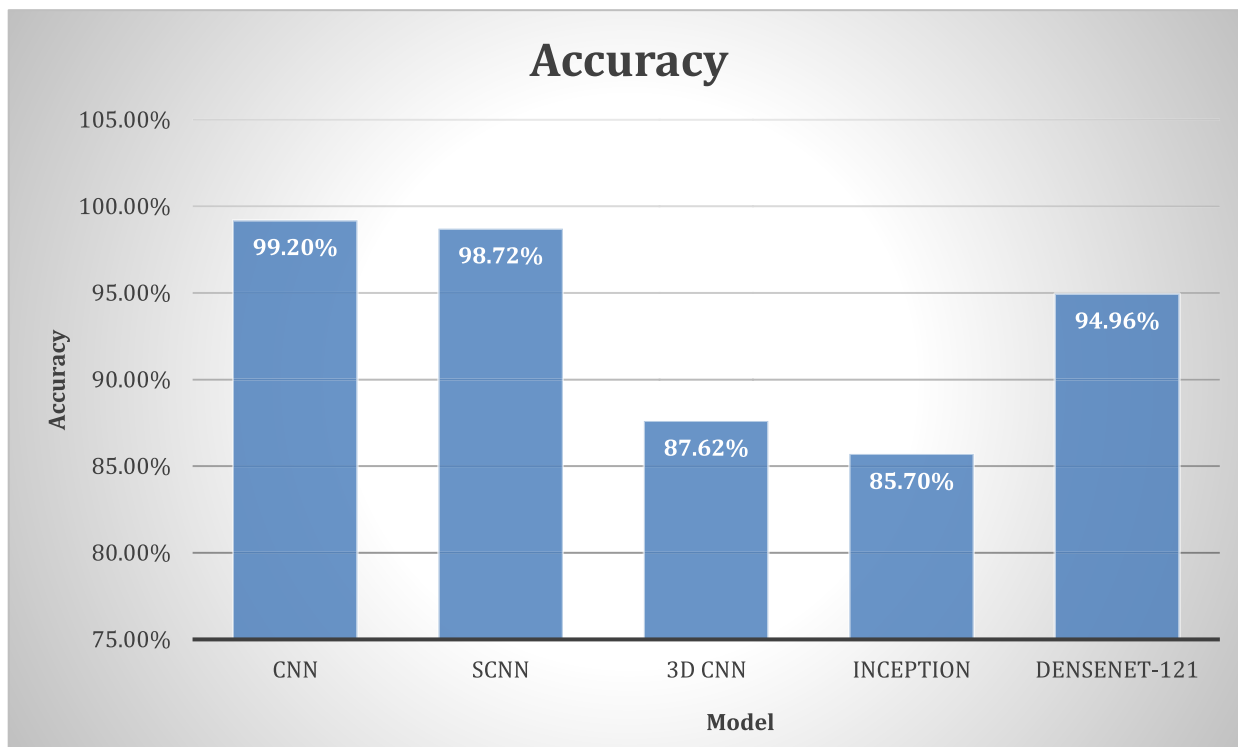


FIGURE 12. Graphical comparison of the accuracy of SOTA binary classifiers.

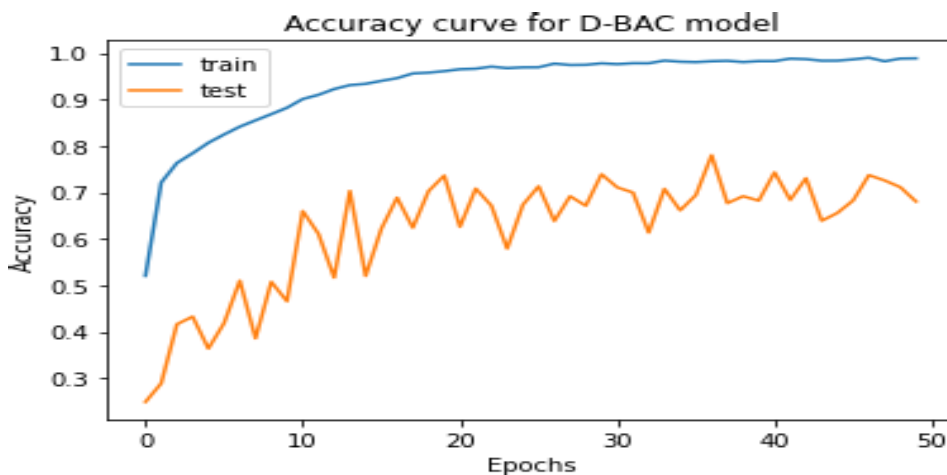


FIGURE 13. Accuracy curve of proposed D-BAC model.

the generated images to deviate from the normal distribution of the original images, which results in mislabeled samples. Custom changes like the resolution of output images, batch size, etc., were done to generate suitable images for the dataset. The images generated from the DCGAN were used to balance the imbalanced dataset. The novel balanced dataset consists of 2500 samples per class. Images augmented using geometric transformations, and GAN-augmented images are displayed in Fig.1 and Fig. 2.

B. DATA PRE-PROCESSING

Input data samples were pre-processed using two methods - Cleaning and Normalization.

1) CLEANING

The GAN-augmented dataset was checked manually for deviant samples. These samples included those that deviated from the normal distribution of other samples belonging to that class label. These deviant samples often reduce the accuracy of the model and are called mislabeled samples. It is necessary to discard these samples so that they do not inhibit the accuracy and performance of the model. Each class had some of the other deviant samples, which didn't represent the normal distribution of that class. The tumor size in MCI scans is minor, so some GAN-generated samples did not correctly capture the tumor size of MCI. The tumor in these augmented scans was absent, and they fell under the N.D. label but were

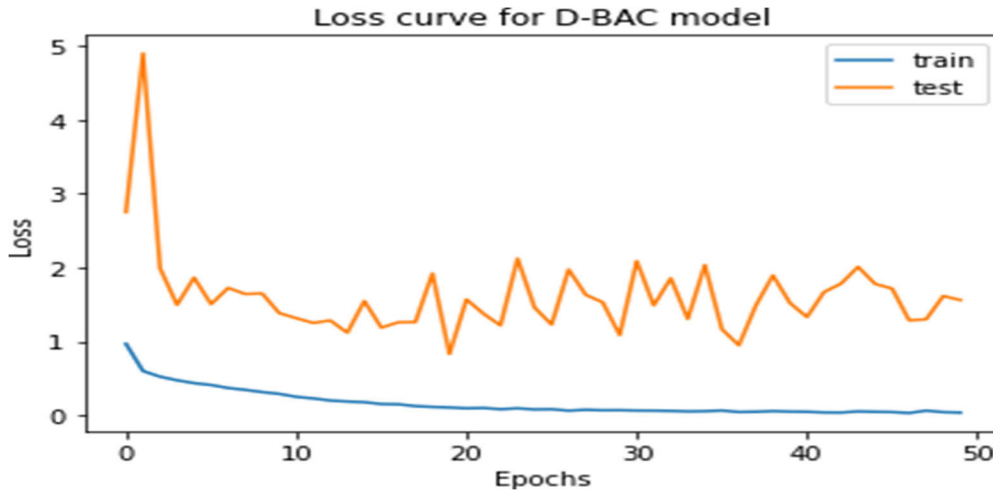


FIGURE 14. Loss curve of proposed D-BAC model.

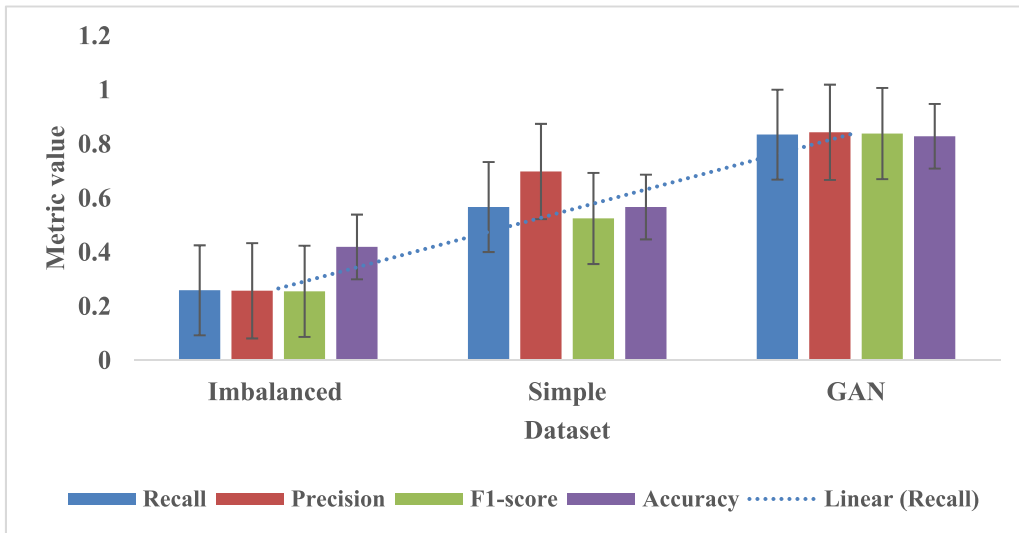


FIGURE 15. Graphical comparison of performance metrics of the 3 datasets.

mislabelled as MCI. Such samples needed to be removed from the MCI category and similarly from other categories as well.

2) NORMALIZATION

Normalization was applied to the input dataset, which consisted of 8,000 images. Normalization reduces the variance and overfitting of the dataset by the CNN model by scaling down the feature values to the range \mathbb{R} (0-1). This is done so that the model can also generalize well to the class of unseen samples (samples not used for training). Generally, input features are unevenly distributed in the feature map. The features are mapped in the range \mathbb{R} (0-255), a wide scale. This causes high variance in the dataset, and the model might start to overfit the dataset if it is not normalized. Overfitting tends to give perfect validation accuracy but inferior test accuracy. This causes a significant drop in the performance of the model after deployment. Therefore, normalization was applied to the dataset to scale the features down to a much narrower scale.

C. CONVOLUTIONAL NEURAL NETWORK (CNN) MODEL

A conventional Convolutional Neural Network (CNN) model was used to classify the disease into different classes. 6 Convolutional layers were used with a kernel size of 1 and a stride size of 3 for feature extraction and learning. 5 Pooling layers were used with a pool size of 2 to drop excess features and speed up computation. Batch normalization and dropout regularization with a value of 0.5 was applied after every convolutional layer to minimize variance. Finally, a flatten layer and four dense layers were added to the end of the model to flatten all the features into a single layer. This particular architecture of CNN was selected as the authors wanted to assess the performance of a rudimentary and straightforward CNN model on the custom dataset. More sophisticated and optimized CNN models and their performance on the dataset have been discussed later in the paper.

Four output nodes in the final layer represent each class of dementia. The model was trained over 2,418,660 parameters. A Rectified Linear Unit (ReLU) activation function was used

TABLE 6. Comparison of various performance metrics for the 3 datasets.

Metric/Dataset	Precision	Recall	F1-score	Accuracy
Original	0.2576	0.2593	0.2584	0.4197
Geometrically transformed	0.6985	0.5672	0.6260	0.5672
GAN-augmented	0.8435	0.8348	0.8388	0.8228

TABLE 7. Confusion matrix of the proposed 4-class classification model.

		Predicted			
		ND (0)	MCI (1)	MD (2)	SD (3)
Actual	ND (0)	P_{00}	N_{10}	N_{20}	N_{30}
	MCI (1)	N_{01}	P_{11}	N_{21}	N_{31}
	MD (2)	N_{02}	N_{12}	P_{22}	N_{32}
	SD (3)	N_{03}	N_{13}	N_{23}	P_{33}

Note: “P” represents positive samples, and “N” represents negative sample

in each layer, excluding the last layer. This function is linear in the first quadrant and 0 in all the other quadrants. The last layer uses a SoftMax activation function that predicts each class’s probability and outputs the class with the highest probability. A Categorical Cross-Entropy loss function was used to compute the error, and an Adaptive Moment Initializer (Adam) was used to optimize the model. Fig. 4 shows the CNN architecture used in this study, and Fig. 5 shows the final flattening and dense layers. Table 2 tabulates the layers in the CNN model along with their parameters.

D. GRADCAM FEATURE VISUALIZATION

Gradient-weighted Class Activation Mapping (GradCam) uses the gradients computed in the final CNN layer to generate a heatmap highlighting the features learned from the input MRI scans. This heatmap represents the different spatial conformities that the CNN model learns for different classes of MRI scans during the training phase. The differentiation learned is then used to predict the class of an unseen MRI test scan accurately. Fig. 6 shows the heatmap of some MRI scans taken from the test set used in this research.

The blue color is most prominent for the N.D. scan and least prominent for the S.D. scan. This was referred to as “blue shift” and represented how the CNN model differentiates between different classes of MRI scans.

The spatial conformity learned by the CNN model is represented in these heatmaps, and this helps to accurately distinguish between different classes of test samples and make appropriate predictions. This also helps in the early prediction of dementia or MCI. The CNN model takes a training image as an input and applies operations like convolutions, pooling, and dropout to learn, extract and discard features. The order and conformity of these features in the plane represent a particular class and severity of dementia. This conformity varies with the features of each class, such as the tumor size. Different colored heatmaps represent the

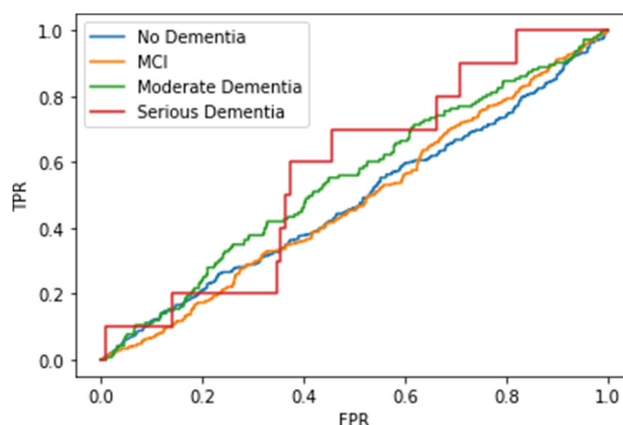


FIGURE 16. Receiver operating characteristics (ROC) curve of the original dataset.

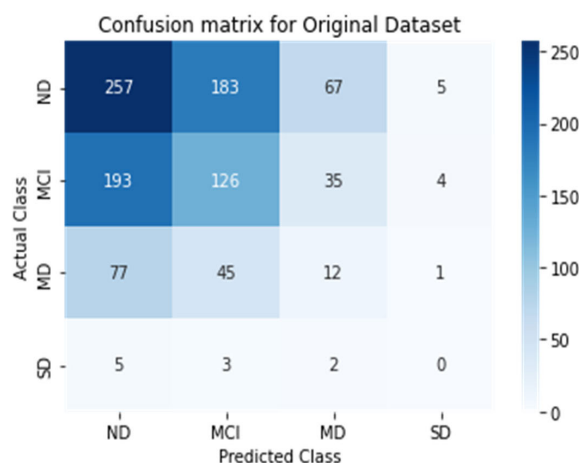


FIGURE 17. Confusion matrix (C.F.) of the original dataset.

different feature distributions in space used by the model to distinguish between different samples. This heatmap helps to make informed predictions about the early onset and severity of dementia. Each MRI scan, along with its intensity value (IV), is presented in Fig. 6.

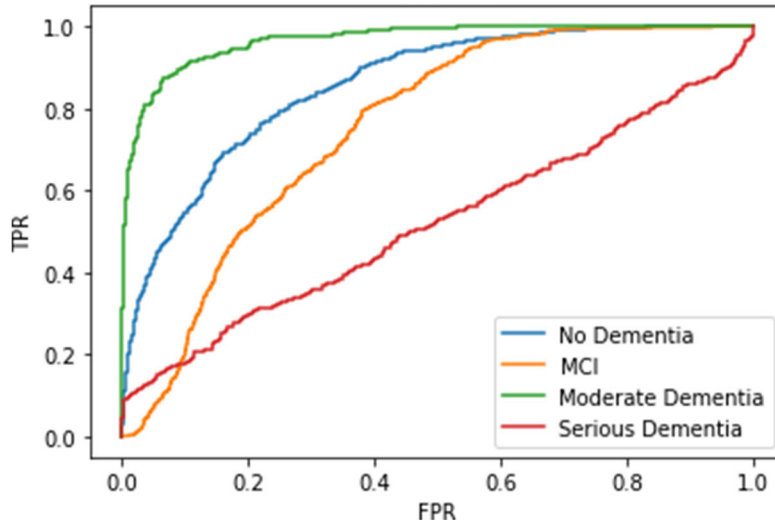


FIGURE 18. Receiver operating characteristics (ROC) curve of dataset augmented using geometric transformations.

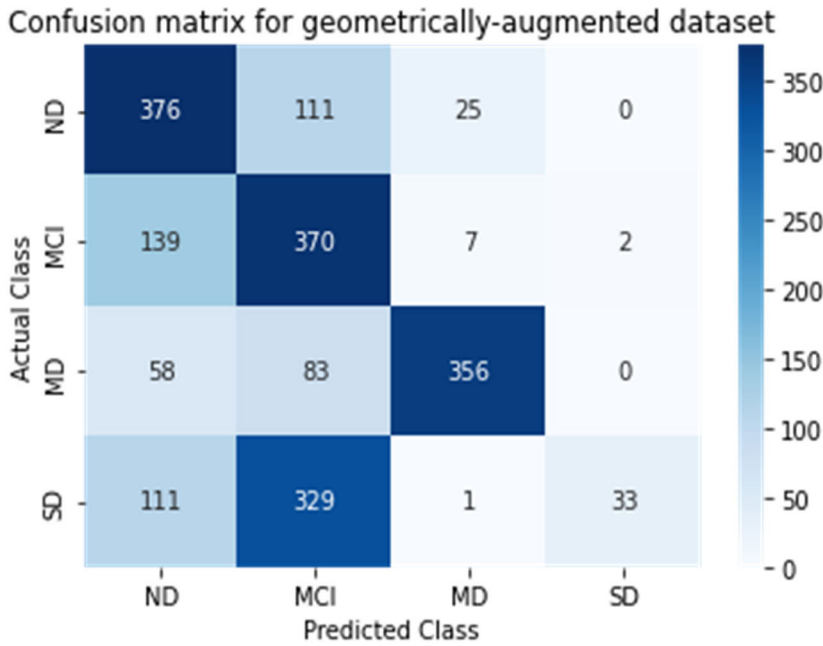


FIGURE 19. Confusion matrix (C.F.) of dataset augmented using geometric transformations.

E. PROGRESSIVE RESIZING (PRO-RESIZING)

Pro-resizing was applied on readily available open-source CNN architectures like ResNet-18, ResNet-34, VGG-16, and VGG-19. These models were trained on the custom dataset proposed in this study, and results with and without pro-resizing were compared. Progressive resizing is a technique used to resize input images sequentially. This includes starting from images of size 32×32 , resizing them to size 64×64 , and finally to 128×128 and giving them as input to different architectures. Each scaled-up model learns its features and incorporates the features learned by the scaled-down model used before. Fast AI was used to apply progressive resizing and to analyze the performance of some

commonly used pre-trained CNN architectures on the custom dataset proposed in this study. VGG-19 performed exceptionally well as compared to other CNN architectures. Various values for batch size, epochs, and resolution were tried, and the best results were considered in this study.

F. PROGRESSIVE-GROWING GAN (PRO-GAN)

Pro-GAN is an extension of the DCGAN model used in the D-BAC system. It has been used as an additional measure to ensure the stability of the generated images and to ensure that they maintain their spatial conformity. Like progressive resizing, Pro-GAN constantly increments the size of the images that are input to it. This is done so that the model generator

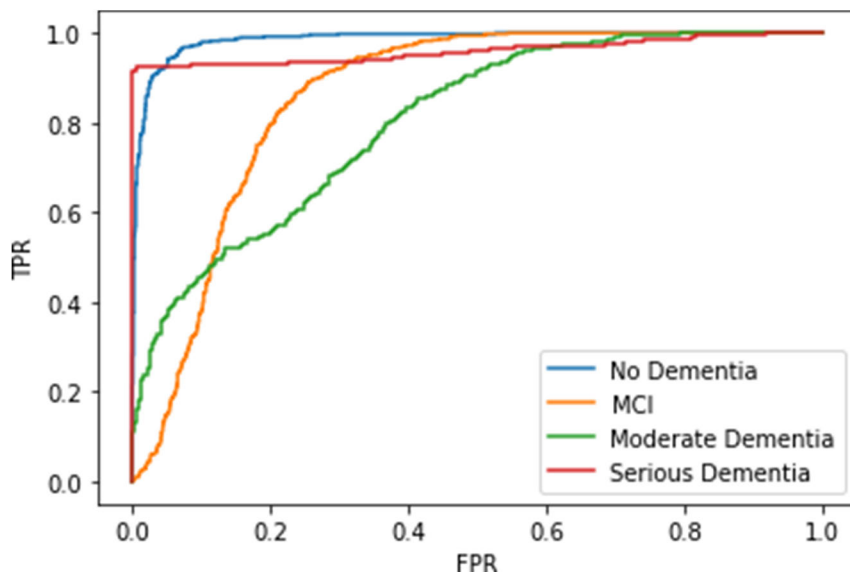


FIGURE 20. Receiver operating characteristics (ROC) curve of D-BAC model.

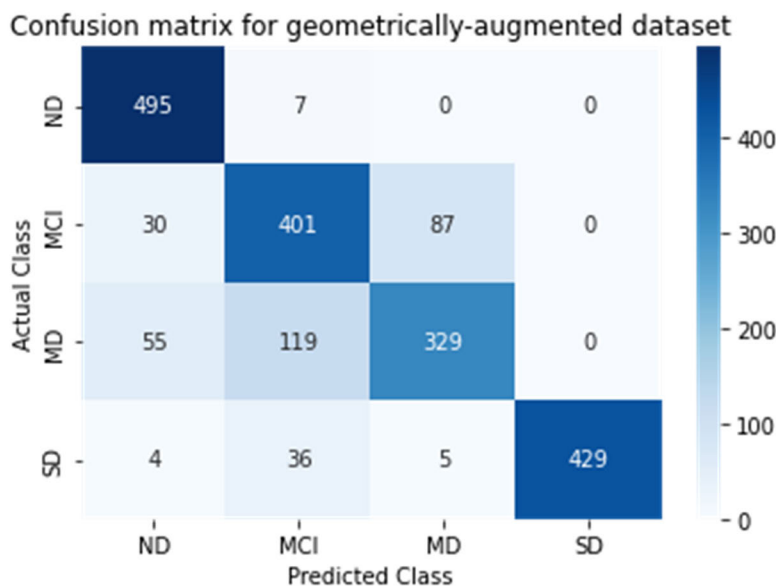


FIGURE 21. Confusion matrix (C.F.) of D-BAC model.

can generate various images of different resolutions until it reaches the desired size and resolution image. This approach has proven to be very effective in this study for generating high-quality synthetic images from original MRI scans.

G. DATASET

The Dementia dataset was acquired from a Kaggle repository [27]. This dataset consisted of 2560 samples for the No Dementia (N.D.) category, 1790 samples for the Mild Cognitive Impairment (MCI) category, 717 samples for Moderate Dementia (M.D.) category, and only 52 samples for Serious Dementia (S.D.) category. Fig. 8 represents this dataset. These categories were highly imbalanced. DCGAN-based augmentation was applied to synthesize new images that were appended to the original dataset after data cleaning. Fig. 9

represents the custom-augmented dataset. This novel Dementia dataset proposed in this research consists of 2500 samples for each class, which helps balance all the categories with the same number of samples and addresses the biasing problem. This novel dataset can also be used for future research in this field.

H. COMPUTATIONAL COMPLEXITY

The model was trained, tested, and evaluated using Google Colab, a cloud-based python IDE developed and maintained by Google. The model training takes about 50 minutes. Testing takes about 15 minutes for the original dataset, 1 hour 30 minutes and 30 minutes for the dataset augmented using geometric transformation, and 1 hour 42 minutes and 38 minutes for the GAN-augmented dataset. Testing takes

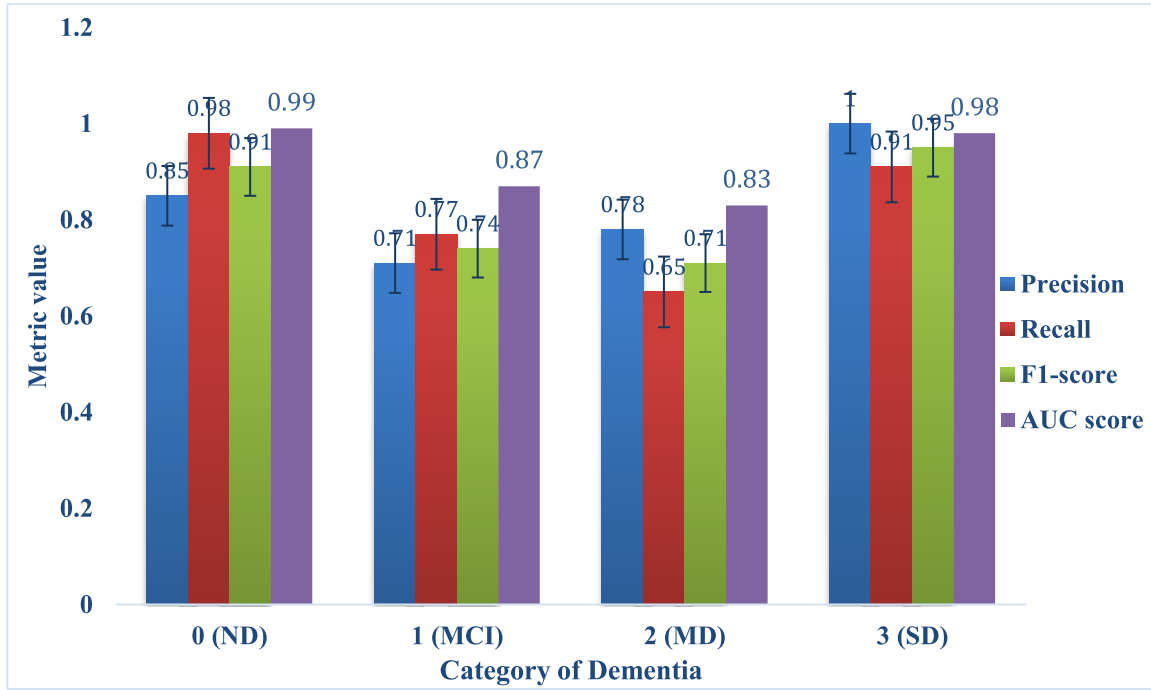


FIGURE 22. Graphical representation of performance metrics for each class of dementia.

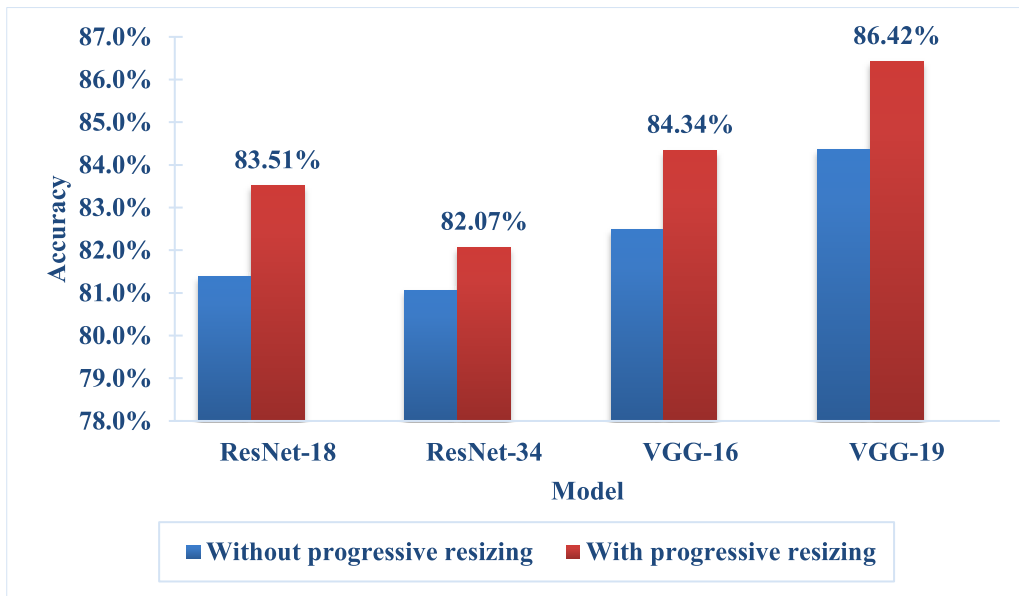


FIGURE 23. Comparison of model accuracies with and without progressive resizing.

15 minutes, 30 minutes, and 38 minutes for the three datasets, respectively. A Dell Inspiron 15 laptop was used to access Colab. Additional hardware and GPUs weren't involved in this research.

IV. SYSTEM ARCHITECTURE

This paper proposes a novel architecture that combines a DCGAN model for image augmentation and a CNN model for classification. Fig. 10 represents this architecture.

The original dataset was acquired from a Kaggle repository [27] and was balanced using DCGAN- based

augmentation. This novel balanced dataset obtained after augmentation was manually analyzed, and mislabelled samples were discarded. Normalization was further applied to the dataset to avoid overfitting, and then this dataset was used for training. This new dataset consisted of 8000 training images and 2000 testing images. This initial stage is termed as the "Data acquisition and augmentation" stage.

The model was trained for 50 epochs using the novel dataset generated above. 50 epochs were chosen because the authors noticed that the model started to overfit the dataset above 50 epochs which resulted in a high accuracy on the

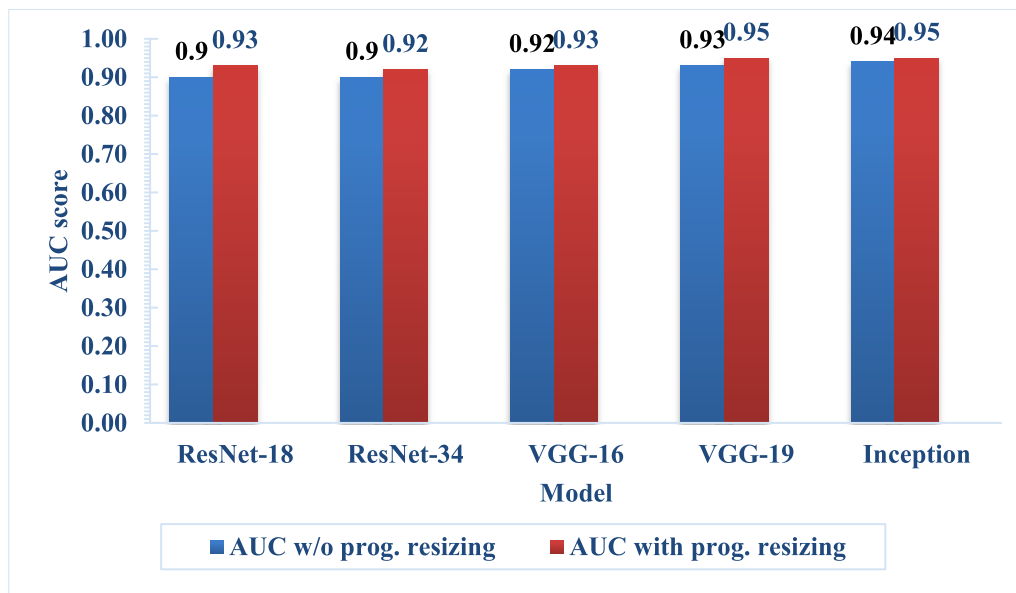


FIGURE 24. Comparison of AUC scores with and without progressive resizing.

training set but a low accuracy on the test set. The model underfit the dataset when trained on epochs lower than 50 which resulted in a low training as well as testing accuracy. A CNN model was used to extract essential features, and the loss/error was computed. The loss was minimized using Adam optimizer. The output of this stage was the final ML model, which could predict early onset of Dementia and classify MRI scans into four different categories depending upon the tumor size – N.D., MCI, MD, and S.D. This stage is termed as the “Training and Optimization” stage.

The third stage involved the Grad-cam-based feature visualization. The weights and features from the last convolutional layer were taken and visualized using Gradient-weighted Class Activation Mapping (Grad-Cam). This stage is termed the “Model Visualization” phase.

The final stage involved testing and making predictions on the test dataset, which consisted of 2000 MRI scans divided equally between the four categories. This is called the “Testing and Prediction” stage. The model could make predictions on this test dataset with the help of the learning done in the training phase.

V. RESULTS AND DISCUSSION

The results for both the augmentation techniques – Simple augmentation and GAN-based augmentation have been compared below. Simple augmentation involves augmenting images by using conventional methods like flipping, scaling, and cropping of images. This technique yields poor results as the spatial conformity of the images are altered during augmentation. The tumor size is crucial in predicting early dementia and determining its severity, but it gets altered during augmentation, leading to poor results. Simple augmentation can be used for binary classification where the classification is not done based on the tumor size, and only the

present is essential. However, it cannot be used for multi-class classification as a multiple-class CNN model differentiates the severity of dementia depending upon the size of the tumor. Therefore, the D-BAC system proposes a GAN-based augmentation as it maintains the spatial conformity of the original images and retains the tumor size in the augmented images. This helps to predict MCI accurately along with the severity of dementia. This method has yielded significantly better results as compared to simple augmentation. The CNN model could achieve 97% accuracy on the training set and SOTA accuracy of 80.5% on the testing set. This is an improvement of over 30% compared to M. Liu’s DSML model [8] and an improvement of over 25% compared to M. Liu’s enhanced DSML model called the DM2L model [8], which are the existing SOTA models in multi-class dementia classification.

Table 4 compares the D-BAC model proposed in this paper with other SOTA models in multi-label Dementia classification. The dataset augmented using DCGAN gives the best results and classification accuracy of 80.5% as the spatial conformity and the normal data distribution of the augmented images remain unaltered. Simple augmentation yields a poor accuracy score as it alters the conformity of the images by scaling, cropping, and resizing the images. The original imbalanced dataset is also compared in Table 4 with other models. It has the lowest accuracy score of 43.5% as the model is biased towards the densely-populated labels and doesn’t generalize well to samples from the sparsely-populated labels. Table 5 and Fig. 12 compare the SOTA binary classifiers in dementia classification.

Various performance metrics like Precision (P), Recall (R), F1-score (F1), Accuracy (A), and Confusion Matrix (C.F.) were used to test and evaluate the performance of the D-BAC system. These metrics compute the relationship between the

TABLE 8. Classification report of performance metrics of the proposed D-BAC model.

Metric	Precision	Recall	F1-score	AUC score	No. of samples
0 (ND)	0.85	0.98	0.91	0.99	512
1 (MCI)	0.71	0.77	0.74	0.87	518
2 (MD)	0.78	0.65	0.71	0.83	497
3 (SD)	1.00	0.91	0.95	0.98	474
Accuracy	-	-	0.82	-	2001
Macro avg.	0.84	0.83	0.83	0.92	2001
Weighted avg.	0.84	0.83	0.83	0.92	2001

True Positives (T.P.), the False Positives (F.P.), the True Negatives (T.N.), and the False Negatives (F.N.) classified by the model. Eq. (1)-(4) represent how these metrics are calculated. The results for all the metrics are tabulated in Table 6.

Precision is the number of false positives classified by the model. Eq. (1) depicts the formula for precision.

$$P = \frac{TP}{TP + FP} \quad (1)$$

The recall is the number of false negatives classified by the model. Eq. (2) depicts the formula for recall.

$$R = \frac{TP}{TP + FN} \quad (2)$$

Eq. (3) depicts the formula for the F1-score.

$$F1 = \frac{2PR}{P + R} \quad (3)$$

Eq. (4) depicts the formula for accuracy.

$$A = \frac{TP + TN}{TP + FP + TN + FN} \quad (4)$$

Table 6 compares the following evaluation metrics –

Precision, Recall, F1-score, and Accuracy for all the 3 datasets studied in this research. The metric values were significantly better for the GAN-augmented dataset than other datasets. The total false positives were low, and the total true positives were high after the CNN classification. The CNN model overfits the imbalanced dataset due to a variation in the sample density for the 4 labels, which amounts to a high number of false positives and the low number of true positives, leading to low-performance metric values. Table 7 represents the confusion matrix (C.F.) used for a 4-class classifier. Furthermore, the ROC curves and the C.F. matrices for the 3 datasets are presented in Fig. 16-21.

The performance metrics, ROC-AUC, and confusion matrix results for each dataset are divided into 3 discrete sections –

- A. Original Dataset,
- B. Simple-augmented dataset and,
- C. GAN-augmented dataset (DCGAN)

A. ORIGINAL DATASET

The ROC curve for the original imbalanced dataset is presented in Fig.16. The ROC curve obtained here is not ideal

as the False Positive Rate (FPR) of the CNN model is very high and True Positive Rate (TPR) is very low. An ideal ROC curve should bend towards the top left corner of the graph, which would signify a high TPR and a low FPR. This doesn't seem to be the case for the original dataset as the different categories of Dementia contain different sample populations, and they are highly imbalanced. The CNN model trained on this dataset starts to overfit the data and fails to generalize well to test data leading to low accuracy and performance scores. The Confusion Matrix (C.F.) for the original dataset is plotted in Fig.17. The diagonal from the left corner represents the total number of correctly classified samples, and the rest of the squares represent misclassified samples or False Positives (F.P.). The False Positives are pretty high for the original dataset, due to which the accuracy score is only 43%. The prediction accuracy of early Dementia or MCI (class 1) for this dataset is only 38%.

1) LIMITATIONS AND THREATS TO VALIDITY FOR AN ORIGINAL DATASET

The obtained results will become invalid if additional MRI scans are injected into the original dataset resulting in a balanced dataset. A balanced dataset would yield significantly better results as well as address the problem of overfitting while training.

B. DATASET AUGMENTED USING GEOMETRIC TRANSFORMATIONS

The ROC curve for the dataset augmented using geometric transformations is presented in Fig.18. Although this curve is better than the original dataset, it fails to accurately predict the early onset of Dementia and its severity. The accuracy score obtained by the CNN model for this dataset is only 57%. The authors think this might be that the feature distribution of the images is altered when augmented like this. This increases the False Positive Rate (FPR) when the model is tested on a test set. The Confusion Matrix (C.F.) for the dataset augmented using geometric transformations is plotted in Fig. 19. The False Positive Rate (FPR) has dropped significantly compared to the C.F. of the original dataset as the dataset is now balanced, and a higher number of image samples are available for training. But this dataset achieves only 57% accuracy, which is not satisfactory or acceptable when working in the medical field to predict early Dementia

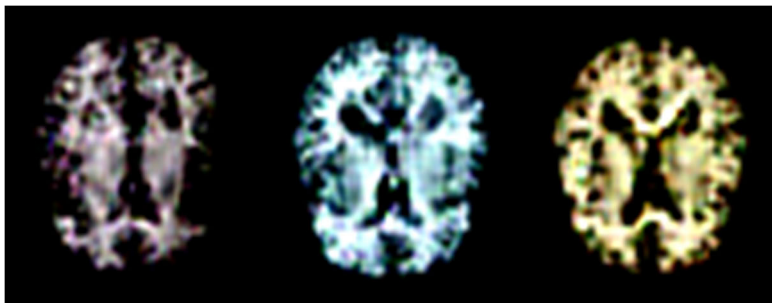


FIGURE 25. Images generated after using Pro-GAN with the original MRI scans.

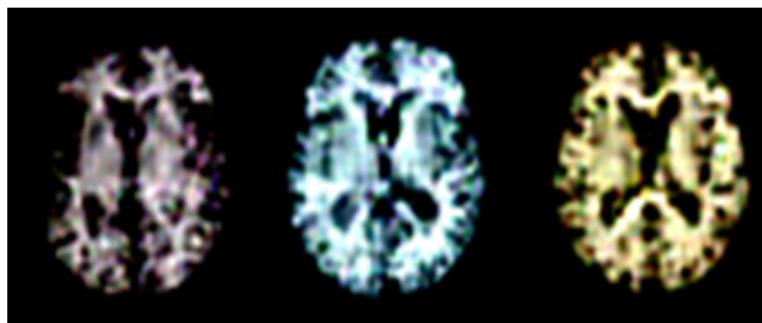


FIGURE 26. Pro-GAN images rotated 180° in the anterior-posterior view.

or its severity. The accuracy score for the prediction of early-onset or MCI for this dataset is only 42%, which is not a significant improvement over the original dataset. Hence, a novel GAN-augmented dataset is proposed to overcome and address this problem and better predict MCI and Dementia severity.

1) LIMITATIONS AND THREATS TO VALIDITY FOR DATASET AUGMENTED USING GEOMETRIC TRANSFORMATIONS

These results would become invalid if one were to take utmost care to prevent geometrically transformed samples while augmenting the dataset using this method or by discarding them if, at all, they were present in the dataset after careful analysis of the dataset in the pre-processing stage. Geometrically transformed samples are mislabeled because the tumor size gets altered while augmenting, and these samples are placed in the incorrect class. The results above would become invalid if other augmentation techniques other than transformations can be implemented, which can prevent geometrically-transformed sample.

C. DCGAN-AUGMENTED DATASET

The ROC curve obtained for the novel D-BAC system is presented in Fig. 20. The model achieved an accuracy of 80.5% on this dataset, and the True Positive Rate (TPR) for this dataset is also high, as can be seen from the curve. The reason is that the spatial conformity of the images remains unaltered when augmented using DCGAN. The CNN model can predict No Dementia (N.D.) and Serious Dementia (S.D.) with nearly 100% accuracy for this dataset. Fig.21 represents the Confusion Matrix (C.F.) for the proposed D-BAC model. The CNN model achieves an accuracy score of 82%,

State-Of-The-Art (SOTA) on a multi-label Dementia prediction. The True Positive Rate (TPR) is significantly better than Fig.17 and Fig.19, contributing to the massive accuracy improvement. The prediction accuracy of MCI (class 1) is 77% for the D-BAC model, which is the highest compared to existing literature and highest amongst the 3 datasets.

1) LIMITATIONS AND THREATS TO VALIDITY FOR DCGAN-AUGMENTED DATASET

The obtained results could become invalid if the DCGAN model used for synthesizing and augmenting images couldn't generalize well to the input dataset and, therefore, produced poor quality images that were very different from the actual images supplied. This could lead to images that have been mislabeled, and such images can lead to a poorly trained model or even an underfitted model. Each image generated by the DCGAN needs to be assessed for its affinity to that particular class.

Furthermore, Table 8 presents the evaluation metric scores for each label using the novel D-BAC model. All the values obtained are SOTA when compared to existing literature and models.

Table 9 tabulates the results before and after progressive resizing was applied to different open-source architectures trained using the novel dataset proposed in this study. Each one of them yielded excellent accuracy scores. VGG-19 is a commonly-used CNN architecture that uses 19 layers for feature learning and classification. It yields an accuracy score of 86.42% after progressive resizing, which is an improvement of nearly 5% compared to the CNN model proposed in this study. This shows that the GAN-augmented dataset is highly scalable and can be used with different CNN architectures,

TABLE 9. The accuracy obtained after applying progressive-resizing on the D-BAC model.

Model	Without Progressive resizing	AUC scores w/o prog. resizing	With Progressive resizing	AUC scores with prog. resizing
ResNet-18	81.39%	0.90	83.51%	0.93
ResNet-34	81.05%	0.90	82.07%	0.92
VGG-16	82.49%	0.92	84.34%	0.93
VGG-19	84.36%	0.93	86.42%	0.95
Inception	85.12%	0.94	86.33%	0.95

TABLE 10. Appendix.

Abbreviation	Full form
ND	No Dementia
MCI	Mild Cognitive Impairment
MD	Moderate Dementia
SD	Serious Dementia
AD	Alzheimer's Disease
CNN	Convolutional Neural Network
DNN	Deep Neural Network
GAN	Generative Adversarial Network
DCGAN	Deep Convolutional GAN
MRI	Magnetic Resonance Imaging
D-BAC	DCGAN-based Augmentation and Classification

not just the one proposed in this study. Further research can be done on this topic to find out which CNN or LSTM model yields the best accuracy on the proposed DCGAN dataset.

The images augmented using Progressive-growing GAN are displayed in Fig. 25. This is another augmentation technique that was applied to augment MRI scans using the original dataset to ensure the stability of the generated images. This shows that other GAN-based augmentation techniques can also generate a new dataset from the original dataset.

VI. CONCLUSION

Since there is a need to explore multiclass classification in dementia along with a sufficient number of MRI images, the paper presented a novel D-BAC system that uses a GAN-based data augmentation technique for training and classification of dementia into various categories depending upon the severity. The paper addressed the critical problem of imbalanced datasets using GAN augmentation to balance the class labels and created a newly balanced dataset. The balancing dataset has been done with the help of two techniques: using geometric transformations, and the second method uses GAN. The GAN-based dataset proved to be superior to geometric transformations as the spatial conformity of the image is unaltered, which helps the CNN model generalize well to unseen test samples. The GAN-augmented dataset achieved an accuracy of 81% using a conventional CNN model and 86% using VGG-19. This is a massive improvement over the existing state-of-the-art models, which achieved only 49% and 52% accuracy. Early-onset of dementia or MCI has been

studied in the paper, and the proposed model could predict MCI with an accuracy of 74%. A novel GradCAM-based approach was also used to analyze the CNN model visually. The “blue shift” presented in this paper displayed the progressive features in different stages of dementia which would help the medical practitioners make decisions. Finally, progressive resizing was applied to the GAN-augmented dataset, and dementia classification was done using ResNet-18, ResNet-36, VGG-16, and VGG-19. Out of these, an accuracy of 83% was obtained using VGG-16, and an accuracy of 86% was obtained using VGG-19.

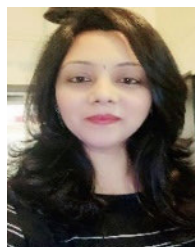
The future scope of this research includes applying other data augmentation techniques to improve the accuracy of the proposed model. Optimization of the total number of features to speed up the computation time. Tuning hyperparameters to find better optimizers and features to improve early dementia prediction

REFERENCES

- [1] K. Davis-Owusu, E. Owusu, L. Marcenaro, C. Regazzoni, L. Feijs, and J. Hu, “Towards a deeper understanding of the behavioral implications of bidirectional activity-based ambient displays in ambient assisted living environments,” in *Enhanced Living Environments* (Lecture Notes in Computer Science), vol. 11369. Cham, Switzerland: Springer, 2019, doi: [10.1007/978-3-030-10752-9_6](https://doi.org/10.1007/978-3-030-10752-9_6).
- [2] J. Gaugler, B. James, T. Johnson, A. Marin, and J. Weuve, “Alzheimer's disease facts and figures,” *Alzheimer's Dementia*, vol. 14, no. 3, pp. 367–429, 2018.
- [3] M. Pudane, S. Petrovica, E. Lavendelis, and H. K. Ekenel, “Towards truly affective AAL systems,” in *Enhanced Living Environments* (Lecture Notes in Computer Science), vol. 11369, I. Ganchev, N. Garcia, C. Dobre, C. Mavromoustakis, and R. Goleva, Eds. Cham, Switzerland: Springer, 2019, doi: [10.1007/978-3-030-10752-9_7](https://doi.org/10.1007/978-3-030-10752-9_7).

- [4] H. Li, M. Habes, D. A. Wolk, and Y. Fan, "Alzheimer's disease neuroimaging initiative," *J. Alzheimer's Dementia*, vol. 15, no. 8, pp. 1059–1070, Aug. 2019.
- [5] S. Basaia, F. Agosta, L. Wagner, E. Canu, G. Magnani, R. Santangelo, and M. Filippi, "Automated classification of Alzheimer's disease and mild cognitive impairment using a single MRI and deep neural networks," *NeuroImage, Clin.*, vol. 21, Jan. 2019, Art. no. 101645, doi: 10.1016/j.nicl.2018.101645.
- [6] M. Amin-Naji, H. Mahdavinataj, and A. Aghagolzadeh, "Alzheimer's disease diagnosis from structural MRI using Siamese convolutional neural network," in *Proc. 4th Int. Conf. Pattern Recognit. Image Anal. (IPRIA)*, Mar. 2019, pp. 75–79, doi: 10.1109/IPRIA.2019.8786031.
- [7] M. Kavitha, N. Yudistira, and T. Kurita, "Multi instance learning via deep CNN for multi-class recognition of Alzheimer's disease," in *Proc. IEEE 11th Int. Workshop Comput. Intell. Appl. (IWCIA)*, Nov. 2019, pp. 89–94, doi: 10.1109/IWCIA47330.2019.8955006.
- [8] F. Li and M. Liu, "Alzheimer's disease diagnosis based on multiple cluster dense convolutional networks," *Computerized Med. Imag. Graph.*, vol. 70, pp. 101–110, Dec. 2018, doi: 10.1016/j.compmedimag.2018.09.009.
- [9] F. J. Martinez-Murcia, A. Ortiz, J.-M. Gorriz, J. Ramirez, and D. Castillo-Barnes, "Studying the manifold structure of Alzheimer's disease: A deep learning approach using convolutional autoencoders," *IEEE J. Biomed. Health Informat.*, vol. 24, no. 1, pp. 17–26, Jan. 2020, doi: 10.1109/JBHI.2019.2914970.
- [10] X. W. Gao, R. Hui, and Z. Tian, "Classification of CT brain images based on deep learning networks," *Comput. Methods Programs Biomed.*, vol. 138, pp. 49–56, Jan. 2017, doi: 10.1016/j.cmpb.2016.10.007.
- [11] 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," in *Proc. IEEE 1st Conf. Multimedia Inf. Process. Retr. (MIPR)*, Apr. 2018, pp. 140–145, doi: 10.1109/MIPR.2018.00032.
- [12] U. Senanayake, A. Sowmya, and L. Dawes, "Deep fusion pipeline for mild cognitive impairment diagnosis," in *Proc. IEEE 15th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2018, pp. 1394–1397, doi: 10.1109/ISBI.2018.8363832.
- [13] D. Jarrett, J. Yoon, and M. van der Schaar, "Dynamic prediction in clinical survival analysis using temporal convolutional networks," *IEEE J. Biomed. Health Informat.*, vol. 24, no. 2, pp. 424–436, Feb. 2020, doi: 10.1109/JBHI.2019.2929264.
- [14] A. Fedorov, R. D. Hjelm, A. Abrol, Z. Fu, Y. Du, S. Plis, and V. D. Calhoun, "Prediction of progression to Alzheimer's disease with deep InfoMax," in *Proc. IEEE EMBS Int. Conf. Biomed. Health Informat. (BHI)*, May 2019, pp. 1–5, doi: 10.1109/BHI.2019.8834630.
- [15] J. M. Valverde, V. Imani, A. Abdollahzadeh, R. De Feo, M. Prakash, R. Ciszek, and J. Tohka, "Transfer learning in magnetic resonance brain imaging: A systematic review," 2021, arXiv:2102.01530.
- [16] M. S. Alam, K.-C. Kwon, M.-U. Erdenebat, M. Y. Abbass, M. A. Alam, and N. Kim, "Super-resolution enhancement method based on generative adversarial network for integral imaging microscopy," *Sensors*, vol. 21, no. 6, p. 2164, Mar. 2021, doi: 10.3390/s21062164.
- [17] D. Lu, K. Popuri, G. W. Ding, R. Balachandar, and M. F. Beg, "Multimodal and multiscale deep neural networks for the early diagnosis of Alzheimer's disease using structural MR and FDG-PET images," *Sci. Rep.*, vol. 8, p. 5697, Apr. 2018, doi: 10.1038/s41598-018-22871-z.
- [18] H. Shin et al., "GANDALF: Generative adversarial networks with discriminator-adaptive loss fine-tuning for Alzheimer's disease diagnosis from MRI," in *Proc. MICCAI*, 2020.
- [19] T. Jo et al., "Deep learning in Alzheimer's disease: Diagnostic classification and prognostic prediction using neuroimaging data," *Frontiers Aging Neurosci.*, 2019.
- [20] Z. Cui, Z. Gao, J. Leng, T. Zhang, P. Quan, and W. Zhao, "Alzheimer's disease diagnosis using enhanced inception network based on brain magnetic resonance image," in *Proc. IEEE Int. Conf. Bioinf. Biomed. (BIBM)*, Nov. 2019, pp. 2324–2330, doi: 10.1109/BIBM47256.2019.8983046.
- [21] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial networks," 2014, arXiv:1406.2661.
- [22] S. Manjunath, A. Nathaniel, J. Druce, and S. German, "Improving the performance of fine-grain image classifiers via generative data augmentation," 2020, arXiv:2008.05381.
- [23] S. Kazeminiya, C. Baur, A. Kuijper, B. van Ginneken, N. Navab, S. Albarqouni, and A. Mukhopadhyay, "GANs for medical image analysis," 2018, arXiv:1809.06222.
- [24] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," 2015, arXiv:1511.06434.
- [25] A. Radford et al., "Unsupervised representation learning with deep convolutional generative adversarial networks," in *Proc. ICLR*, 2016.
- [26] *Reference Dataset*. Accessed: Apr. 9, 2021. [Online]. Available: <https://www.kaggle.com/tourist55/alzheimers-dataset-4-class-of-images>
- [27] N. Srivastava, E. Mansimov, and R. Salakhutdinov, "Unsupervised learning of video representations using LSTMs," 2015, arXiv:1502.04681.
- [28] Maaten, Laurens van der, and Geoffrey Hinton, "Visualizing data using T-SNE," *J. Mach. Learn. Res.*, vol. 9, no. 86, pp. 2579–2605, 2008.
- [29] C. K. Fisher, A. M. Smith, and J. R. Walsh, "Machine learning for comprehensive forecasting of Alzheimer's disease progression," *Sci. Rep.*, vol. 9, no. 1, 2019, Art. no. 13622, doi: 10.1038/s41598-019-49656-2.
- [30] J. Albright, "Forecasting the progression of Alzheimer's disease using neural networks and a novel preprocessing algorithm," *Alzheimer's Dementia: Transl. Res. Clin. Intervent.*, vol. 5, no. 1, pp. 483–491, Jan. 2019, doi: 10.1016/j.trci.2019.07.001.
- [31] G. Lee, K. Nho, B. Kang, K. A. Sohn, and D. Kim, "Predicting Alzheimer's disease progression using multi-modal deep learning approach," *Sci. Rep.*, vol. 9, no. 1, p. 1952, 2019, doi: 10.1038/s41598-018-37769-z.
- [32] A. Abrol, Z. Fu, Y. Du, and V. D. Calhoun, "Multimodal data fusion of deep learning and dynamic functional connectivity features to predict Alzheimer's disease progression," in *Proc. 41st Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2019, pp. 4409–4413, doi: 10.1109/EMBC.2019.8856500.
- [33] X. Hao, Y. Bao, Y. Guo, M. Yu, D. Zhang, S. L. Risacher, A. J. Saykin, X. Yao, and L. Shen, "Multi-modal neuroimaging feature selection with consistent metric constraint for diagnosis of Alzheimer's disease," *Med. Image Anal.*, vol. 60, Feb. 2020, Art. no. 101625, doi: 10.1016/j.media.2019.101625.
- [34] Y. Freund and R. E. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting," *J. Comput. Syst. Sci.*, vol. 55, no. 1, pp. 119–139, Aug. 1997, doi: 10.1006/jcss.1997.1504.
- [35] J. Islam and Y. Zhang, "Deep convolutional neural networks for automated diagnosis of Alzheimer's disease and mild cognitive impairment using 3D brain MRI," in *Brain Informatics*, 2018, pp. 359–369, doi: 10.1007/978-3-030-05587-5_34.
- [36] A. Sabo, S. Mehdizadeh, A. Iaboni, and B. Taati, "Estimating parkinsonism severity in natural gait videos of older adults with dementia," 2021, arXiv:2105.03464.
- [37] J. Ma, *Associations Between Finger Tapping, Gait and Fall Risk With Application to Fall Risk Assessment*. Santa Clara, CA, USA: Hitachi (China) Research & Development Corporation, 2020. [Online]. Available: <https://arxiv.org/abs/2006.16648>
- [38] S. Luz, F. Haider, S. D. L. Fuente, D. Fromm, and B. MacWhinney, *Detecting Cognitive Decline Using Speech: The ADReSSoChallenge*. Shanghai, China: Interspeech, 2021, doi: 10.21437/Interspeech.2021-1220.
- [39] M. Rohanian, J. Hough, and M. Purver, *Multi-Modal Fusion With Gating Using Audio, Lexical and Disfluency Features for Alzheimer's Dementia Recognition From Spontaneous Speech*. Shanghai, China: Interspeech, 2021, doi: 10.21437/Interspeech.2020-2721.
- [40] T. Cohen and S. Pakhomov, *A Tale of Two Perplexities: Sensitivity of Neural Language Models to Lexical Retrieval Deficits in Dementia of the Alzheimer's Type*. Stroudsburg, PA, USA: Association for Computational Linguistics, 2021. [Online]. Available: <https://arxiv.org/abs/2005.03593>
- [41] S. El-Sappagh, J. M. Alonso, S. M. R. Islam, A. M. Sultan, and K. S. Kwak, "A multilayer multimodal detection and prediction model based on explainable artificial intelligence for Alzheimer's disease," *Sci. Rep.*, vol. 11, no. 1, p. 2660, Dec. 2021, doi: 10.1038/s41598-021-82098-3.
- [42] J. Venugopalan, L. Tong, H. R. Hassanzadeh, and M. D. Wang, "Multimodal deep learning models for early detection of Alzheimer's disease stage," *Sci. Rep.*, vol. 11, no. 1, p. 3254, Dec. 2021, doi: 10.1038/s41598-020-74399-w.
- [43] M. Odusami, R. Maskeliūnas, R. Damaševičius, and T. Krilavičius, "Analysis of features of Alzheimer's disease: Detection of early stage from functional brain changes in magnetic resonance images using a finetuned ResNet18 network," *Diagnostics*, vol. 11, no. 6, p. 1071, Jun. 2021, doi: 10.3390/diagnostics11061071.

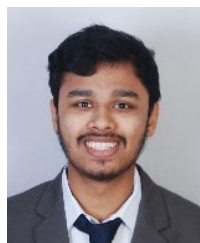
- [44] C. Han, L. Rundo, K. Murao, T. Noguchi, Y. Shimahara, Z. Á. Milacski, S. Koshino, E. Sala, H. Nakayama, and S. Satoh, "MADGAN: Unsupervised medical anomaly detection GAN using multiple adjacent brain MRI slice reconstruction," *BMC Bioinf.*, vol. 22, no. S2, pp. 1–20, Apr. 2021, doi: [10.1186/s12859-020-03936-1](https://doi.org/10.1186/s12859-020-03936-1).
- [45] S. Spasov, L. Passamonti, A. Duggento, P. Lió, and N. Toschi, "A parameter-efficient deep learning approach to predict conversion from mild cognitive impairment to alzheimer's disease," *NeuroImage*, vol. 189, pp. 276–287, Apr. 2019, doi: [10.1016/j.neuroimage.2019.01.031](https://doi.org/10.1016/j.neuroimage.2019.01.031).
- [46] C. Han, L. Rundo, R. Araki, Y. Furukawa, G. Mauri, H. Nakayama, and H. Hayashi, "Infinite brain MR images: PGGAN-based data augmentation for tumor detection," in *Neural Approaches to Dynamics of Signal Exchanges (Smart Innovation, Systems and Technologies)*, vol. 151. Singapore: Springer, 2020, pp. 291–303, doi: [10.1007/978-981-13-8950-4_27](https://doi.org/10.1007/978-981-13-8950-4_27).
- [47] C. K. Chong and E. T. W. Ho, "Synthesis of 3D MRI brain images with shape and texture generative adversarial deep neural networks," *IEEE Access*, vol. 9, pp. 64747–64760, 2021, doi: [10.1109/ACCESS.2021.3075608](https://doi.org/10.1109/ACCESS.2021.3075608).



SHRUTI PATIL received the M.Tech. degree in computer science and the Ph.D. degree in the domain of data privacy from Pune University. She was an Industry Professionalist in the past. She is currently associated as a Professor with the Symbiosis Institute of Technology and a Research Associate with SCAA, Pune, Maharashtra. She has three years of industry experience and ten years of academic experience. She has expertise in applying innovative technology solutions to real world problems. She is working in the application domains of healthcare, sentiment analysis, emotion detection, and machine simulation via which she is also guiding several UG, PG, and Ph.D. students as a Domain Expert. She has published more than 30 research articles in reputed international conferences and Scopus/Web of Science indexed journals and books with more than 100 citations. Her research interests include applied artificial intelligence, natural language processing, acoustic AI, adversarial machine learning, data privacy, digital twin applications, GANS, and multimodal data analysis.



VARUN JAIN is currently pursuing the B.Tech. degree in electronics and telecommunication with the Symbiosis Institute of Technology, Symbiosis International University, Pune. His research interests include artificial intelligence, deep learning, generative adversarial networks, and the IoT.



OM NANKAR is currently pursuing the B.Tech. degree in information technology with the Symbiosis Institute of Technology, Symbiosis International (Deemed University), Pune. His research interests include computer vision, deep learning, machine learning, and general adversarial networks.



DARYL JACOB JERRISH is currently pursuing the B.Tech. degree in computer science with the Symbiosis Institute of Technology, Symbiosis International (Deemed University), Pune. His research interests include deep learning, machine learning, computer vision, general adversarial networks, and data engineering.



SHILPA GITE received the Ph.D. degree in using deep learning for assistive driving in semi-autonomous vehicles from Symbiosis International (Deemed University), Pune, India, in 2019. She is currently working as an Associate Professor with the Computer Science Department, Symbiosis Institute of Technology, Pune. She is also working as an Associate Faculty at the Symbiosis Centre of Applied AI (SCAAI). She has around 13 years of teaching experience. She is guiding Ph.D. students in biomedical imaging, self-driving cars, and natural language processing areas. She has published more than 50 research papers in international journals and 20 Scopus indexed international conferences. Her research interests include deep learning, machine learning, medical imaging, and computer vision. She was a recipient of the Best Paper Award at 11th IEMERA Conference held virtually at Imperial College, London, in October 2020.



KETAN KOTECHA is currently an Administrator and a Teacher of deep learning. He has expertise and experience of cutting-edge research and projects in AI and deep learning for the last 25 (more than) years. He has published more than 100 articles in a number of excellent peer-reviewed journals on various topics ranging from cutting edge AI, education policies, teaching learning practices, and AI for all. He has published three patents and also delivered keynote speeches at various national and international forums, including at the Machine Intelligence Laboratory, USA, IIT Bombay under the World Bank Project, the International Indian Science Festival organized by the Department of Science and Technology, Government of India, and many more. His research interests include artificial intelligence, computer algorithms, machine learning, and deep learning. He was a recipient of the two SPARC projects worth INR 166 lakhs from MHRD Government of India in AI in collaboration with Arizona State University, USA, and The University of Queensland, Australia. He was also a recipient of numerous prestigious awards, like the Erasmus+ Faculty Mobility Grant to Poland, the DUO-India Professors Fellowship for research in responsible AI in collaboration with Brunel University, U.K., the LEAP Grant at Cambridge University, U.K., the UKIERI Grant with Aston University, U.K., and a Grant from the Royal Academy of Engineering, U.K., under the Newton Bhabha Fund. He is an Academic Editor of *PeerJ Computer Science* journal. He is also an Associate editor of IEEE ACCESS journal.