

# MRI-based mild cognitive impairment and Alzheimer's disease classification using an algorithm of combination of variational autoencoder and other machine learning classifiers

Journal of Alzheimer's  
Disease Reports  
Volume 8: 1434–1452  
© The Author(s) 2024  
Article reuse guidelines:  
sagepub.com/journals-permissions  
DOI: 10.1177/25424823241290694  
journals.sagepub.com/home/alr



Subhrangshu Bit<sup>1</sup> , Pritam Dey<sup>1</sup>, Arnab Maji<sup>2</sup> , for the Alzheimer's Disease Neuroimaging Initiative\* and Tapan K Khan<sup>1</sup>

## Abstract

**Background:** Correctly diagnosing mild cognitive impairment (MCI) and Alzheimer's disease (AD) is important for patient selection in drug discovery. Research outcomes on stage diagnosis using neuroimages combined with cerebrospinal fluid and genetic biomarkers are expensive and time-consuming. Only structural magnetic resonance imaging (sMRI) scans from two internationally recognized datasets are employed as input as well as test and independent validation to determine the classification of dementia by the machine learning algorithm.

**Objective:** We extract the reduced dimensional latent feature vector from the sMRI scans using a variational autoencoder (VAE). The objective is to classify AD, MCI, and control (CN) using MRI and without any other information.

**Methods:** The extracted feature vectors from MRI scans by VAE are used as input conditions for different advanced machine-learning classifiers. Classification of AD/CN/MCI are conducted using the output of VAE from MRI images and different artificial intelligence/machine learning classifier models in two cohorts.

**Results:** Using only MRI scans, the primary goal of the study is to test the ability to classify AD from CN and MCI cases. The current study achieved classification accuracies of AD versus CN 75.45% (F1-score = 79.52%), AD versus MCI 81.41% (F1-Score = 87.06%), and autopsy-confirmed AD versus MCI 92.75% (F1-Score = 95.52%) in test sets and AD versus CN 86.16% (F1-score = 92.03%) and AD versus MCI 70.03% (F1-Score = 82.1%) in validation data set.

**Conclusions:** By overcoming the data leakage problem, the autopsy-confirmed machine learning classification model is tested in two independent cohorts. External validation by an independent cohort improved the quality and novelty of the classification algorithm.

## Keywords

Alzheimer's disease, extra tree, light gradient boosting model, machine learning, magnetic resonance imaging, mild cognitive impairment, random forest, support vector machine linear kernel, variational autoencoder, XGB classifier

Received: 20 September 2024; accepted: 20 September 2024

## Introduction

Like other research areas, artificial intelligence (AI), machine learning (ML), and deep learning (DL) in image analysis are growing trends in biomedical research. The last decades of the exponential growth of computer power and theoretical breakthroughs in extensive data handling power have accelerated the development of the research area of biomedical imaging. Artificial neural networks and deep-rooted computer vision applications enable DL to support solving the shortcomings of biomedical imaging techniques such as magnetic resonance imaging (MRI), electroencephalography,

<sup>1</sup>Biolmaginix LLC, Morgantown, WV, USA

<sup>2</sup>Department of Chemistry, Indian Institute of Technology, Kanpur, UP, India

\*Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in the analysis or writing of this report. A complete listing of ADNI investigators can be found at: [http://adni.loni.usc.edu/wp-content/uploads/how\\_to\\_apply/ADNI\\_Acknowledgement\\_List.pdf](http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf).

## Corresponding author:

Tapan K Khan, Biolmaginix LLC, 410 Mallard Run, Morgantown, WV 26508, USA.  
Email: [khantapan67@gmail.com](mailto:khantapan67@gmail.com)



magnetoencephalography, positron emission tomography (PET), etc. One of the significant advantages of ML is the ability to predict where other standard statistical models fail. In addition, DL methods have the capabilities of quantitative image analysis, region-specific segmentation, and determining the acceptable structural changes concerning subtle changes in pathophysiology in the brain before widespread atrophy and irreversible brain damage.<sup>1–3</sup>

Approximately 50 million people live with Alzheimer's disease (AD) worldwide, and estimates show that this number will reach 152 million in 2050. AD, the most common form of dementia consisting of 60–80%, is an irreversible, slow, but progressive neurodegenerative condition that gives rise to structural and functional changes in the brain, with often no sign of permanent recovery. The disease is diagnosed from a history of cognitive decline and clinical cognitive testing, such as the Mini-Mental State Examination (MMSE) and the Mini-Cog test, and exclusion criteria, which indicate an impairment in cognitive abilities rather than the structural changes in the brain that cause it. The most common early symptom of AD is recent memory loss, as the disease impacts brain parts associated with learning. The National Institute on Aging and Alzheimer's Association 2011 (NIA-AA 2011) introduced the concept of preclinical AD, which arises before mild cognitive impairment (MCI) and progresses to advanced stages of AD.<sup>4</sup> MCI is defined as the prodromal stage of cognitive decline that is significantly higher than expected due to normal brain aging. The time between MCI state to AD spans many years, but every MCI patient does not convert to AD.<sup>5</sup> A meta-analysis found that the annual conversion rate from MCI to dementia, AD, and vascular dementia was 9.6%, 8.1%, and 1.9%, respectively, in specialist clinical settings and 4.9%, 6.8%, and 1.6% in community studies.<sup>5</sup> Even after ten years of follow-up, many MCI cases do not progress to the dementia stage. However, there is increasing evidence that neuropathological changes begin decades before the manifestation of the symptoms.<sup>6</sup>

Accurate detection of MCI and its conversion to AD with minimal invasion, like brain imaging, is one of the critical factors for future therapeutic intervention. A machine learning brain imaging-based systemic review on AD versus MCI estimates an accuracy of 75.4% by support vector machine (SVM), and a slightly better result was achieved (78.5%) by convolutional neural networks (CNN).<sup>7</sup> The study found that a combination of MRI and PET achieved overall better classification accuracy than studies that only used one neuroimaging technique. However, in practice, more expensive PET might be out of reach for most insurance companies and not readily available geographically. A combination of CNN and ensemble learning on the MRI approach achieved a better result.<sup>7</sup> The binary classification produced an accuracy of 84% for AD versus healthy control (CN), 79% for MCI-converter versus CN, and 62% for MCI-converter

versus MCI-nonconverter.<sup>8</sup> However, none were verified by a second cohort or specifically cared to reduce biases due to data leakage in AI/ML algorithms.

## Medical context

AD therapeutic trials have failed in recent years because the disease-modifying trials were initiated at an advanced stage of the disease when irreversible neuronal damage had already taken place.<sup>9</sup> AD detection at the MCI stage may provide a crucial window of opportunity to intervene with disease-modifying therapy. Structural MRI (sMRI) provides helpful information about the atrophy of the brain regions by aging and AD. It is still the most used neuroimaging biomarker for the differentiation of dementia in normal aging, MCI, AD, and non-AD dementias. An affected brain is a common feature of familial and sporadic AD.<sup>10</sup> High-resolution 3-D sMRI scans can determine the atrophy of critical brain areas such as the parahippocampal gyrus, hippocampus, amygdala, posterior association cortex, and subcortical region.<sup>11–13</sup> However, the visual rating of hippocampal atrophy and manual volumetric analysis of the hippocampus, several techniques for quantitative assessment of brain volume and atrophy in MRI images, such as automated whole-brain volumetry, quantitative region of interest-based volumetry, the quantitative voxel-based technique, tensor-based morphometric technique, and global atrophy quantification technique are inadequate to estimate the conversion of different stages of AD-related dementia. PET and cerebrospinal fluid (CSF) biomarkers could be better predictors of AD path progression. However, MRI has the potential to identify cortical regions of abnormality known to be affected in AD nearly a decade before clinical symptoms of dementia emerges, providing necessary preclinical evidence of neurodegeneration. Moreover, a recent interest in measuring Amyloid-related imaging abnormalities by specific drug treatments can be detected in brain MRI.<sup>14</sup> Multiple ML models (Logistic regression, XGBoost, and Random Forest) using only Electronic Health Records from the Stanford Health Care data (1999–2022) predicted MCI to AD conversion only <65% accurately<sup>15</sup> and re-enforced the superiority of brain imaging data for diagnosing and predicting MCI to AD conversion.

## Contributions

Other existing approaches of MRI image analysis require reducing variations by intricate multi-step, often manual preprocessing pipelines, such as segmentation, cortical reconstruction, and outlining region-of-interest. The potential of variational autoencoder (VAE) as a promising alternative to predict early AD progression was first

demonstrated for distinguishing healthy/diseased (diseased class includes both MCI and AD patients) states.<sup>16</sup> A second study used a convolutional adversarial autoencoder.<sup>17</sup> Both studies reported comparable classification accuracies. However, neither study tested for gold-standard autopsy-confirmed cases or validated by a separate cohort. VAE-based data augmentation helps resolve high dimension, low sample size problems in healthcare where practitioners must deal most of the time with (very) low sample sizes along with very high dimensional data (e.g., 3-D neuroimaging data). VAE-based data augmentation to address class imbalance has recently been used.<sup>18</sup> A conditional VAE algorithm has been reported to identify brain dysfunction in AD.<sup>19,20</sup> In a non-brain imaging segment, VAE has been employed to distinguish dementia and non-dementia using a clock drawing test.<sup>21</sup> VAE has been used for AD detection in a variety of cases, including genome-wide single nucleotide polymorphisms,<sup>22</sup> tau Flortaurin-PET,<sup>23</sup> and in the identification of white matter macro and microstructural abnormalities.<sup>24</sup> Autoencoder systems have been used to diagnose other psychiatric disorders.<sup>25–27</sup> The model we propose comprises two sections: the first part is VAE, which does feature extraction from MRI scans. The extracted latent representation is fed into a discriminative deep-learning machine algorithm. The output gives different structural features that enable the diagnosis of CN/MCI/AD. In this study, we used only 3D sMRI scans from two internationally recognized datasets [Alzheimer's Disease Neuroimaging Initiative (ADNI) and Open Access Series of Imaging Studies (OASIS)] as input and determined the current stage of dementia (CN/MCI/AD). It is a methodology that leverages the power of VAEs to extract latent distributions and input that to specific machine learning classification to enable us to determine the optimal model parameters for the classification of AD versus CN, AD versus MCI, and MCI versus CN. More specific contributions are: 1) The VAE model outputs latent representations of subject MRIs which are fed into ML classifiers to classify AD/MCI/CN; 2) It is the first successful validated attempt to use VAE in MRI brain imaging processing to classify and validate by a separate cohort. The outcomes of other studies that used VAE were not validated by an independent cohort; and 3) The algorithm we developed is validated with National Institutes of Health (NIH) 'Gold Standard' autopsy-confirmed AD cases and can differentiate AD from MCI patients.

### ***Related work: sMRI and ML for the study of AD diagnosis***

ML and DL techniques, such as SVM, artificial neural network, CNN, etc., are very useful for diagnosing AD.<sup>28,29</sup> The accuracy of classification, however, depends on the type of problem (CN versus AD/CN versus MCI/

MCI versus AD).<sup>30</sup> According to several reports, the best classifiers can discriminate between CN and AD subjects with accuracies in the ~90% range but have considerably lower accuracies when discriminating between control and MCI subjects. However, some of the studies need to be reproducible due to the lack of defined published frameworks and implementation details not reported.<sup>31</sup> Lastly, some of these papers may report a biased performance due to inadequate or unclear validation or model selection procedures. Some classification studies must be validated with multiple data sets.<sup>7</sup> Region-specific information is required for several classification studies.<sup>32</sup> The best classifiers combine optimum features from different modalities, including CSF biomarkers, MRI, fluorodeoxyglucose (FDG)-PET, cognitive measures, and factors such as age and *APOE4* allele status.<sup>7</sup> Implementing such models in clinics and other applicability will be expensive because PET imaging is costly and not readily available, and CSF sample collection is very invasive. The objective is to classify AD, MCI, and CN using MRI and without any other information (e.g., age >60, sex, educational background, no psychometric test) in heterogeneous populations (e.g., a combination of Hispanic, Black, and Caucasian ethnicity). Several AI/ML classification studies have been published where only MRI scans were used.<sup>33–41</sup> Those studies have limitations like not taking care of data leakages, not being validated with secondary data sets, and AD versus MCI classification accuracy for most of them was below 80%. Moreover, most of them were not tested in autopsy-validated sample sets. Only one such kind of study, autopsy-validated, needed to be more accurate to classify AD versus MCI cases.<sup>42</sup>

A significant challenge of using MRI scans is high-dimensional images. To address this, in this study, we extract the reduced dimensional latent feature vector from the sMRI scans using a VAE. The objective is to classify AD, MCI, and CN using MRI and without any other information (e.g., age >60, sex, educational background, no psychometric test) in heterogeneous populations (e.g., a combination of Hispanic, Black, and Caucasian ethnicity). The extracted feature vectors from MRI scans are used as input conditions for different advanced AI/ML classifiers. Classification of AD versus CN, AD versus MCI, and autopsy-AD versus MCI are conducted using the output of VAE from MRI images and different AI/ML classifier models in two different cohorts. Data leakage is an invariable problem in AI/ML image classification. To overcome it, we split ADNI data into training and test sets at the very beginning and later used it to test the algorithm and validated the algorithm again in a different independent cohort (OASIS). Using only MRI scans, the primary goal of the study is to test the ability to classify AD from CN and MCI cases. The current ML classification is tested with patients of NIH 'Gold Standard' autopsy-validated, and it is the first successful attempt to classify AD versus

CN/MCI patients using a combined algorithm consisting of VAE and other ML classifiers. External validation is required to improve the quality and novelty of the classification algorithm. The ML classification model is also tested in an independent cohort (OASIS).

## Methods

### Study population

All neuroimaging, clinical, and autopsy diagnostic information used in this study were in a deidentified format and obtained upon external request. Sources ensured compliance with ethical issues. Therefore, this study was exempted from local institutional review (IRB).<sup>43</sup> This work used state-of-the-art data from the ADNI (Table 1) and OASIS study population (Table 2). The data set was properly labeled collections of MRI images. Each data set has three groups of patients (CN, MCI, and AD) with no difference in age ( $p$ -value  $\geq 0.05$ ). For the study, we use a standard database released by ADNI named the TADPOLE Challenge. We used all the MRI scans and their corresponding subjects from subsets D1 and D2 for the current study. We follow the exact instructions mentioned in the challenge and download the standard data sets from the LONI. After logging in, we go to Download → Study Data → Test Data → Data for Challenges and download ‘Tadpole Challenge Data’. However, since this is study data, it does not include the MRI scans, so we leverage the scripts available in the challenge to find all MRI scan IDs and download them separately.

The ADNI ([adni.loni.usc.edu](http://adni.loni.usc.edu)) was launched in 2003 as a public-private partnership, led by Principal Investigator Michael W. Weiner, MD. The original goal of ADNI was to test whether serial MRI, PET,

other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of MCI and early AD. The current goals include validating biomarkers for clinical trials, improving the generalizability of ADNI data by increasing diversity in the participant cohort, and to provide data concerning the diagnosis and progression of Alzheimer’s disease to the scientific community. For up-to-date information, see [adni.loni.usc.edu](http://adni.loni.usc.edu).

We further investigated the performance of our model on autopsy-confirmed patients’ MRI scans (ADNI data set). The inclusion of patients in this cohort is based on the values of NP\_THAL and NP\_BRAAK being greater than 2 and III, respectively. As an external validation cohort, we used the OASIS-1 data set.<sup>44</sup> We downloaded the MRI scans of all 416 subjects and classified them based on Clinical dementia rating (CDR) scores, as given below. Since all the subjects do not have valid CDR scores, being ‘NA’, we removed them from our test dataset. CDR-Value: 0.0 corresponds to CN, 0.5 corresponds to MCI, 1.0 and 2.0 correspond to AD.

To avoid data leakage during training and validation, the ADNI data was split into training and test data sets. Table 1A data was used for training, and Table 1B was used for the testing. For further validation, we tested the classification models using Table 2 (data from OASIS-1). The number of images from the TADPOLE dataset used in the model training is CN = 1895, MCI = 1880, and AD = 898 (including autopsy-confirmed: 44). There is a slight imbalance by the reduced number of AD images. However, such a slight imbalance will not affect results.

### ADNI data

The ADNI dataset contains several collections of MRI images.<sup>43</sup> In order to benchmark the current algorithm, we selected a standardized subset of data created by ADNI in the form of a challenge TADPOLE and used all the modalities of MRI images for all the subjects included in the challenge. Clinical diagnosis of CN, MCI and AD are made as follows.

*Diagnostic criteria of ADNI TADPOLE Data* (<https://tadpole.grand-challenge.org/Data/>): **Neuropsychological**

**Table 2.** Open Access Series of Imaging Studies (OASIS) study population.

Clinical Dementia Rating (CDR)		
Diagnosis	Number of subjects	Number of MRI scans
CN	135	526
MCI	70	270
AD	30	117

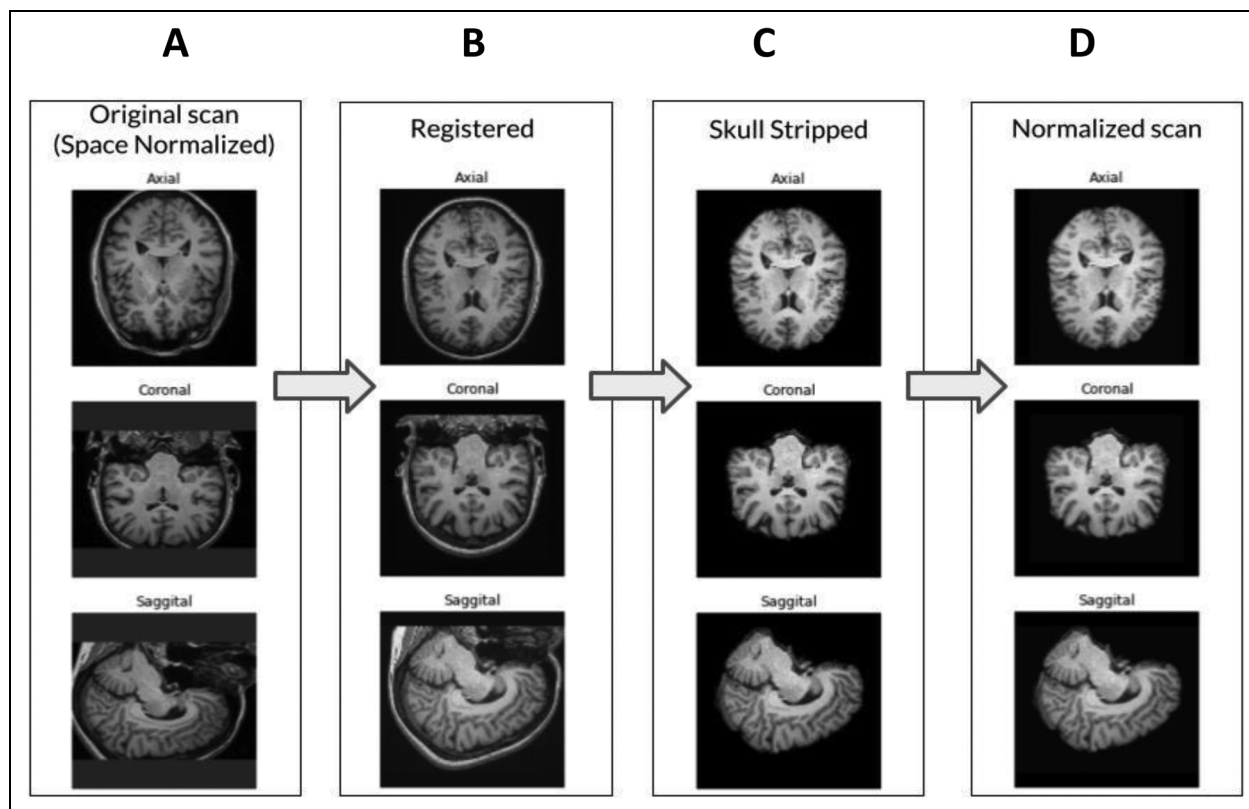
CN: CDR = 0.0, MCI: CDR = 0.5; AD: CDR = 1 and 2.

AD, Alzheimer’s disease; MCI, mild cognitive impairment; CN, age-matched control.

**Table 1.** Data source: Alzheimer’s Disease Neuroimaging Initiative (ADNI).

A. Training data set		
Diagnosis	Number of subjects	Number of MRI scans
CN	380	1895
MCI	361	1880
AD	280	851
B. Testing data set		
Diagnosis	Number of subjects	Number of MRI scans
CN	33	155
MCI	42	233
AD	41	122
Autopsy-confirmed AD	14	44

AD, Alzheimer’s disease; MCI, mild cognitive impairment; CN, age-matched control.



**Figure 1.** Illustration of an example of a preprocessing pipeline on typical sMRI scans. (A) Space normalization ensures that the spatial structure of all the images of the dataset is as similar as possible with each scan resampled to a  $(1 \times 1 \times 1) \text{ mm}^3$  of voxel space. (B) Registration adapts the sMRI scan to another reference image, which is called an atlas (we use T1 weighted MNI152), seeking that the same regions of both represent the same anatomical structure. (C) Skull stripping removes the information from the skull that appears on structural MRI images. (D) Min-max normalization reduces the pixel values varying between 0-1 to maintain similar pixel variation among the sMRI scans.

**tests:** CDR, Sum of Boxes, Alzheimer's Disease Assessment Scale (ADAS)11, ADAS13, MMSE, Rey Auditory Verbal Learning Test (RAVLT), Montreal Cognitive Assessment (MoCA), Everyday Cognition (ECog).

**MRI ROIs (Freesurfer):** Measures of brain structural integrity, volumes, cortical thicknesses surface areas.

**FDG PET:** Region-of-interest (ROI) averages: Measures of cell metabolism, AV45 PET ROI averages: measures amyloid-beta load in the brain, AV1451 PET ROI averages – measures tau load in the brain, diffusion tensor imaging (DTI)-ROI measures microstructural parameters related to cells and axons (cell radial diffusivity, axonal diffusivity, mean diffusivity, axial diffusivity, radial diffusivity, etc.).

**CSF biomarkers:** Amyloid and tau levels.

**APOE status:** APOE4 levels

**Demographic information:** age, gender, education, etc.

**Autopsy-confirmed AD:** Antemortem clinical dementia and two main neuropathologic abnormalities of amyloid plaques and the neurofibrillary tangles in the brain at post-mortem autopsy plus the above criteria of AD.

## OASIS data

The Open Access Series of Imaging Studies (OASIS) is a publicly available series of neuroimaging data sets widely recognized and readily available for study and analysis.<sup>44</sup> It is a project specifically aimed at making neuroimaging data sets of the brain freely available to the scientific community without any obligation. The OASIS data set used only CDR to diagnose AD/MCI/CN clinically. Therefore, there are considerable risks of using only the CDR score for diagnosis as it has been performed in the OASIS compared to the ADNI TADPOLE data set, where numerous factors have been used for the clinical diagnosis of AD/MCI/CN. CDR is an estimation of the clinical judgment of the clinicians that consists of six cognitive and behavioral domains (memory, judgment and problem-solving, orientation, community affairs, home and hobbies performance, and personal care). There are several limitations of CDR cores to use in the clinical diagnosis of AD/MCI/CN. A CDR score of 0.5 may not be accurate for diagnosing MCI cases. A multi-center trial observed that interrater reliability of CDR scores performed moderately to high nut

found limitations in detecting early dementia.<sup>45</sup> Therefore, there is still a risk of using the CDR score for diagnosis that has been used in the OASIS dataset.

### *Description of the distribution of the demographics in the different groups*

A detailed description of demographics information comprising of sex, race & ethnicity and education levels have been described in the Supplemental File 3. There is no bias in sex, race & ethnicity and education levels between training and testing groups.

### *Data inclusion criteria*

The data set of patients considered for the study is a subset of the data set provided under ‘The Alzheimer’s Disease Prediction Of Longitudinal Evolution (TADPOLE) Challenge’. The subset of the patients belongs under the TADPOLE-provided ‘standard’ datasets D1, D2, and D3. However, in this study, our analysis is based only on the structural MRI scans with the ADNI-provided labels CN, MCI, and AD. The reason for using MCI labels is to create an adversary that makes the classification task difficult for AD patients. The MRI scans were filtered to match to be T1 scans.

When multiple scans per subject were available, they were added to increase the training ability. However, classification of train and test subsets were performed at the subject level. Both ADNI and OASIS have multiple scans per subject. We used ADNI for training the VAE model. OASIS data was used only for validation.

### *MRI preprocessing and harmonization*

The images from ADNI have considerable variability within the database because MRI scans were obtained from a multi-centered project. The database had been preprocessed beforehand with multiple techniques. Therefore, images need to be resampled to a joint isotropic resolution and registered to a standard atlas. Image registration consists of adapting a specific image to another reference image, which is called an atlas, seeking that the same regions of both represent the same anatomical structures.<sup>46</sup> This study took the T1 weighted MNI152 template as a reference image. A schematic presentation of MRI image processing steps is illustrated in Figure 1. Before image registration, we performed the spatial normalization preprocessing step to ensure that the spatial structure of all the dataset’s images was as similar as possible (Figure 1). Thus, in the first step, images and the atlas were resampled to an isotropic resolution of 2 mm<sup>3</sup>. For example, in a preprocessing pipeline on typical sMRI scans, the space normalization ensured that the spatial structure of all the images of the

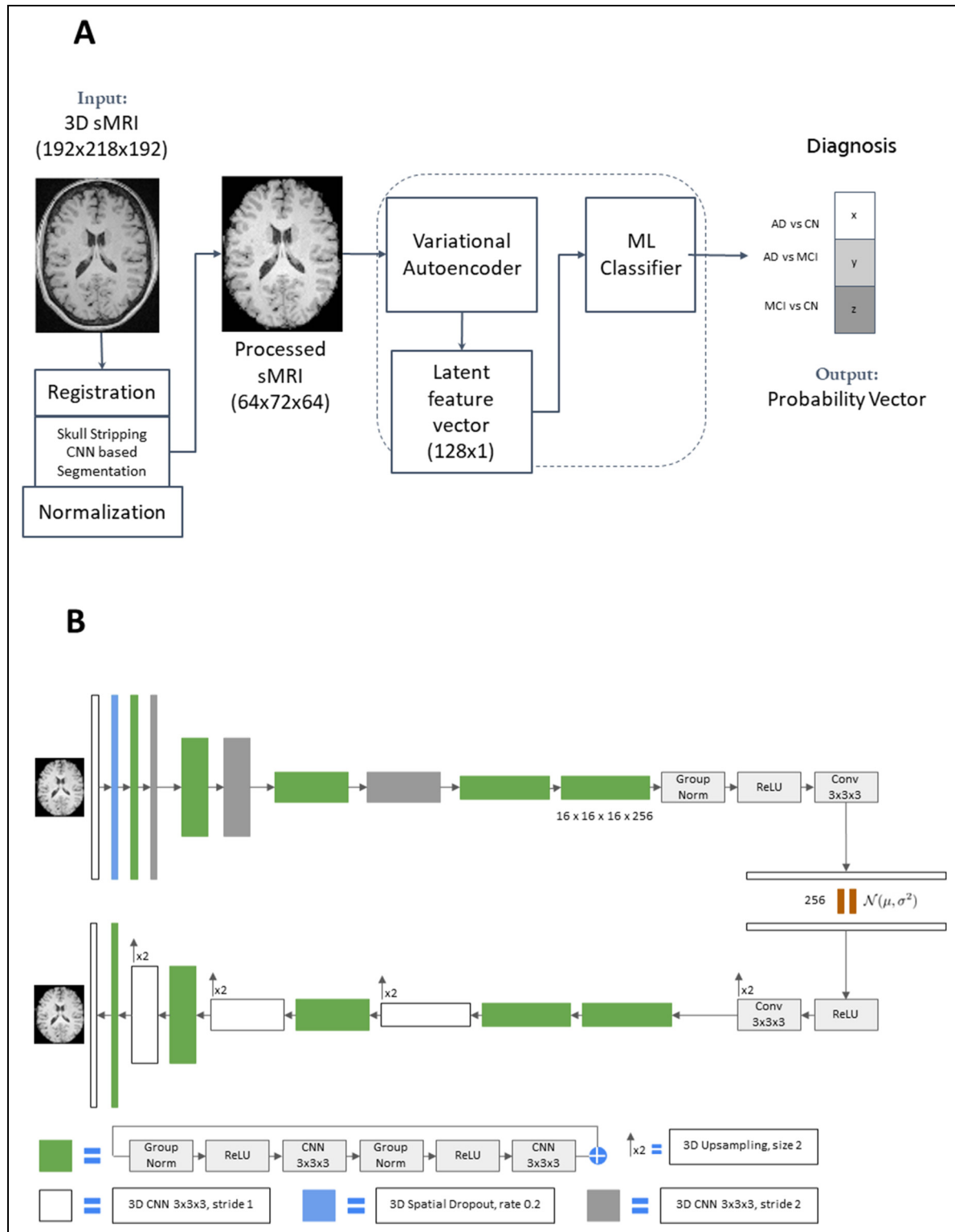
dataset was as similar as possible, with each scan resampled to a (1 × 1 × 1) mm<sup>3</sup> of voxel space (Figure 1A). Registration adapted the sMRI scan to another reference image atlas in T1 weighted MNI152, seeking that the same regions of both represent the same anatomical structure (Figure 1B). Skull stripping removed the information from the skull that appears on structural MRI images (Figure 1C). Min-max normalization reduced the pixel values varying between 0–1 to maintain similar pixel variation among the sMRI scans (Figure 1D). The primary goal is to obtain an image containing only the information relevant to the task. It is established that none of the most relevant biomarkers are found in the skull in the case of AD. Thus, different techniques are used to extract the skull and other non-brain regions<sup>47,48</sup> or directly use an image data set with the skull already stripped.<sup>49</sup> Here, we have stripped the skull part from images of the dataset using the Brain Extraction Tool from the FSL package [<https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/>].<sup>50</sup> Since not all MRI data is obtained and preprocessed in the same way before being stored in the database, finding the ideal fractional intensity threshold that measures the aggressiveness of the algorithm in removing image components that do not represent the brain tissue would not be feasible. So, after multiple tests, a threshold of 0.2 was used since it could keep a correct balance for most images without being perfect for all. We perform intensity normalization beyond these two procedures specific to medical images (Figure 1). In this study, we used the min-max normalization technique to reduce the pixel values between 0 and 1. All those operations helped the auto-encoder identify the regions of interest (Figure 2).

### *Model design*

To find the characteristics of the disease state and capture the probabilities, we approach the problem from a discriminative perspective. As shown in Figure 2, the model is comprised of two sections, (A) a schematic representation of the flowchart of the study, (B) the detailed description of the VAE, which simultaneously extracts a latent representation and reduces the dimension of a 3D sMRI scan. The extracted latent representation is then inputted to a different machine-learning classifier. Given an initial representation of an MRI scan, the final output constitutes the probabilities of possible diagnosis.

### *Feature extraction from MRI scans using variational autoencoder*

Extracting relevant features from the vast amount of complicated information in MRI scans is an important step.<sup>50</sup> We approached the problem from a discriminative perspective of modeling, which is one type of unsupervised learning that deals with complicated data distributions. It could be interpreted as learning a discriminating process by



**Figure 2.** A schematic illustration of the flow diagram illustrating a process for diagnosing Alzheimer's disease (AD) versus mild cognitive impairment (MCI), AD versus non-demented control (CN), and MCI versus CN. (A) MRI 3-dimensional scans are registered as input, followed by skull stripped, convolutional neural network (CNN)-based segmentation, and normalized. Inside the dotted box, in the first step, processed MRI scans are used as input in the Variational Autoencoder (VAE) to produce the output of latent feature vectors. In the second step, the latent feature vectors are used in different machine learning (ML) models to classify AD versus MCI, AD versus CN, and MCI versus CN using MRI scans. (B) Illustration of VAE image processing steps consisting of an encoder and a decoder in a latent Gaussian model where all are parametrized by a discriminating CNN with the computer architecture of a variational autoencoder and decoder utilized. ReLU: rectified linear unit.

which the observation data arose.<sup>51</sup> Feature extraction from MRI scans using VAE for this study has been presented in Figure 2B. Mathematical derivation of the use of VAE for MRI image processing consisting of an encoder and a decoder in a latent Gaussian model where all are parametrized by a discriminative CNN has been derived in the Supplemental File 1.

**Encoder part.** The encoder part uses ResNet blocks, where each block consists of two convolutions with normalization and ReLU, followed by additive identity skip connection.<sup>52</sup> For normalization, we use Group Normalization,<sup>53</sup> which shows better than BatchNorm performance when the batch size is small (batch size is 1 in our case). We followed a common CNN approach to progressively downsize image dimensions by 2 and increase feature size by 2. For downsizing, we use stride convolutions. All convolutions are  $3 \times 3 \times 3$ , with the initial number of filters equal to 32. The encoder endpoint has size 512 (256 for mean and 256 for variance) and is 4096 times spatially smaller than the input image.

**Decoder part.** Starting from the encoder endpoint output, we first reduce the input to a low dimensional space of 256 (128 to represent mean and 128 to represent standard deviation). Then, a sample is drawn from the Gaussian distribution  $N(\mu, \Sigma)$ . The sample is fed as input to the decoder structure like the encoder. Each decoder level begins with upsizing: reducing the number of features by a factor of 2 (using  $1 \times 1 \times 1$  convolutions) and doubling the spatial dimension (using 3D bilinear sampling). The decoder's end has the same spatial size as the original image.

## Computational methods

The code for the framework is developed using Python (version 3.10.x). The VAE framework uses the PyTorch library, and the machine learning classifiers are utilized from PyCaret. We make use of the PyCaret library for cross-validation and fine-tuning the classifiers. For preprocessing the MRI scans, we use the FSL library and its BET tool, while the normalization is done in Python. The data from ADNI has been collected in NIfTI format and read and analyzed in Python using the nibabel library and its dependencies. Furthermore, we use the sci-kit-learn library to generate t-SNE embeddings and matplotlib to visualize the same. The computer codes are available to interested readers upon request. Multiple ML models were explored using PyCaret on the embeddings generated by the VAE model. In this study, we report the best-performing models. To tune the model, a 3-fold cross-validation on the image embeddings from the training subjects was employed to select the optimal hyperparameters. The details of hyperparameter spaces, which were used to find the optimal hyperparameters as well as the finally chosen hyperparameters, are incorporated in the Supplemental File 4. The Bayesian search algorithm from the scikit-optimize library was used

for this purpose. ADNI data from Table 1B and OASIS data from Table 2 were used for validation purposes. ML classifiers used in this study are ExtraTree, Light Gradient Boosting Machine (LGBM), SVM-Linear Kernel, eXtreme Gradient Boosting (XGB), and Random Forest.

## Computer specification

The training, testing, and code development of the model were run on a computer with an Intel Core i7-12700KF processor running at 3.6 gigahertz frequency using 16 gigabytes of DDR4 SDRAM with a speed of 3000 megahertz. The system was equipped with a dedicated 8 gigabytes of NVIDIA GeForce RTX-3070 GPU video memory (RAM), running Ubuntu (a Linux distribution) version 22.04.

## Results

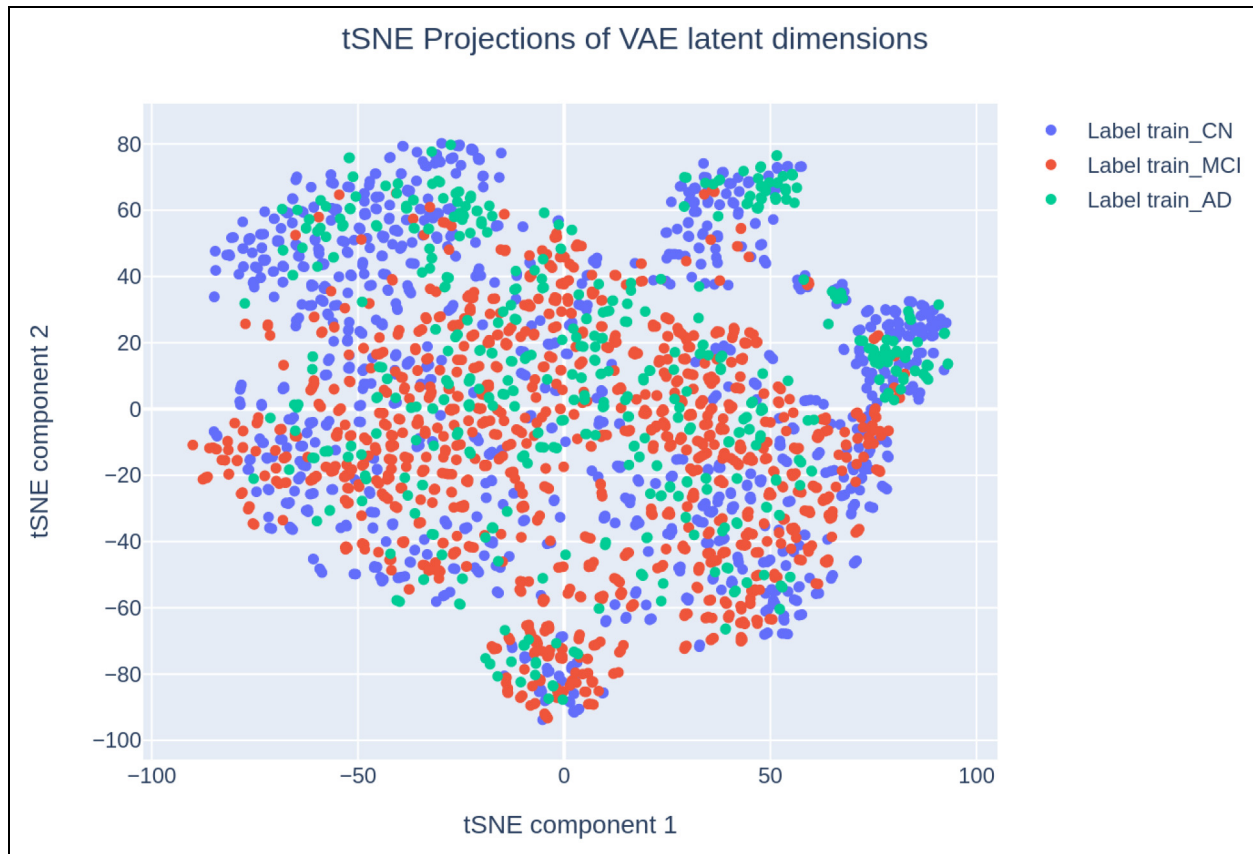
### Cluster identification by t-distributed stochastic neighborhood embedding

The testing and assessing of feature extraction from MRI scans using VAE was conducted on t-distributed stochastic neighborhood embedding (t-SNE) of the latent vectors from the VAE generator of all classes. We found complete image harmonization in the t-SNE embeddings of different classes from three overlapping clusters: CN, AD, and MCI (Figure 3, Supplemental Figure 1). Clustering results on the t-SNE plots indicate the samples were generated from a normal distribution (Feature [1]: t-SNE embedded dimension 1 and Feature [2]: t-SNE embedded dimension 2; AD, CN, MCI). We identified three different diagnoses from the latent features encoded by VAE with a hierarchical presentation. Thereby, we could bring class balance among training samples of CN, MCI, and AD classes. The reduced dimensional latent representation of MRI scans generated using VAE is further reduced to two dimensions to visualize the class-wise embeddings and clusters. 2-dimensional t-SNE embedding plots assessed the confounding relationship between the diagnostic status and a specific form of metadata. There was no apparent distant cluster formation of metadata after the post-processing of MRI embeddings in both cohorts (ADNI and OASIS) (Supplemental Figure 1).

### Control/mild cognitive impairment/Alzheimer's disease classification

The algorithm we developed is a combination of VAE and the addition of one of the different advanced ML classifiers (ExtraTree, LGBM, SVM-Linear Kernel, XGB, and Random Forest). First, the machine was trained with ADNI data presented in Table 1A. After the training, the algorithm was tested using the ADNI data presented in





**Figure 3.** A schematic view of cluster identification by t-distributed stochastic neighborhood embedding (t-SNE) of the latent vectors. It shows the projection of variational autoencoder (VAE) latent dimensions. The reduced dimensional latent representation of MRI scans generated using VAE is further reduced to two dimensions to visualize the class-wise embeddings and clusters. Individual points depict projection of MRI scans from a single patient. t-SNE plot is derived from training samples from ADNI cohort. (Feature [1]: t-SNE embedded dimension 1 and Feature [2]: t-SNE embedded dimension 2; AD, Alzheimer's disease; CN, control; MCI, late mild cognitive impairment).

Table 1B. The data from the OASIS (Table 2) were used to validate the algorithm independently. The performance of different machine learning models is presented in terms of accuracy, precision, recall, and F1 score for the ADNI test data set (Tables 3 and 4). The classification validation results are presented in Table 5 using OASIS data. The receiver operating characteristic (ROC) is traced from the true positive rate and false positive rate of prediction. The study presents each case's diagnostic ability in terms of classification performance (Figures 4–6; Tables 3–5). The most important finding is a classification of MCI with autopsy-confirmed patients. Figure 6 and Table 3 depict the classification performance of autopsy-confirmed AD (Autopsy-AD) versus MCI. XGB classifier performed best (Accuracy = 92.78%; Precision = 91.42%, Recall = 100%, F1-Score = 95.52%), and other classifiers always performed with an accuracy of >70%. The model we generated has been validated with the OASIS cohort (Table 5). Several published studies comprising volumetric measures, cortical thickness, and gray matter (Table 6) did not validate their

classification in a separate cohort. In the case of AD versus MCI, the current algorithm outperformed existing published data (Table 6). Autopsy-confirmed AD patients with serial MRI scans are not readily available. The highest accuracy value in the autopsy-AD versus MCI case confirmed the algorithm's genuine power. The performances of some ML models are slightly less accurate (Table 3). Which is also reflected in ROC traces (Figures 4–6). Some of the MRI scans are not perfectly aligned in the case of the CN data set of the TADPOLE data format. Abnormality may cause a drop in accuracy in the case of AD versus CN classification.

Two interesting characteristics of the current algorithm are that it is the best performing autopsy-confirmed AD cases and less accurate for MCI cases. The algorithm has been validated in a well-recognized separate cohort. In general, autopsy-confirmed AD cases are the 'Gold Standard' of AD diagnosis by NIH and Alzheimer's Association. However, the most uncertain case is MCI. Results from the present algorithm confirm the diagnostic

**Table 3.** Summarized classification by the current algorithm using three machine learning models in combination with variational autoencoder in ADNI data set.

Classification	Model	Accuracy (%)	Recall (%)	Precision (%)	F1-Score (%)
CN versus AD	LGBM – Classifier	75.45	85.16	74.58	79.52
	SVM – Linear Kernel	75.45	83.87	75.14	79.27
	XGB – Classifier	74.01	86.45	72.43	78.82
	ExtraTree – Classifier	67.51	81.94	67.2	73.84
MCI versus AD	LGBM – Classifier	81.41	95.28	80.14	87.06
	ExtraTree – Classifier	80.85	93.56	80.44	86.51
	XGB – Classifier	78.31	92.7	78.26	84.87
	Random Forest – Classifier	78.03	92.7	77.98	84.71
MCI versus Autopsy AD	XGB – Classifier	92.78	91.42	100	95.52
	ExtraTree – Classifier	77.26	85.41	87.28	86.33
	LGBM – Classifier	73.65	81.97	86.04	83.96
	Random Forest – Classifier	73.65	83.69	84.78	84.23

LGBM, light gradient boosting machine; SVM, support vector machines; XGB, extreme gradient boosting.

**Table 4.** Best Algorithm for specific classification of Alzheimer's disease (AD).

Patient combination	Accuracy (%); F1-Score (%)	Classifier
AD versus CN	75.45; 79.52	LGBM
Autopsy-confirmed AD versus MCI	92.75; 95.52	XGB
AD versus MCI	81.41; 87.06	LGBM

criteria of the NIH-AA 2011 (Working Group: New guidelines for the diagnosis of Alzheimer's disease<sup>4</sup>; and IGW-2 2014 (International Working Group 2: New guidelines for the diagnosis of Alzheimer's disease<sup>54</sup>).

### Data leakage

Data leakage is an invariable problem in AI/ML image classification.<sup>31,55</sup> To address this issue, we have taken the following precautions as mentioned by Wen et al.<sup>31</sup> The data (ADNI) was split into training and test sets at the very beginning at the subject level. The training set was used to train the model, and the test data set was used for testing the model. A separate cohort data set (OASIS) was used to validate the model. Splitting between training and test sets is performed at the subject level (see Table 1A and 1B, Table 2). Splitting at the subject level is essential because, otherwise, the results are biased due to data leakage.

### Class imbalances in classification results

There is a slight imbalance to a reduced number of AD images. However, such a slight imbalance will not affect results. We intend to distinguish MCI versus autopsy AD during the test only, not in training. However, since we

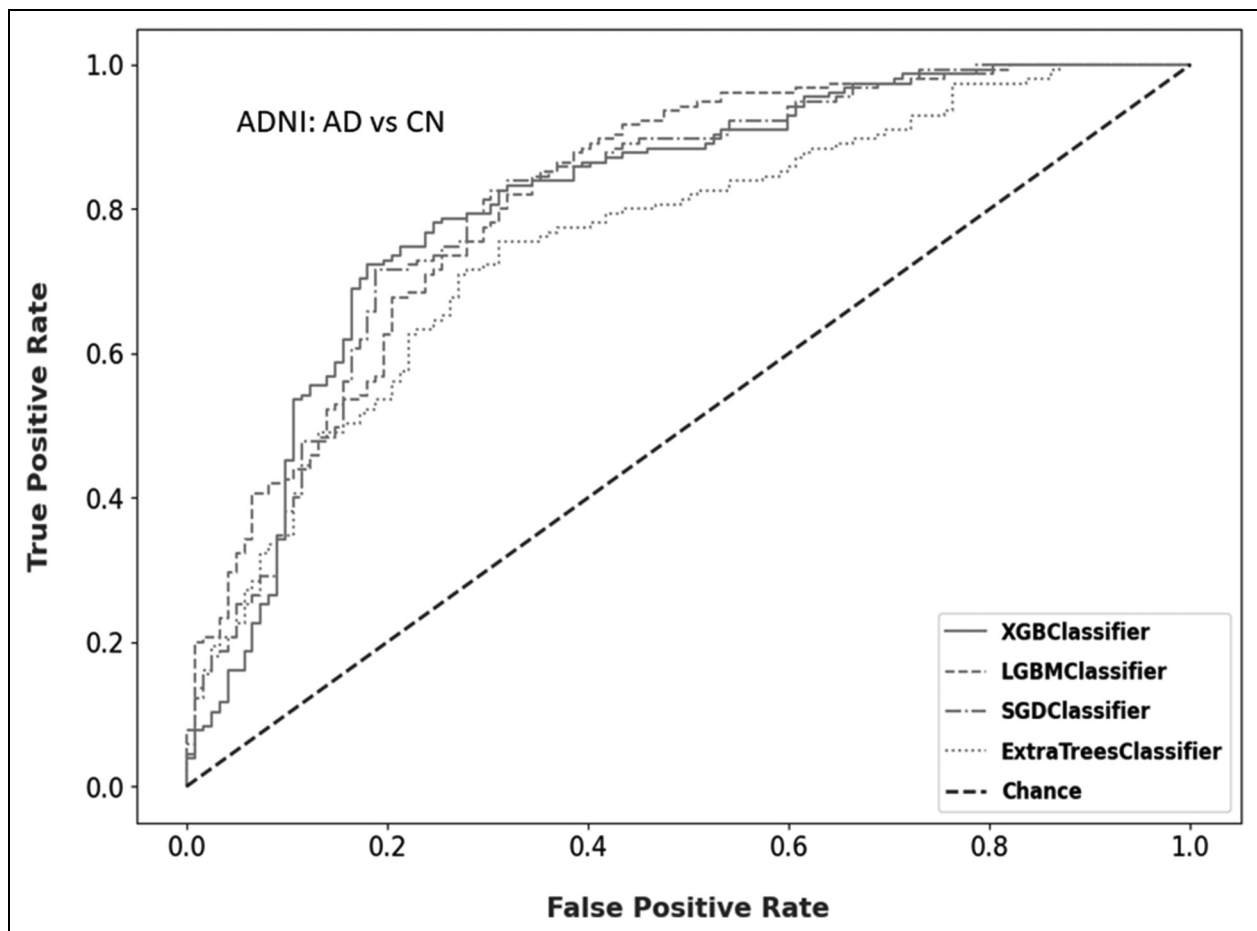
only use the autopsy-confirmed AD cases in the testing subset, we do not induce any class imbalance while training the model. Moreover, the 2-components of the tSNE embeddings do not show any formation of clusters specific to a class. We see smaller clusters formed by the images from the same subject (Figure 3 and Supplemental Figure 1). This suggests that latent encodings generated by the VAE model are not biased toward a specific class of images. Thus, the imbalance in the count of images does not affect the encoding model during training.

## Discussion

The non-invasive sMRI has been used extensively in volumetric atrophy measurement, and it does not require radioactivity exposure like PET, can easily be repeated in the same subjects with no harm, are relatively inexpensive, and can be operated with scanner machines available in almost any clinics. The European Medicine Agency has been approved as a biomarker for AD clinical trials. There is ample scope for ML/DL to be used for sMRI analysis platform support to detect the different stages of dementia in the AD continuum, from non-demented control to preclinical AD to MCI and AD. There are several attempts to classify AD and MCI using only an MRI scan by a deep neural network without NIH 'Gold Standard' autopsy validation.<sup>34–41</sup> A systemic review of ML of neuroimaging found that the accuracy was highest for differentiating AD versus healthy controls and poor for differentiating MCI versus AD.<sup>29</sup> Most of them do not have NIH 'Gold Standard' autopsy validation. Only one published autopsy-validated study showed that the accuracy of classifying AD versus MCI cases was not adequate.<sup>42</sup> The current study with autopsy-validated produced MCI versus autopsy-confirmed AD 92.78% accuracy with very high precision, F1-score, and recall value (Precision = 100%; F1-score = 95.52%; Recall = 91.42%). The study

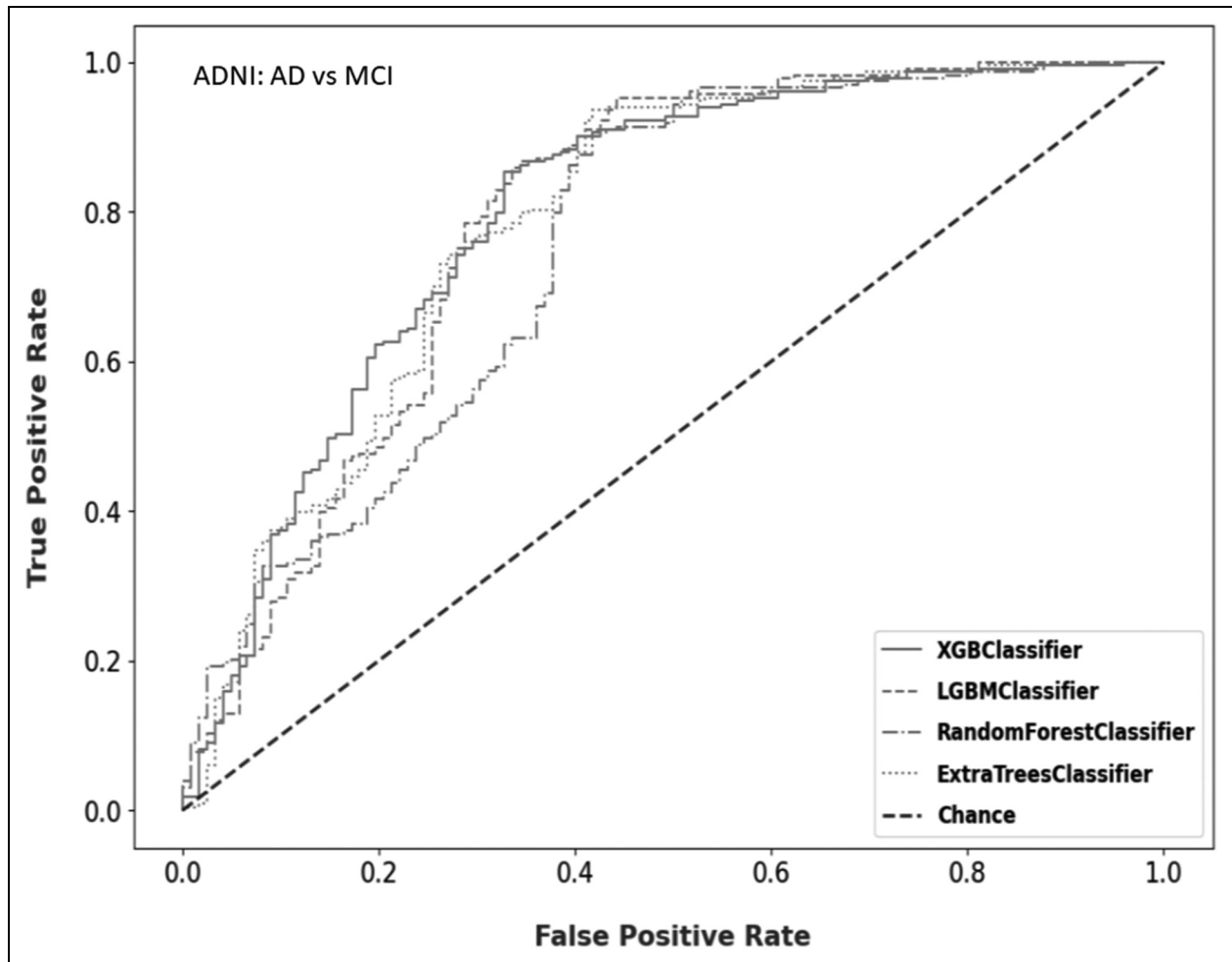
**Table 5.** Validation of the Algorithm in a separate cohort (OASIS).

Classification	Model	Accuracy (%)	Recall (%)	Precision (%)	F1-Score (%)
CN versus AD	ExtraTree – Classifier	86.16	97.72	86.97	92.03
	XGB – Classifier	85.85	95.44	88.22	91.69
	LGBM – Classifier	85.54	92.21	90.32	91.25
	SVM – Linear Kernel	84.14	90.49	90.15	90.32
MCI versus AD	XGB – Classifier	70.03	98.52	70.37	82.1
	ExtraTree – Classifier	69.25	98.89	69.71	81.78
	LGBM – Classifier	69.25	98.85	69.92	81.66
	RandomForest – Classifier	68.73	97.41	69.76	81.3

**Figure 4.** Performance of classification of Alzheimer's disease (AD) versus non-demented control (CN). The receiver operating characteristic (ROC) is traced from the true positive rate and false positive rate of prediction.

found that the CN versus AD classification performance is less accurate than the classification between CN versus MCI and MCI versus AD. ADNI clinical diagnoses of CN/MCI/AD were made based on neuropsychological tests (e.g., CDR, Sum of Boxes, ADAS11, ADAS13, MMSE, RAVLT, MoCA, ECog), MRI ROIs (measures of brain structural integrity, volumes, cortical thicknesses surface areas), FDG-PET, DTI-ROI measures, and CSF biomarkers. The VAE may have overestimated the brain features as AD in some of CN cases. On the other hand, it has

been well established that pathological AD features started much before the clinical manifestation of AD. Approximately 30% of cognitively normal older individuals have plaques and tangles in their brains. Some of the AD underlying pathological features in the brain may be detected in CN cases by VAE. As a result, the study found a less accurate classification of CN cases compared to MCI and AD cases. AD-related feature extraction by VAE may be more precise in cases of MCI and AD cases than CN. As a result, the study found better accuracy in

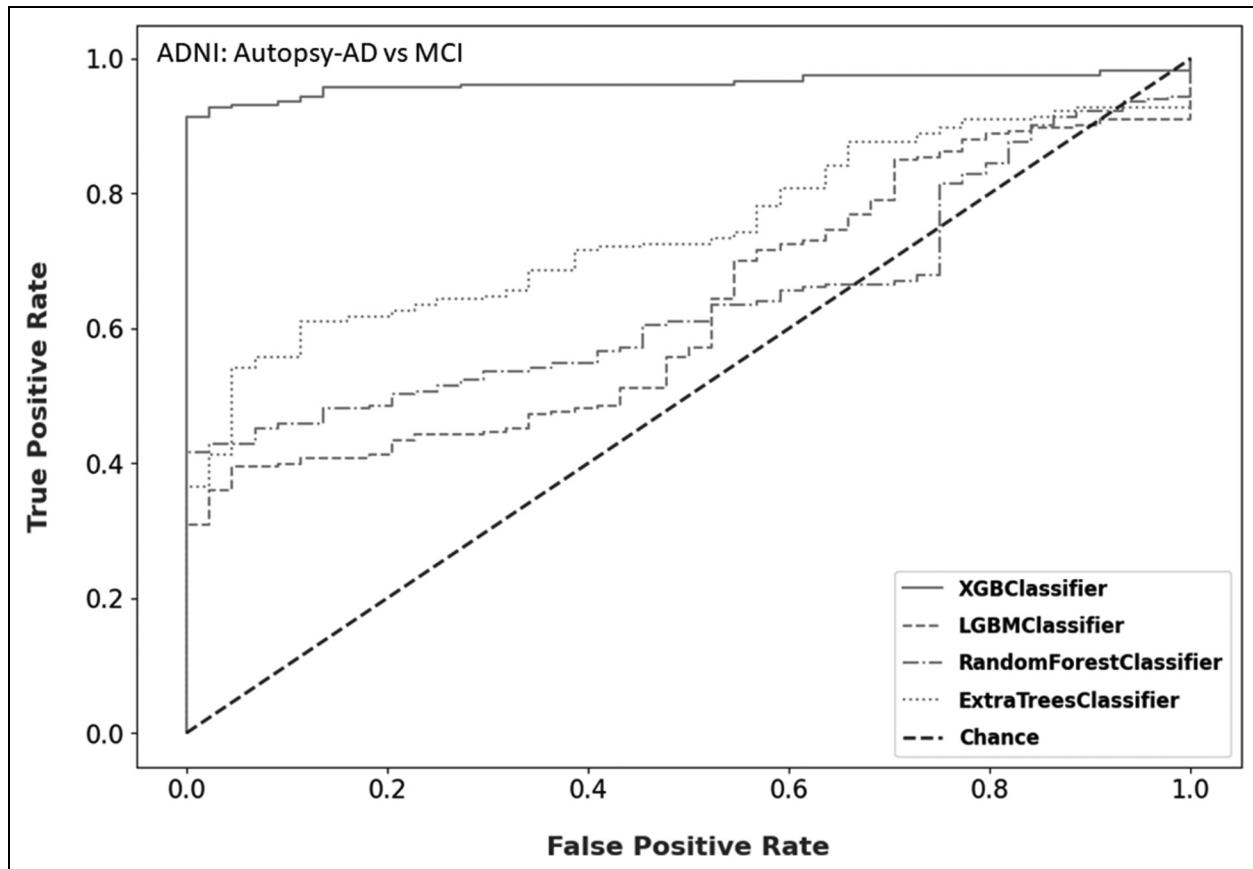


**Figure 5.** Performance of classification of Alzheimer's disease (AD) versus mild cognitive impairment (MCI). The receiver operating characteristic (ROC) is traced from the true positive rate and false positive rate of prediction.

classifying MCI and AD cases. It is to be noted that invasive CSF biomarkers are also not very accurate, particularly amyloid- $\beta_{42}$ . The conversion rate of MCI to AD is 10–15% per year, and 80% of MCI will eventually be converted to AD.<sup>35</sup> The diagnosis of MCI versus AD and CN is always tricky. Some of the MCI cases can be reversed back to normal cognitive condition, and several may convert to non-AD dementia cases.

The encoder in VAE, consisting of 3D CNN blocks, plays a pivotal role in compressing the input MRI data into a lower-dimensional representation commonly known as a latent space. Its primary objective is to discern and encapsulate the critical features and intricate patterns inherent in MRI brain images. This encoding process is the foundation for various downstream tasks such as data compression, feature extraction, and denoising. To ensure that the model effectively captures the underlying structures in MRI brain images, various data augmentation techniques, including translation, scaling, rotation, flipping the 3D image around an axis, and adding Gaussian noise,

were applied with specific probabilities. The decoder in the current MRI autoencoder, mirroring the encoder, contains 3D CNN blocks and aims to reconstruct the compressed latent space back into the original MRI data format. It employs a symmetrical architecture to the encoder, reversing the encoding process to generate output images that closely resemble the non-augmented original MRI data. The loss function incorporates the mean squared error (MSE) and Kullback-Leibler (KL) divergence terms. The MSE term is calculated within the masked region (the region where the brain tissue is present in the original image) of both the model's output and the original image, measuring pixel-wise reconstruction fidelity. Simultaneously, the KL divergence term is instrumental in regularizing the latent space representations. It encourages them to follow a standard Gaussian distribution with a mean of zero and a standard deviation of one. This dual-loss approach ensures accurate reconstruction and effective latent space regularization, contributing to the overall model's robustness and performance in handling MRI brain data. Our novel approach using a VAE excels



**Figure 6.** Performance of classification of autopsy-confirmed Alzheimer's disease (Autopsy-AD) versus mild cognitive impairment (MCI). The receiver operating characteristic (ROC) is traced from the true positive rate and false positive rate of prediction.

at extracting crucial features from MRI brain images self-supervised. By incorporating the VAE framework, the model learns to capture the data's inherent structure and salient features without the need for explicit feature labeling or supervision. This self-supervised learning process is especially advantageous in MRI data, where identifying crucial features can be complex and multifaceted. The VAE's ability to implicitly discover and represent these features within the latent space not only streamlines the modeling process but also opens the door to finding previously unknown or subtle patterns and information within the MRI data, ultimately enhancing our understanding of the underlying neurobiology.

We conduct a set of comprehensive experiments on ADNI. The results demonstrate that the use of VAE can achieve robust classification. In the data augmentation procedure, VAE exploits the features from 3-D MRI scans and encodes the images into 2-D latent space. It compresses the data, but at the same time, it concentrates on the latent features of the MRI scans. Unlike principal component analysis or independent component analysis, VAE encodes 3-D input into a form Gaussian distribution, not like a point or regular distinct entity. Data augmentation is the art of increasing the size of a given data set by creating

synthetic labeled data. Using VAE, the study achieved a better or comparable accurate classification with much better F-1 scores than one of the best and most recent DL-based AD/MCI dementia classification studies. Multimodal deep learning has achieved a better classification.<sup>56</sup>

The current algorithm outcomes perfectly found that diagnosis of MCI is less accurate than autopsy-confirmed AD. It is obvious that almost all MCI patients will not be converted to AD eventually. In fact, there are reports that some of the MCI cases may come back to the non-dementia phase. A study found that ~16% of MCI cases went reverse back to normal or near-normal cognition approximately 1-year later.<sup>57</sup> (MCI can also be divided into two categories: amnesic MCI and multimodal MCI or non-amnesic MCI. MCI with primarily memory deficits is called amnesic MCI. Multimodal MCI includes MCI with problems in thinking skills, inability to make sound decisions and judgments, and failure to take the sequential steps needed to perform relatively complex tasks. In general, individuals with amnesic MCI eventually develop AD, and those with multimodal MCI develop non-AD dementia. MCI also has been divided into two categories based on clinical criteria, such as MMSE and Clinical CDR, late MCI, and

**Table 6.** Comparison of results with other published MRI-based machine learning models.

Feature inputs	Data and Results	Remarks	References
MRI: unprocessed	Data source: ADNI and OASIS ML Classifier: XGB Results: AUC Accuracy = 0.93 (MCI versus Autopsy-confirmed AD) ML Classifier: LGBM Results: AUC Accuracy = 0.81 (MCI versus AD)	External validation of the model by an independent cohort improved the quality and novelty of the classification algorithm.	Current study
MRI: Whole brain volume measures	Data source: Sample collected on site for study. ML classifier: SVM Result: AUC ROC = N/A Accuracy = 0.50 (pMCI versus sMCI)	Model was not verified with other data sets.	Plant et al. (2010) <sup>33</sup>
MRI: GM volumes	Data source: ADNI ML classifier: SVM Results: AUC ROC = 0.74 Accuracy = N/A (pMCI versus cMCI)	Model was not verified with other data sets.	Chincarini et al. (2011) <sup>34</sup>
MRI: 3D hippocampal morphometric measures	Data source AddNeuroMed; ML classifier: SVM with RBF kernel Results: AUC ROC = N/A Accuracy = 0.80 (pMCI versus cMCI)	Model was not verified with other data sets.	Costafreda et al. (2011) <sup>35</sup>
MRI: Hippocampal volume Other information: Demographic, APOE, genotypes, CSF biomarkers	Data source: ADNI ML classifier: SVM Results: AUC ROC = 0.68 Accuracy = 0.68 (pMCI versus cMCI)	Model was not verified with other data sets. The model will not be practical because need genetic testing plus invasive CSF collection.	Apostolova et al., (2014) <sup>10</sup>
MRI: 3D brain volumes	Data source: ADNI ML classifier: SVM Results: AUC ROC = Accuracy = 0.97 (pMCI versus sMCI)	Model was not verified with other data sets.	Guerrero et al. (2014) <sup>36</sup>
MRI: 3D brain volumes	Data source: ADNI ML classifier: SVM Results: AUC ROC = 0.87 Accuracy = 0.79 (pMCI versus sMCI)	Model was not verified with other data sets.	Zhang et al. (2021) <sup>37</sup>
MRI: cortical thickness	Data source: ADNI ML classifier: SVM Results: Results: AUC ROC = N/A Accuracy = 0.68 (pMCI versus cMCI)	Model was not verified with other data sets.	Cho et al. (2012) <sup>38</sup>
MRI: cortical thickness Other information: demographic variables, and APOE4 genotype	Data source: ADNI ML classifier: RF Results: AUC ROC = 0.83 Accuracy = 0.82 (pMCI versus sMCI)	Model was not verified with other data sets.	Lebedev et al. (2014) <sup>39</sup>
MRI: cortical thickness	Data source: ADNI ML classifier: Graph neural network Results: AUC ROC = N/A Accuracy: Conversion from: eMCI to AD: 0.79 IMCI to AD: 0.65	Model was not verified with other data sets.	Wee et al. (2019) <sup>40</sup>

(continued)

**Table 6.** Continued.

Feature inputs	Data and Results	Remarks	References
MRI: GM density	Data source: ADNI ML classifier: SVM Results: AUC ROC = N/A Accuracy = 0.80 (pMCI versus sMCI)	Model was not verified with other data sets.	Wen et al. (2021) <sup>41</sup>

AD, Alzheimer's disease; ADNI, Alzheimer's Disease Neuroimaging Initiative; APOE, Apolipoprotein; AUC, area under the curve; cMCI, converting MCI; CSF, cerebrospinal fluid; eMCI, early MCI; GM, gray matter; LGBM, Light Gradient Boosting Machine; IMCI, late MCI; MCI, mild cognitive impairment; ML, machine learning; MRI, magnetic resonance imaging; N/A, not available; OASIS, Open Access Series of Imaging Studies; pMCI, progressive MCI; RBF, radial basis function; RF, random forest; ROC, receiver operating characteristic; sMCI, stable MCI; SVM, support vector machine; XGB, eXtreme Gradient Boosting.

early MCI.<sup>58</sup> MCI proceeds to AD with an annual rate of 10–12%,<sup>59</sup> and an individual with MCI is expected to convert to AD within five years (about 5–25% per year).<sup>60</sup> Another study found a slightly higher conversion rate (10–25%) from MCI to AD.<sup>61</sup> Patients with MCI who are progressing to AD have converting-MCI, and those who continue to stay at the MCI stage have stable MCI. Individuals with converting-MCI might be better candidates for inclusion in clinical studies of AD drugs to test their efficacy in slowing progression to moderate and severe AD.

More than 99% of AD drug trials failed previously. Three primary reasons for failure are the selection of wrong targets, biological heterogeneity of disease mechanism, and poor patient selection. An ideal trial design should have a placebo-AD arm with a measurable cognitive decline due to AD-related brain pathology changes faster than the treatment arm. The successful completion of this project will help patient stratification in AD clinical trials. Two major marketing segments will benefit from this research work: the clinical facilities and the AD drug development companies. The clinical facilities include radiologists, neurologists, and geriatric psychologists' medical centers with MRI facilities. They will benefit from early detection of AD converted from MCI using only MRI scans. Expected confirmed results would help the caregiver with the right plan. We deliberately selected early AD patients by considering the subset of scans for which the date of study/visit is under 18 months from the baseline. By focusing on accurate diagnosis of AD at an earlier stage would help better patient stratification in AD clinical trials. The second beneficiary is development industries to stratify the suitable MCI cases to convert to AD. The early-stage AD patient selection before widespread neuronal loss is a key for AD drug development using MRI data. The proposed project has the potential to disrupt the targeted market segments. NIA and AA have incorporated MRI as one of the AD biomarkers to detect AD for research and clinical trial purposes in 2011.<sup>4</sup> The European Medicines Agency has introduced MRI data as the primary biomarker data for patient stratification. Moreover, U.S. Food and Drug Administration clearance is not necessary for this marketing segment.

An estimated 47 million Americans (1 in 5 above the age of 50 years) have preclinical AD and are at risk of developing cognitive impairment in the future. There is a scope for improvement of this ML algorithm by using the region-specific difference between AD and control brain MRI images as inputs. We intend to improve the ML algorithm to capture subtle changes in brain structure that occur before widespread neuronal loss. Such characteristics of an algorithm will enable us to predict the predementia phase before the onset of AD in the preclinical stage of the disease.

The study has not demonstrated how AD can be distinguished from other non-AD dementias (e.g., multi-infarct dementia, dementia due to Parkinson's disease, Lewy body dementia, frontotemporal dementia, etc.). The study will continue to work to formulate new algorithms to distinguish AD from other non-AD dementias using region-specific MRI scans. Moreover, brain autopsy studies found more than 50% of cases of AD pathologies remain co-morbid with other forms of neurodegeneration. In the future, we plan to introduce more specific argument base and brain-region-specific differentiating parameters for how AD pathology can be distinguished in the presence of another dementia. In a future study, the present algorithm will apply other factors that are risk factors for AD pathology, for example, predisposition of AD risk factors genes (*APOE4*, *BIN1*, *CLU*, *ABCA7*, *CRI*, *PICALM*, *CD33*, *MS4A*, *SORL1*, etc.), demography (education, sex, ethnicity) and PET scan data. The multimodal algorithm model will be able to improve the diagnosis of CN versus MCI versus AD and can be able to diagnose AD from other non-AD dementia cases. The goal is to extend the capability of the algorithm to predict the pre-clinical stage of AD. This is the first complete study where VAE has been employed to classify CN/MCI/AD. However, in a recent report, VAE has been utilized to explore multimodal AD progression modeling.<sup>18</sup> In our ongoing study, we are utilizing VAE for predicting dementia state in combination with other ML modalities. In the future to improve classification of AD, MCI, and CN we intend to add multimodal approach, e.g., MRI with other variables (e.g., age >60,

sex, educational background, and psychometric test) in heterogeneous populations (e.g., a combination of Hispanic, Black, and Caucasian ethnicity).

Only MRI-based ML models are included.

### Limitations

Only the CDR score for the clinical diagnosis has been performed in the OASIS dataset compared to the ADNI TADPOLE data set, where numerous factors have been employed to diagnose AD/MCI/CN clinically. Therefore, the OASIS data set may be less accurate than the clinical diagnosis of the ADNI TADPOLE data set. One of the important aspects of the study is to examine the performance in 'gold standard' autopsy-confirmed AD cases. The study has only 14 autopsy-confirmed cases in a single cohort (ADNI TADPOLE). There were no autopsy-confirmed cases in the OASIS cohort. Therefore, the capability of the classifiers has not been tested in the data set obtained from multi-cohorts. We are now planning to validate and generate more data by testing the algorithm in a third cohort, National Alzheimer's Coordination Committee (NACC) patients. NACC functions as the centralized data repository and collaboration and communication hub for the NIA's Alzheimer's Disease Research Centers program, which currently includes 33 centers and 4 exploratory centers across the US. NACC cohorts have non-AD dementia data. The algorithm has not been tested for non-AD dementia cases and predictivity of conversion of MCI-converter and non-converter cases. We are now working in this direction.

### Conclusion

The study concluded that 3-D MRI scans could successfully diagnose the course of AD-related dementia after feature extraction of images by VAE and using AI/ML algorithms. The model comprises two sections, and the first part is VAE, which features extraction from MRI scans. The extracted latent representation is then fed into a discriminative ML model. We propose a methodology that leverages the power of VAEs to extract latent distributions and the discriminating power of different machine learnings. The current study achieved classification AD versus CN with an accuracy of 75.45% (F1-score = 79.52%), AD versus MCI with an accuracy of 81.41% (F1-Score = 87.06%), and autopsy-confirmed AD versus MCI with an accuracy of 92.75% (F1-Score = 95.52) using ADNI data set. By overcoming the data leakage problem, the ML classification model is tested in two independent cohorts. When tested in an independent cohort (OASIS), the study achieved the classification of AD versus CN with an accuracy of 86.16% (F1-score = 92.03%) and AD versus MCI with an accuracy of 70.03% (F1-Score = 82.1%). External validation has improved the quality and

novelty of the classification algorithm. The current study has three novelties: 1) The study is 'gold standard' autopsy validated. The algorithm is tested in autopsy-confirmed cases which is very rare in AI/ML AD imaging studies; 2) The algorithm is validated by a separate independent cohort; and 3) The study has taken care of data leakage by separating training and testing data sets at subject level before training.

In the future, predictions of the current and future state of a patient's AD can be computed from our model by generating latent representations of the patient's sMRI scan (current) using VAE. Sampling from the generated conditional distribution helps augment the dataset. The ability to sample from any conditional distribution is one advantage a modeling framework based on VAE and machine learning has over alternative nongenerative models.


### Acknowledgments


The research was partly supported by the Intramural Research Program of BioImaginix LLC (T.K.).

Data collection and sharing for the Alzheimer's Disease Neuroimaging Initiative (ADNI) is funded by the National Institute on Aging (National Institutes of Health Grant U19AG024904). The grantee organization is the Northern California Institute for Research and Education. In the past, ADNI has also received funding from the National Institute of Biomedical Imaging and Bioengineering, the Canadian Institutes of Health Research, and private sector contributions through the Foundation for the National Institutes of Health (FNIH) including generous contributions from the following: AbbVie, Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics.

### ORCID iDs

Subhrangshu Bit  <https://orcid.org/0000-0002-3566-9500>

Arnab Maji  <https://orcid.org/0000-0003-0462-8324>

Tapan K Khan  <https://orcid.org/0000-0002-0737-3884>

### Statements and declarations

#### Author contributions/CRediT

Subhrangshu Bit (Data curation; Formal analysis; Methodology; Software; Writing – original draft; Writing – review & editing); Pritam Dey (Data curation; Formal analysis; Software; Validation); Arnab Maji (Methodology; Resources); Tapan Khan (Conceptualization; Data curation; Formal analysis; Funding acquisition; Investigation; Methodology; Project administration;



Resources; Software; Supervision; Validation; Visualization; Writing – original draft; Writing – review & editing).

### Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

### Declaration of conflicting interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

### Supplemental material

Supplemental material for this article is available online.

### References

1. Titano JJ, Badgeley M, Schefflein J, et al. Automated deep-neural-network surveillance of cranial images for acute neurologic events. *Nat Med* 2018; 24: 1337–1341.
2. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. *Nat Med* 2019; 25: 44–56.
3. Bernal J, Kushibar K, Asfaw DS, et al. Deep convolutional neural networks for brain image analysis on magnetic resonance imaging: a review. *Artif Intell Med* 2019; 95: 64–81.
4. Sperling RA, Aisen PS, Beckett LA, et al. Toward defining the preclinical stages of Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimers Dement* 2011; 7: 280–292.
5. Mitchell AJ and Shiri-Feshki M. Rate of progression of mild cognitive impairment to dementia – meta-analysis of 41 robust inception cohort studies. *Acta Psychiatr Scand* 2009; 119: 252–265.
6. Jack CR, Bennett DA, Blennow K, et al. NIA-AA research framework: toward a biological definition of Alzheimer's disease. *Alzheimers Dement* 2018; 14: 535–562.
7. Grueso S and Viejo-Sobera R. Machine learning methods for predicting progression from mild cognitive impairment to Alzheimer's disease dementia: a systematic review. *Alzheimers Res Ther* 2021; 13: 62.
8. Pan D, Zeng A, Jia L, et al. Early detection of Alzheimer's disease using magnetic resonance imaging: a novel approach combining convolutional neural networks and ensemble learning. *Front Neurosci* 2020; 14: 59.
9. Khan TK. An algorithm for preclinical diagnosis of Alzheimer's disease. *Front Neurosci* 2018; 12: 75.
10. Apostolova LG, Hwang KS, Medina LD, et al. Cortical and hippocampal atrophy in patients with autosomal dominant familial Alzheimer's disease. *Dement Geriatr Cogn Disord* 2011; 32: 118–125.
11. Frisoni GB, Fox NC, Jack CR, et al. The clinical use of structural MRI in Alzheimer disease. *Nat Rev Neurol* 2010; 6: 67–77.
12. Vemuri P and Jack CR. Role of structural MRI in Alzheimer's disease. *Alzheimers Res Ther* 2010; 2: 23.
13. Dickerson BC and Wolk DA. MRI Cortical thickness biomarker predicts AD-like CSF and cognitive decline in normal adults. *Neurology* 2012; 78: 84–90.
14. Sima DM, Phan TV, Van Eyndhoven S, et al. Artificial intelligence assistive software tool for automated detection and quantification of amyloid-related imaging abnormalities. *JAMA Netw Open* 2024; 7: e2355800.
15. Fouladvand S, Noshad M, Periyakoil VJ, et al. Machine learning prediction of mild cognitive impairment and its progression to Alzheimer's disease. *Health Sci Rep* 2023; 6: e1438.
16. Basu S, Wagstyl K, Zandifar A, et al. Early prediction of Alzheimer's disease progression using variational autoencoders. In: Shen D, Liu T and Peters TM, et al. (eds) *Medical image computing and computer assisted intervention – MICCA. Lecture Notes in Computer Science*, vol. 11767. Cham: Springer, 2019, pp.205–213.
17. Cobbinah BM, Sorg C, Yang Q, et al. Reducing variations in multi-center Alzheimer's disease classification with convolutional adversarial autoencoder. *Med Image Anal* 2022; 82: 102585.
18. Martí-Juan G, Lorenzi M, Piella G, et al. MC-RVAE: multi-channel recurrent variational autoencoder for multimodal Alzheimer's disease progression modelling. *Neuroimage* 2023; 268: 119892.
19. Kumar S, Payne PRO and Sotiras A. Normative modeling using multimodal variational autoencoders to identify abnormal brain volume deviations in Alzheimer's disease. *Proc SPIE Int Soc Opt Eng* 2023; 12465: 1246503.
20. Wang X, Zhou R, Zhao K, et al. Normative modeling via conditional variational autoencoder and adversarial learning to identify brain dysfunction in Alzheimer's disease. In: *IEEE 20th International Symposium on Biomedical Imaging (ISBI) 2023*, Cartagena, Colombia, pp.1–4.
21. Bandyopadhyay S, Dion C, Libon DJ, et al. Variational autoencoder provides proof of concept that compressing CDT to extremely low-dimensional space retains its ability of distinguishing dementia. *Sci Rep* 2022; 12: 7992.
22. Vivek S, Faul J, Thyagarajan B, et al. Explainable variational autoencoder (E-VAE) model using genome-wide SNPs to predict dementia. *J Biomed Inform* 2023; 148: 104536.
23. Hong J, Kang SK, Alberts I, et al. Image-level trajectory inference of tau pathology using variational autoencoder for Flortaucipir PET. *Eur J Nucl Med Mol Imaging* 2022; 49: 3061–3072.
24. Feng Y, Chandio BQ, Villalon-Reina JE, et al. Deep normative tractometry for identifying joint white matter macro- and micro-structural abnormalities in Alzheimer's disease. *BioRxiv* 2024. <https://doi.org/10.1101/2024.02.05.578943>. [Preprint]. Posted February 06, 2024.
25. Pinaya WHL, Mechelli A and Sato JR. Using deep autoencoders to identify abnormal brain structural patterns in neuropsychiatric disorders: a large-scale multi-sample study. *Hum Brain Mapp* 2019; 40: 944–954.
26. Huang Q, Qiao C, Jing K, et al. Biomarkers identification for Schizophrenia via VAE and GSDAE-based data augmentation. *Comput Biol Med* 2022; 146: 105603.
27. Yamaguchi H, Hashimoto Y, Sugihara G, et al. Three-dimensional convolutional autoencoder extracts features of structural brain images with a “diagnostic label-free”

- approach: application to schizophrenia datasets. *Front Neurosci* 2021; 15: 652987.
28. Moradi E, Pepe A, Gaser C, et al. Machine learning framework for early MRI-based Alzheimer's conversion prediction in MCI subjects. *Neuroimage* 2015; 104: 398–412.
  29. Pellegrini E, Ballerini L, Hernandez MDCV, et al. Machine learning of neuroimaging for assisted diagnosis of cognitive impairment and dementia: a systematic review. *Alzheimers Dement (Amst)* 2018; 10: 519–535.
  30. Beheshti I, Demirel H and Matsuda H. Classification of Alzheimer's disease and prediction of mild cognitive impairment-to-Alzheimer's conversion from structural magnetic resource imaging using feature ranking and a genetic algorithm. *Comput Biol Med* 2017; 83: 109–119.
  31. Wen J, Thibeau-Sutre E, Diaz-Melo M, et al. Convolutional neural networks for classification of Alzheimer's disease: overview and reproducible evaluation. *Med Image Anal* 2020; 63: 101694.
  32. Zhang Y, Dong Z, Phillips P, et al. Detection of subjects and brain regions related to Alzheimer's disease using 3D MRI scans based on eigenbrain and machine learning. *Front Comput Neurosci* 2015; 2: 66.
  33. Plant C, Teipel SJ, Oswald A, et al. Automated detection of brain atrophy patterns based on MRI for the prediction of Alzheimer's disease. *Neuroimage* 2010; 50: 162–174.
  34. Chincarini A, Bosco P, Calvini P, et al. Local MRI analysis approach in the diagnosis of early and prodromal Alzheimer's disease. *Neuroimage* 2011; 58: 469–480.
  35. Costafreda SG, Dinov ID, Tu Z, et al. Automated hippocampal shape analysis predicts the onset of dementia in mild cognitive impairment. *Neuroimage* 2011; 56: 212–219.
  36. Guerrero R, Wolz R, Rao AW, et al. Manifold population modeling as a neuro-imaging biomarker: application to ADNI and ADNI-GO. *Neuroimage* 2014; 94: 275–286.
  37. Zhang J, Zheng B, Gao A, et al. A 3D densely connected convolution neural network with connection-wise attention mechanism for Alzheimer's disease classification. *Magn Reson Imaging* 2021; 78: 119–126.
  38. Cho Y, Seong JK, Jeong Y, et al. Individual subject classification for Alzheimer's disease based on incremental learning using a spatial frequency representation of cortical thickness data. *Neuroimage* 2012; 59: 2217–2230.
  39. Lebedev AV, Westman E, Van Westen GJP, et al. Random forest ensembles for detection and prediction of Alzheimer's disease with a good between-cohort robustness. *Neuroimage Clin* 2014; 6: 115–125.
  40. Wee CY, Liu C, Lee A, et al. Cortical graph neural network for AD and MCI diagnosis and transfer learning across populations. *Neuroimage Clin* 2019; 23: 101929.
  41. Wen J, Samper-González J, Bottani S, et al. Reproducible evaluation of diffusion MRI features for automatic classification of patients with Alzheimer's disease. *Neuroinformatics* 2021; 19: 57–78.
  42. Kautzky A, Seiger R, Hahn A, et al. Prediction of autopsy verified neuropathological change of Alzheimer's disease using machine learning and MRI. *Front Aging Neurosci* 2018; 10: 06.
  43. Weiner MW, Veitch DP, Aisen PS, et al. Update of the Alzheimer's Disease Neuroimaging Initiative: a review of papers published since its inception. *Alzheimers Dement* 2015; 11: e1–120.
  44. Marcus DS, Fotenos AF, Csernansky JG, et al. Open access series of imaging studies: longitudinal MRI data in nondemented and demented older adults. *J Cogn Neurosci* 2010; 22: 2677–2684.
  45. Rockwood K, Strang D, MacKnight C, et al. Interrater reliability of the Clinical Dementia Rating in a multicenter trial. *J Am Geriatr Soc* 2000; 48: 558–559.
  46. Klein A, Andersson J, Ardekani BA, et al. Evaluation of 14 nonlinear deformation algorithms applied to human brain MRI registration. *Neuroimage* 2009; 46: 786–802.
  47. Suk H-I and Shen D. Deep learning-based feature representation for AD/MCI classification. In: Salinesi C, Norrie MC and Pastor Ó (eds) *Advanced information systems engineering*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, pp.583–590.
  48. Sarraf S, DeSouza DD, Anderson J, et al. Deep AD: Alzheimer's disease classification via deep convolutional neural networks using MRI and fMRI. *bioRxiv* 2017; doi: <https://doi.org/10.1101/070441> [Preprint]. Posted January 14, 2017.
  49. Korolev S, Safiullin A, Belyaev M, et al. Residual and plain convolutional neural networks for 3D brain MRI classification. Residual and plain convolutional neural networks for 3D brain MRI classification. In: *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*, Melbourne, VIC, Australia, 2017, pp.835–838.
  50. Smith SM. Fast robust automated brain extraction. *Hum Brain Mapp* 2002; 17: 143–155.
  51. Bishop CM. *Pattern recognition and machine learning*. New York: Springer-Verlag, 2006, pp 38–46.
  52. He K, Zhang X, Ren S, et al. Identity mappings in deep residual networks. In: Leibe B, Matas J and Sebe N, et al. (eds) *Computer vision – ECCV 2016. ECCV 2016. Lecture notes in computer science, vol 9908*. Cham: Springer, 2016, pp.630–645.
  53. Wu Y and He K. Group normalization. *arXiv* 2018. <https://doi.org/10.48550/arXiv.1803.08494>
  54. Dubois B, Feldman HH, Jacova C, et al. Advancing research diagnostic criteria for Alzheimer's disease: the IWG-2 criteria. *Lancet Neurol* 2014; 13: 614–629.
  55. Kapoor S and Narayanan A. Leakage and the reproducibility crisis in machine-learning-based science. *Patterns (NY)* 2023; 4: 100804.
  56. Qiu S, Miller MI, Joshi PS, et al. Multimodal deep learning for Alzheimer's disease dementia assessment. *Nat Commun* 2022; 13: 3404.
  57. Koepsell TD and Monsell SE. Reversion from mild cognitive impairment to normal or near-normal cognition: risk factors and prognosis. *Neurology* 2012; 79: 1591–1598.

58. Petersen RC, Smith GE, Waring SC, et al. Mild cognitive impairment: clinical characterization and outcome. *Arch Neurol* 1999; 56: 303–308.
59. Petersen RC, Doody R, Kurz A, et al. Current concepts in mild cognitive impairment. *Arch Neurol* 2001; 58: 1985–1992.
60. Grand JH, Caspar S and Macdonald SW. Clinical features and multidisciplinary approaches to dementia care. *J Multidiscip Healthc* 2011; 4: 125–147.
61. Mitchell AJ and Shiri-Feshki M. Rate of progression of mild cognitive impairment to dementia—meta-analysis of 41 robust inception cohort studies. *Acta Psychiatr Scand* 2009; 119: 252–265.