An End-to-End 3D ConvLSTM-based Framework for Early Diagnosis of Alzheimer's Disease from Full-Resolution Whole-Brain sMRI Scans

Selene Tomassini[†]*, Nicola Falcionelli[†], Paolo Sernani[†], Henning Müller[‡] and Aldo Franco Dragoni[†]
[†]Department of Information Engineering, Engineering Faculty, Marche Polytechnic University (UnivPM), Ancona, Italy
[‡]Information Systems Institute, University of Applied Sciences Western Switzerland (HES-SO), Sierre, Switzerland
*Corresponding author: Selene Tomassini (s.tomassini@pm.univpm.it)

***Abstract*—Alzheimer's Disease (AD) is the most prevailing form of dementia, killing more people than prostate and breast cancers combined. Structural Magnetic Resonance Imaging (sMRI) is widely used for the analysis of progressive brain aggravation and its clinical utility in discriminating AD is well established. Even if an effective cure does not exist yet, early detection is fundamental for slowing down the worsening of symptoms. Thus, the aim of the present work is to propose an end-to-end 3D Convolutional Long Short-Term Memory (ConvLSTM)-based framework for early diagnosis of AD from full-resolution whole-brain sMRI scans. The proposed framework was applied to 427 full-resolution whole-brain sMRI scans belonging to both OASIS and ADNI databases in order to provide a less dataset-specific approach. Results show that our framework is performing well in discriminating AD from Cognitively Normal (CN) patients, reaching a classification accuracy of 86%, sensitivity of 96%, f1-score of 88% and AUC of 93% on the test data. The tests were performed on a scalable GPU cloud service and are publicly available to guarantee reproducibility. Since the proposed framework performs well without domain-specific knowledge from AD as well as computationally-costly processes such as segmentation, it can be applied to other mental disorders using whole-brain sMRI scans as input data.**

***Index Terms*—Alzheimer's Disease, Deep Learning, Diagnosis, End-to-End Approach, Scalable GPU Cloud, Structural Magnetic Resonance Imaging, 3D Convolutional Long Short-Term Memory**

I. Introduction

Over 50 million people worldwide suffer from dementia and Alzheimer's Disease (AD) is the most prevailing form [1], [2], killing more people than prostate and breast cancers combined [3]. AD is a neurological brain disorder that begins with Mild Cognitive Impairment (MCI) and worsens progressively until death, as it ends up destroying the brain area that controls breathing and heart functions [4], [5].

Structural Magnetic Resonance Imaging (sMRI) is widely used for the analysis of progressive cognitive aggravation [6], as it offers a non-invasive and painless method of studying the anatomy of the brain and its structural changes using radio waves and a strong magnetic field [7], [8]. The clinical utility of sMRI in discriminating AD from other cognitive diseases is well established [1], [9].

Even if an effective cure does not exist and a final diagnosis of AD is difficult, early detection is fundamental for slowing down the worsening of symptoms [2], [4]. Through early detection, the quality of life of AD patients can also improve [6], [10]. For this reason, Machine Learning (ML) as well as Deep Learning (DL)-based computer-aided systems have been investigated to automatically diagnose AD [11]. Conventional ML methods require extensive pre-processing to obtain handcrafted features in combination with different types of classifiers such as Support Vector Machines (SVMs), which may increase both the computational time and the error rate [12]. Moreover, these technical choices may limit the flexibility of learning methods that cannot freely identify new features to improve the performance. DL-based methods are often achieving very good results, being better suited for generalizing even under subtle anatomical changes [9], [13]. The main differences from conventional ML methods are that DL may require minimal pre-processing and no feature engineering, resulting in a less bias-prone and more end-to-end process [9].

To the best of our knowledge, a study that leverages exclusively on a bidirectional 3D Convolutional Long Short-Term Memory (ConvLSTM) trained end-to-end on a scalable GPU cloud service to investigate the presence of AD is no yet existing. The aim of the present work is to propose an end-to-end 3D ConvLSTM-based framework for early diagnosis of AD from full-resolution whole-brain sMRI scans. The source code of the implementation is available on GitHub [1].

II. Related Work

The development of an algorithm able to automatically classify the brain anatomical changes caused by AD is of great interest to many research groups. Out of the large number of DL architectures, approaches mainly rely on Convolutional Neural Networks (CNNs) for detecting AD in MRI data [12]–[14]. 2D CNNs are widely used on 2D brain MRI slices to detect AD. Nevertheless, 2D CNNs take a single MRI slice as input, failing to leverage context from adjacent slices as

The research here presented was supported by the project "Using 3D CNNs for the Recognition of Lung Cancer Histotypes Directly from CT Scans" funded by CARIVERONA Foundation and by the project "Deep Learning for Early Medical Diagnosis: A Novel Methodology for Different Clinical Scenarios" funded by Department of Information Engineering, Engineering Faculty, Marche Polytechnic University, Ancona, Italy.

[1]https://github.com/airtlab/ConvLSTM4AD.

Published in Proceedings of the 2021 IEEE 34th International Symposium on Computer-Based Medical Systems (CBMS), Aveiro, Portugal; 7-9 June 2021, pp. 74-78, which should be cited to refer to this work.
DOI: 10.1109/CBMS52027.2021.00081

the brain anatomy remains unused. Voxel information from adjacent slices can be very useful. For this reason, 3D CNNs operate in the 3D space as they can examine groups of images as a whole, having the power to catch the anatomical changes of the brain [1]. Both 3D CNN and 3D ConvLSTM use 3D convolution operations in their inner structure [1]. Since they rely on different convolution mechanisms, the latter is able to capture more discriminative features [1], [8].

Focusing exclusively on patient-level classification, meaning that the scan is used at once and the classification is performed at the patient level, Luo et al. [11] provided an automatic AD detection system based on brain sMRI scans taken from the ADNI 1 dataset of the Alzheimer's Disease Neuroimaging Initiative (ADNI) database. They trained a 3D CNN-based model, gaining a good AD detection with a sensitivity of 100% and a specificity of 93%. Korolev et al. [15] compared two approaches, performing their experiments on whole-brain sMRI scans belonging to the ADNI database. In detecting AD from cognitively healthy patients, the model they called VoxCNN reached an accuracy of 79% and an Area Under the receiver operating characteristic Curve (AUC) of 88%. Ullah et al. [3] came up with a 3D CNN-based model to detect AD and dementia from whole-brain sMRI scans selected from the Open Access Series of Imaging Studies (OASIS) database. They achieved an accuracy of 80.25%. Basaia et al. [9] presented a 3D CNN-based model leveraged on a single cross-sectional brain sMRI scan belonging to the ADNI database. In AD versus healthy patients classification, their method reached an accuracy of 99.2%, sensitivity of 98.9% and specificity of 99.5%. More recently, Xia et al. [1] proposed a hybrid framework for AD detection, where both 3D CNN and 3D ConvLSTM were employed on whole-brain sMRI scans taken from the ADNI 1 dataset. In detecting AD from cognitively healthy patients, the best model achieved an accuracy of 94.19%, sensitivity of 93.75%, specificity of 94.57% and AUC of 96%.

III. Proposed Framework

Our framework (Fig. 1) is composed of necessary pre-processing steps followed by training/classification. In particular, the minimal pre-processing does not affect the end-to-end philosophy of our method, as better explained in subsection III-B.

A. Data Collection

In our data collection process, we considered two datasets, OASIS-3 and ADNI 1-Screening, belonging to the OASIS [2] and ADNI [3] databases.

From OASIS-3, we selected one raw axial T1-weighted (T1w) sMRI scan for each patient, resulting in 275 (130 CN and 145 AD) scans, each having an original shape of 176x256x256, 1-mm pixel size and 1-mm slice thickness. In case of multiple scans per patient, we dwelt on the scan done first in order to avoid an intra-patient bias. All selected sMRI scans belong to anonymized patients from 42 to 97 years old and were done with 3 Siemens scanners [16]. We stored them as NIFTI files.

For ADNI 1, we focused on patients who had the first screening scan (ADNI 1-Screening). We selected one raw axial T1w sMRI scan for each patient, resulting in 152 (84 CN and 68 AD) scans, each having an original shape of 256x256x5, 1 or 1.2-mm pixel size and 5-mm slice thickness. All selected sMRI scans belong to anonymized patients from 55 to 90 years old and were done with 2 GE Medical Systems scanners. We stored them as NIFTI files.

B. Data Preparation

Brain sMRI scans capture extraneous parts of the head, which may interfere as noise. To overcome this problem, some studies use the FMRIB Software Library (FSL) [4] for brain MRI pre-processing [10].

First, we extracted the brain, which is one of the most important pre-processing steps for eliminating non-brain tissue, requiring no domain-specific knowledge [14]. To do so, we removed all non-brain tissue from brain sMRI data using the Brain Extraction Tool (BET) of the FSL by setting the fractional intensity threshold to 0.3 in order to reduce the bias without discarding brain voxels.

After brain extraction, we performed registration in order to have consistent dimensions for both datasets. Registration is fundamental to align multiple images for verifying their spatial correlation in terms of anatomy [14]. By registering sMRI images to a unique template, we expect similar structures to be roughly in the same spatial location, allowing the network to identify the most relevant features. As reported in subsection III-A, OASIS-3 and ADNI 1-Screening have different original shapes. To overcome this dimensionality challenge and make data homogeneous, we performed registration on the OASIS-3 dataset taking as reference dimension one of the ADNI 1-Screening series, as shown in Fig. 2.

After registration, we merged all brain sMRI scans belonging to both datasets into a dimensionally-uniform dataset. The merged dataset has a total number of 427 whole-brain sMRI scans (214 CN and 213 AD patients), each having a post-registration shape of 256x256x5.

We then resampled all sMRI data to 1-mm pixel size and slice thickness, and performed input normalization. Both resampling and input normalization lead to an easier comparison in order to interpret data onto a common shape and size [14].

Eventually, we labelled these full-resolution whole-brain sMRI data in AD and CN patients and randomly split them into training, validation and test sets with a ratio of 6:2:2.

C. Model Architecture

In designing the architecture of the proposed neural network, we found the best trade-off between speed, complexity

[2]https://www.oasis-brains.org/.

[3]http://adni.loni.usc.edu/. The investigators within ADNI provided data but did not participate in analysis and writing of this paper. 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.

[4]https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/FSL.

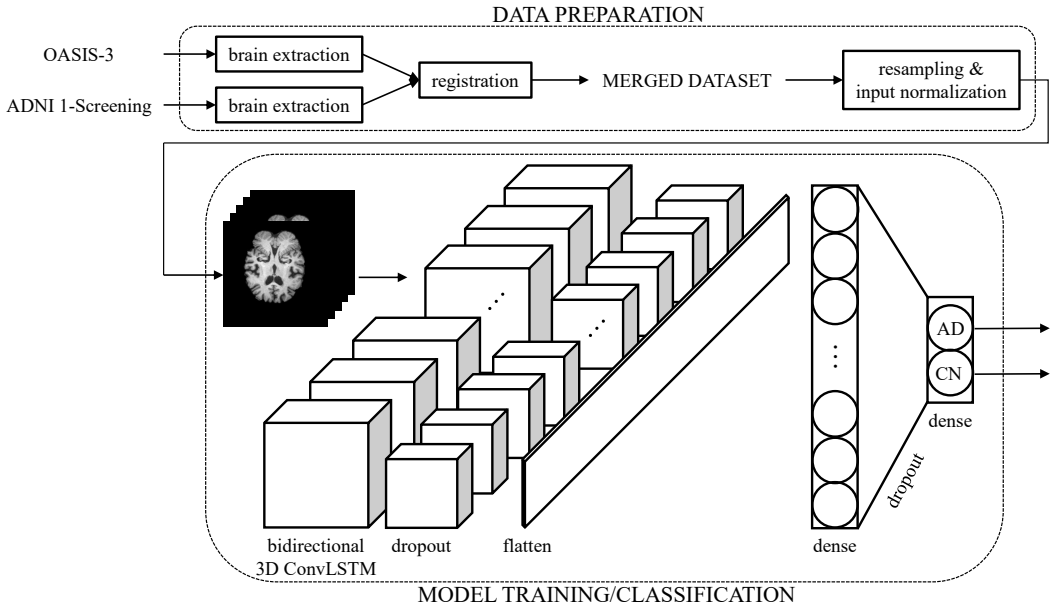

Figure 1: Representation of the proposed framework composed of necessary pre-processing steps followed by training/classification. In the output layer of the neural network, AD and CN mean Alzheimer's Disease and Cognitively Normal patients, respectively.

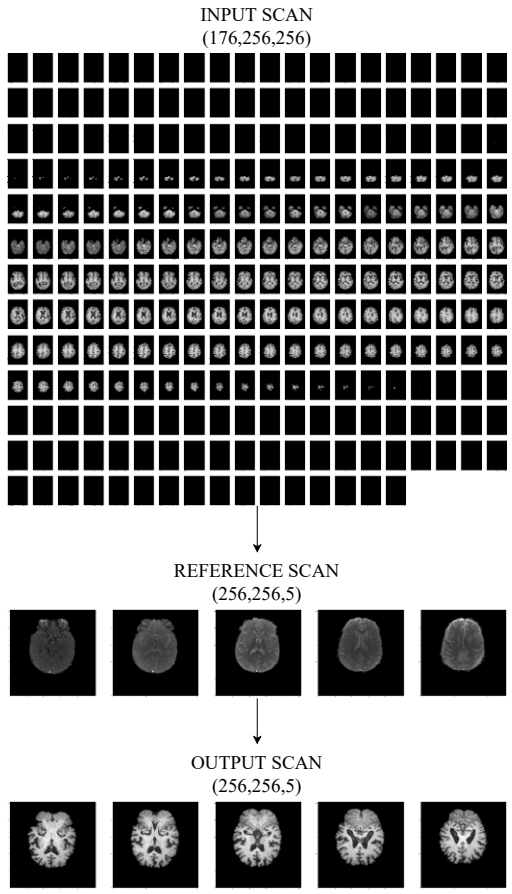

Figure 2: Registration performed on one whole-brain sMRI scan of the OASIS-3 dataset, taking a whole-brain sMRI scan of the ADNI 1-Screening dataset as reference.

and performance. In this regard, we designed a sequential model with 6 layers.

The first layer is a bidirectional 3D ConvLSTM layer that is the backbone of our model. It has 32 convolutional filters, each having a size of 3. We chose a ConvLSTM as both spatial and temporal features take part in the classification [8]. ConvLSTM is a special Recurrent Neural Network (RNN) that uses convolution filters, thus being not only able to model the long-term interactions but also to explore the spatial information [1]. The relevant features are automatically extracted from whole-brain sMRI data through convolution operations. During the convolution, the network learns which filters activate when seeing a specific type of feature at a spatial position in the input. The LSTM architecture allows each sequence to be treated in its entirety, remembering information selectively and applying the information of the already processed frames in the processing of subsequent ones. LSTM, indeed, comprises three gates, as shown in Fig. 3: the input gate transmits new input information into the cell; the forget gate selectively forgets information; the output gate stores relevant information [1], [8]. The cell itself decides when to allow the reading and updating of the information via the three gates [14].

The second layer is a Dropout layer with a rate of 50% to limit overfitting. Dropout is a widely-used regularization technique that randomly eliminates a few units from the network (in this case 50%) during training, thus reducing the overall model complexity.

The third layer is a Flatten layer that flattens all the extracted features into a mono-dimensional tensor.

The fourth layer is a Dense layer. A Rectified Linear Unit (ReLU) activation function is used to help the model consider non-linear effects and interactions, as it demonstrates faster

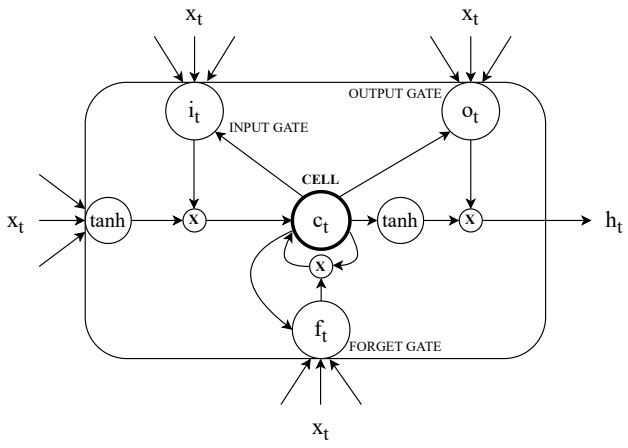

Figure 3: The input, output and forget gates of a LSTM unit.

training as well as better results than sigmoid function [17].

The fifth layer is another Dropout layer with a rate of 50% for further reducing the complexity of the model.

The sixth layer is another Dense layer with a Softmax activation function that assigns probabilities to each class by having as outputs real values between 0 and 1 (with sum 1).

D. Model Training

To train our model, we chose the parameter combination that led to the best performance on the validation data.

Selecting the GPU hardware acceleration and high RAM setups of the PRO version of Google Colab cloud service [5], we used Keras to train our model end-to-end for 40 epochs with a mini-batch size of 4 and a learning rate of 0.001 using the Stochastic Gradient Descent (SGD) optimizer. During training, the optimization based on a stochastic gradient is fundamental to minimize the loss function while ensuring better efficiency [17]. We set categorical cross entropy as loss function. We used early stopping callback with a patience of 7 to further reduce overfitting, as it stops the training at the point where the validation loss is minimal, which in our case coincided with the 12th epoch.

We chose to save the model with the lowest validation loss instead of the one with the highest validation accuracy, as the validation loss captures exactly the divergence between the predicted output and the ground truth [10]. Eventually, we used it to calculate the performance on the test set.

IV. Results

We performed a patient-level classification, in which AD is detected from CN. The primary evaluation metric we considered is classification accuracy on the test data. We also considered specificity, sensitivity, f1-score and AUC, as it is independent of the threshold chosen for classification. The sensitivity, in particular, is extremely important in the biomedical context as a misclassified pathology is the major issue in medical diagnosis.

[5]https://colab.research.google.com/notebooks/intro.ipynb.

Table I summarizes the performance for AD vs. CN of the proposed method and of the state-of-the-art methods cited in section II in terms of accuracy, specificity, sensitivity, f1-score and AUC. It also reports the total number of patients involved, the database/s to which the data belong, the whole-brain sMRI scan shape and the 3D architecture used.

V. Discussion and Conclusions

Since timely and accurate diagnosis of AD is crucial for early treatment [2], [4], much research has been conducted to find ways for automatically detecting the patterns of brain cell degradation from MRI scans early in the process [12]–[14].

The aim of the present study is to propose an end-to-end 3D ConvLSTM-based framework for early diagnosis of AD from full-resolution whole-brain sMRI scans using a scalable GPU cloud service to ensure its reproducibility. For this aim, we built a framework composed of only necessary pre-processing steps followed by training/classification, which resulted in the development of an end-to-end process. Pre-processing consists of brain extraction, registration, merging of whole-brain sMRI data originating from OASIS-3 and ADNI 1-Screening in a dimensionally-uniform dataset, resampling and input normalization. Classification comprises the end-to-end training, validation and test of full-resolution whole-brain sMRI scans, selecting the GPU hardware acceleration and high RAM setups of the PRO version of Google Colab cloud service and exploiting an architecture whose core layer is a bidirectional 3D ConvLSTM. Results show that the proposed methodology, that first exploits exclusively a 3D ConvLSTM-based architecture to classify samples from both OASIS and ADNI databases, is performing well in discriminating AD from CN patients, as reported in Table I. It reaches an accuracy of 86%, specificity of 74%, sensitivity of 96%, f1-score of 88% and AUC of 93% on the test data. The latter value, in particular, means that the classifier achieves a high performance in separating the samples belonging to AD from CN [14].

The strengths of the proposed method with respect to other DL-based studies in AD detection are multiple. In our study, the model was trained, validated and tested on data belonging to two openly-accessible databases in order to provide a less dataset-specific approach. We performed only minimal pre-processing thanks to the ability of 3D ConvLSTM to automatically select visual features, thus preserving the whole brain anatomy. Since 3D ConvLSTM does not require an input tensor of standard dimensions, we kept the full resolution of whole-brain sMRI scans. We also trained our model completely end-to-end on a scalable GPU cloud service, allowing an easy reproducibility of the entire framework. Moreover, since the proposed framework performs well with no domain-specific knowledge from AD as well as computationally-costly processes such as segmentation, we believe it can be applied to other mental disorders using whole-brain sMRI scans as input data.

One possible limitation relies on the number of slices per scan, which corresponds to 5. Among ADNI 1 subsets, we

Table I: Performance in classifying AD vs. CN of the proposed method and of state-of-the-art methods doing a patient-level classification on whole-brain sMRI data. Values of ACCuracy, SPEcificity, SENsitivity, F1-Score and AUC are given in percentage (%). Number of patients (AD and CN) involved in the evaluation, database/s (OASIS or/and ADNI) to which the data belong, whole-brain sMRI scan shape and 3D architecture used are also reported.

Method	Patients	Database	Shape	Architecture	ACC	SPE	SEN	F1-S	AUC
Luo et al. [11]	81	ADNI	(56,56,5)	CNN	-	93	100	-	-
Korolev et al. [15]	111	ADNI	(110,110,110)	CNN	79	-	-	-	88
Ullah et al. [3]	416	OASIS	-	CNN	80.25	-	-	-	-
Basaia et al. [9]	646	ADNI	-	CNN	99.2	99.5	98.9	-	-
Xia et al. [1]	427	ADNI	(119,119,143)	CNN + ConvLSTM	94.19	94.57	93.75	-	96
Proposed	427	OASIS + ADNI	(256,256,5)	ConvLSTM	86	74	96	88	93

focused on ADNI 1-Screening as we liked to take the scan related to the patient's first visit (screening) because AD detection in the first stages is even more challenging. ADNI 1-Screening, however, has 5 slices per scan. In order to make data homogeneous, we needed 5 slices also for OASIS-3, as explained in the registration step (subsection III-B). We strongly believe that a higher number of slices per scan could lead to even better results.

Our framework is still under development and as part of future work we plan to make the classification results interpretable by means of dedicated visualization techniques [18] in order to understand which brain regions are more involved in the decision-making process. We also plan to increase the cardinality of the samples by looking for other openly-accessible repositories or applying data augmentation techniques, such as horizontal flipping to exchange left and right sides of the brain, as AD affects both hemispheres [1]. Furthermore, it would be interesting to introduce metadata associated to each patient in order to cross-check the results.

To conclude, in this study it was demonstrated that the proposed end-to-end 3D ConvLSTM-based framework represents a promising and easily-reproducible method for early diagnosis of AD from full-resolution whole-brain sMRI scans.

ACKNOWLEDGMENT

The authors would like to thank Alessandro Quercetti and Lorenza Gianvittorio for their contribution in developing the proposed framework.

REFERENCES

[1] Z. Xia, G. Yue, Y. Xu, C. Feng, M. Yang, T. Wang, and B. Lei, "A novel end-to-end hybrid network for Alzheimer's disease detection using 3D CNN and 3D CLSTM," in *2020 IEEE 17th International Symposium on Biomedical Imaging*. IEEE, 2020, pp. 1–4.

[2] O. B. Ahmed, S. Fezzani, C. Guillevin, L. Fezai, M. Naudin, B. Gianelli, and C. Fernandez-Maloigne, "Deepmrs: An end-to-end deep neural network for dementia disease detection using MRS data," in *2020 IEEE 17th International Symposium on Biomedical Imaging*. IEEE, 2020, pp. 1459–1463.

[3] H. T. Ullah, Z. A. Onik, R. Islam, and D. Nandi, "Alzheimer's disease and dementia detection from 3D brain MRI data using deep convolutional neural networks," in *2018 IEEE 3rd International Conference for Convergence in Technology*. IEEE, 2018, pp. 1–3.

[4] J. Sun, S. Yan, C. Song, and B. Han, "Dual-functional neural network for bilateral hippocampi segmentation and diagnosis of Alzheimer's disease," *International Journal of Computer Assisted Radiology and Surgery*, vol. 15, no. 3, pp. 445–455, 2020.

[5] J. Islam and Y. Zhang, "Early diagnosis of Alzheimer's disease: A neuroimaging study with deep learning architectures," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 1881–1883.

[6] S. Afzal, M. Maqsood, F. Nazir, U. Khan, F. Aadil, K. M. Awan, I. Mehmood, and O.-Y. Song, "A data augmentation-based framework to handle class imbalance problem for Alzheimer's stage detection," *IEEE Access*, vol. 7, pp. 115 528–115 539, 2019.

[7] A. Farooq, S. Anwar, M. Awais, and S. Rehman, "A deep CNN based multi-class classification of Alzheimer's disease using MRI," in *2017 IEEE International Conference on Imaging Systems and Techniques*. IEEE, 2017, pp. 1–6.

[8] S. V. Yakkundi and D. Subha, "Convolutional LSTM: A deep learning approach for dynamic MRI reconstruction," in *2020 IEEE 4th International Conference on Trends in Electronics and Informatics*. IEEE, 2020, pp. 1011–1015.

[9] S. Basaia, F. Agosta, L. Wagner, E. Canu, G. Magnani, R. Santangelo, M. Filippi, A. D. N. Initiative *et al.*, "Automated classification of Alzheimer's disease and mild cognitive impairment using a single MRI and deep neural networks," *NeuroImage: Clinical*, vol. 21, p. 101645, 2019.

[10] Z. Guan, R. Kumar, Y. R. Fung, Y. Wu, and M. Fiterau, "A comprehensive study of Alzheimer's disease classification using convolutional neural networks," *arXiv preprint arXiv:1904.07950*, 2019.

[11] S. Luo, X. Li, and J. Li, "Automatic Alzheimer's disease recognition from MRI data using deep learning method," *Journal of Applied Mathematics and Physics*, vol. 5, no. 9, pp. 1892–1898, 2017.

[12] J. Wen, E. Thibeau-Sutre, M. Diaz-Melo, J. Samper-González, A. Routier, S. Bottani, D. Dormont, S. Durrleman, N. Burgos, O. Colliot *et al.*, "Convolutional neural networks for classification of Alzheimer's disease: Overview and reproducible evaluation," *Medical Image Analysis*, vol. 63, p. 101694, 2020.

[13] N. Yamanakkanavar, J. Y. Choi, and B. Lee, "MRI segmentation and classification of human brain using deep learning for diagnosis of Alzheimer's disease: A survey," *Sensors*, vol. 20, no. 11, p. 3243, 2020.

[14] M. B. T. Noor, N. Z. Zenia, M. S. Kaiser, S. Al Mamun, and M. Mahmud, "Application of deep learning in detecting neurological disorders from magnetic resonance images: A survey on the detection of Alzheimer's disease, Parkinson's disease and schizophrenia," *Brain Informatics*, vol. 7, no. 1, pp. 1–21, 2020.

[15] S. Korolev, A. Safiullin, M. Belyaev, and Y. Dodonova, "Residual and plain convolutional neural networks for 3D brain MRI classification," in *2017 IEEE 14th International Symposium on Biomedical Imaging*. IEEE, 2017, pp. 835–838.

[16] P. J. LaMontagne, T. L. Benzinger, J. C. Morris, S. Keefe, R. Hornbeck, C. Xiong, E. Grant, J. Hassenstab, K. Moulder, A. Vlassenko *et al.*, "OASIS-3: Longitudinal neuroimaging, clinical, and cognitive dataset for normal aging and Alzheimer disease," *MedRxiv*, 2019.

[17] H. Allioui, M. Sadgal, and A. Elfazziki, "Deep MRI segmentation: A convolutional method applied to Alzheimer disease detection," *International Journal of Advanced Computer Science and Applications*, vol. 10, pp. 365–371, 2019.

[18] M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional networks," in *European Conference on Computer Vision*. Springer, 2014, pp. 818–833.