Longitudinal Prediction Modeling of Alzheimer Disease using Recurrent Neural Networks

Solale Tabarestani[1], Maryamossadat Aghili[1,2], Mehdi Shojaie[1], Christian Freytes[1], Mercedes Cabrerizo[1], Armando Barreto[1], Naphtali Rishe[2], Rosie E. Curiel[3,5], David Loewenstein[3,5], Ranjan Duara[4,5], and Malek Adjouadi[1,5]

[1] *Center for Advanced Technology and Education, Department of Electrical and Computer Engineering, Florida International University, Miami 33174, FL, USA. e-mail: staba006@fiu.edu*

[2] *School of Computing and Information Sciences, Florida International University, Miami 33174, FL, USA.*

[3] *Center for Cognitive Neuroscience and Aging, University of Miami Miller School of Medicine, Miami, FL, USA.*

[4] *Wien Center for Alzheimer's Disease & Memory Disorders, Mount Sinai Medical Center, Miami, FL, USA.*

[5] *1Florida ADRC (Florida Alzheimer's Disease Research Center at Gainesville. e-mail: madjouadi@fiu.edu*

***Abstract*—This paper proposes an implementation of Recurrent Neural Networks (RNNs) for (a) predicting future Mini-Mental State Examination (MMSE) scores in a longitudinal study and (b) deploying a multiclass multimodal neuroimaging classification process that involves three different known stages of Alzheimer's progression, cognitively normal (CN), Mild Cognitive Impairment (MCI) and Alzheimer's Disease (AD). This multimodal data is fed into two well-studied variations of the RNNs; Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU). The accuracy, F-score, sensitivity, and specificity of the models are reported for the classification task as well as the root mean square error (RMSE) and correlation coefficient for the regression task. The results demonstrate the superiority of the proposed model over state-of-the-art classification and regression techniques of Support Vector Machine (SVM), Support Vector Regression (SVR) and Ridge Regression.**

***Keywords*—Long Short-Term Memory (LSTM), Gated Recurrent Unit (GRU), Classification, Regression, Alzheimer, Prognosis, Diagnosis, multimodal, Neuroimaging, Longitudinal.**

I. Introduction

Alzheimer's disease is an irreversible neurodegenerative disorder that impairs memory, cognitive abilities and behavior [1]. The complex nature of AD biomarkers and the heterogeneity of measurements obtained from various imaging modalities are some of the obstacles faced in seeking effective early detection and planning therapeutic protocols [2].

In addressing the barriers impeding AD research, scientists have proposed statistical and machine learning techniques for robust diagnosis. Until recently, most efforts were dedicated to modeling the disease at a single time point using cross-sectional datasets [3], [4]. However, these approaches could not provide enough information about the future status of patients. At later stages of AD, where the brain has already suffered from atrophy, treatment would too late to be effective. Early diagnosis of the disease allows for early intervention and facilitates development of effective healthcare services. This initiates a new line of research aiming at enhancing the effect of treatment by predicting the onset of the disease before the occurrence of acute neurodegeneration. The objective of these studies is to leverage temporal information from longitudinal data to model the progression of AD. Multiple classification and regression models have been proposed to predict disease progression and level of disease severity. The feature space is either based on the information available at baseline or a concatenation of features from multiple previous time points [5], [6], [7], [8]. The integration of features into a single observation window creates a high dimensional input space which is not only difficult to deal with, but also disregards temporal connections between consecutive time points [9], [10]. With the gradual nature of AD progression, these methods could not efficiently exploit the longitudinal information.

Recurrent Neural Networks (RNNs), introduced in 1986, recently gained popularity due to the intrinsic power in learning long-short term dependencies of sequenced data. These networks share information between series of data points through an additional hidden set of parameters. RNNs are now being implemented in modeling the progression patterns of chronic diseases [11], [12]. In [10], Nguyen et al. trained an RNN-LSTM network over a seven-year period to predict multiple AD biomarkers for one subsequent time point. In another study, Wang et al. applied an RNN architecture with LSTM cells to predict global staging of the Clinical Dementia Rating (CDR) score of the next visit using previous records [13]. Aghili et al. utilized LSTM and GRU models to classify AD subjects using longitudinal records of data over an 11-year period [14].

Using the inherent correlations of sequential data, RNNs proved their potential in predicting AD related biomarkers for a future time point. Although effective, these studies limit themselves to predicting at only a single future interval. This paper broadened the scope and application of the RNNs by predicting the progression of AD over multiple future time points simultaneously. Employing three records of data for each subject, the RNN surpassed other machine learning methods not only in estimating the categorical variable for a multiclass classification task, but also in assessing the numerical value of the AD biomarker. Furthermore, two variations of RNN, GRU and LSTM, are investigated for the challenging task of drawing the delineation boundary of subjects in a multiclass classification scenario and also for predicting the trajectories of cognitive scores for the next two years.

The rest of the paper is structured as follows. Section II begins with a brief description of the data used and the

978-1-7281-0848-3/19/$31.00 ©2019 IEEE

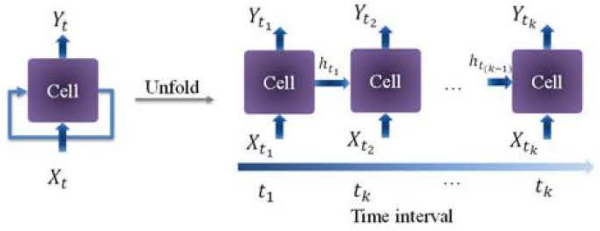

Fig.1. Recurrent Neural Network architecture.

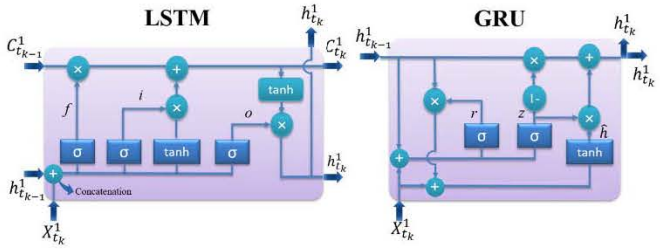

Fig.2. The structure of LSTM and GRU cells

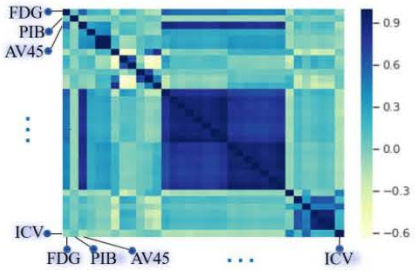

Fig.3. Heat-map of features used in this study

preprocessing steps undertaken. Then, the different variations of RNN are introduced and the proposed methodology is presented. Section III reports on experiments that were conducted and discusses the results obtained, with section IV providing the conclusion.

II. Methodology

A. Recurrent Neural Network (RNN)

Processing sequences of data, RNNs have the capability to effectively incorporate temporal dependencies in longitudinal data. Fig. 1 illustrates an RNN with data sequences of k time steps. At each time point (t_i), besides the input features (X_{t_i}), the internal state (memory) of the cell from the previous time step ($h_{t_{(i-1)}}$) are fed to the cell. Thus, unlike feedforward neural networks, RNNs can identify patterns hidden in sequences of data. However, due to a lack of long-term memory in basic RNNs, each time point is mainly affected by previous intervals in close vicinity. Therefore, they are not capable of leveraging long-term relationships in historical data and older information tends to fade away. This setback is known as "vanishing gradient" in which the network gradually forgets older traces.

To address this issue, GRU and LSTM-based RNN architectures with the capability of capturing long-term memories have been proposed [15], [16]. The structure of LSTM and GRU cells as the building blocks of improved version of RNN are shown in Fig. 2. In an LSTM cell, three gates denoted by sigmoid functions (σ), decide whether the previous cell state (C), the input (X), and the output (h) need to be passed to the next time step. This will make the memorizing capability of the cell more intelligent and durable. The following equations describe the operation principle of an LSTM cell.

$$\begin{aligned} f_{t_k} &= \sigma\left(W_f\left(X_{t_k}, h_{t_{k-1}}\right) + b_f\right) \\ i_{t_k} &= \sigma\left(W_i\left(X_{t_k}, h_{t_{k-1}}\right) + b_i\right) \\ \hat{\imath}_{t_k} &= tanh\left(W_{\hat{\imath}}\left(X_{t_k}, h_{t_{k-1}}\right) + b_{\hat{\imath}}\right) \\ C_{t_k} &= C_{t_{k-1}} * f_{t_k} + \hat{\imath}_{t_k} * i_{t_k} \\ o_{t_k} &= \sigma\left(W_o\left(X_{t_k}, h_{t_{k-1}}\right) + b_o\right) \\ h_{t_k} &= o_{t_k} * \tanh(C_{t_k}) \end{aligned} \tag{1}$$

where t_k refers to the k^{th} time step; X_{t_k}, C_{t_k}, and h_{t_k} represent the input, state, and output of the cell at the k^{th} time step; and f_{t_k}, i_{t_k}, and o_{t_k} are the outputs of the forget, input, and output gates. Also, W and b are the weights of the neural networks.

In the gating mechanism of GRU, two gates known as reset and update gates determine the amount of the current input and output of the previous time step that needs to be preserved. With the same notations of X_{t_k} and h_{t_k} as the input and output of the cell for the k^{th} time step, the mathematical equations of a GRU cell are summarized as follows.

$$\begin{aligned} z_{t_k} &= \sigma\left(W_z\left(X_{t_k}, h_{t_{k-1}}\right) + b_z\right) \\ r_{t_k} &= \sigma\left(W_r\left(X_{t_k}, h_{t_{k-1}}\right) + b_r\right) \\ \hat{h}_{t_k} &= tanh\left(W_{\hat{h}}\left(X_{t_k}, r_{t_k} * h_{t_{k-1}}\right) + b_{\hat{h}}\right) \\ h_{t_k} &= \left(1 - z_{t_k}\right) * h_{t_{k-1}} + z_{t_k} * \hat{h}_{t_k} \\ h_{t_k} &= o_{t_k} * \tanh(C_{t_k}) \end{aligned} \tag{2}$$

where z_{t_k} and r_{t_k} are the outputs of the update and reset gates.

B. Feature Selection

Referring to previous studies [14], which shed light on the possible overfitting of RNNs on the original feature space, feature analysis and ranking has been performed on the data. Consequently, to address the highly correlated features, *L1* feature selection was employed to extract the most important features. Using *L1* method, 25 features with highest variance in the feature space have been selected. The correlation matrix (heat map) of the features is illustrated in Fig 3.

C. Longitudinal AD Prediction using RNN

The proposed framework uses the memorization capability of the LSTM/GRU cell to capture historical dependencies from three records of subjects in order to predict the progression of AD at three next future time points. Therefore, a many-to-many RNN architecture with LSTM/GRU cells has been developed to carry out two tasks of longitudinal multiclass classification and regression.

The structure of the network for the LSTM case is demonstrated in Fig. 4. In the developed network, the three inputs (X_{t_1}, X_{t_2} *and* X_{t_3}) represent the feature space associated with three-time points of M_0 (Baseline), M_6 (after 6 months), and M_{12} (after 12 months). The information is transferred from one time point to the next one using the cell state (C) and output (Y). The outputs Y_{t_i} are the Mini-Mental State Examination

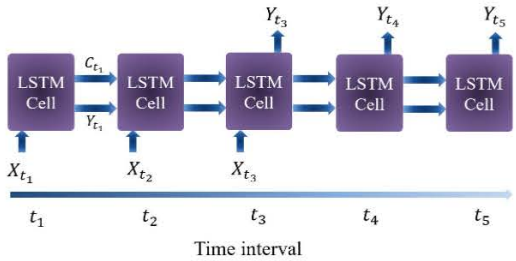

Fig. 4. The RNN architecture used to predict the progression of AD using historical data

TABLE I. STATISTICS OF THE DATASET USED IN THIS STUDY

Category	Subjects (f/m)	Age	Education(y)	MMSE
AD	336 (150/186)	74.93±7.81	15.17 ±2.99	23.18 ± .06
MCI	864 (354/510)	73.03±7.60	15.91±2.85	27.59±1.81
CN	521 (268/253)	74.25±5.79	16.37±2.70	29.06±1.14

(MMSE) score for regression model or status of patients (CN, MCI, and AD) for classification model. The time steps t_4 and t_5 are associated with the future time points M_{24} (24 months after the baseline) and M_{36} (36 months after baseline). The next section discusses the material and experimental results.

III. EXPERIMENTAL SETUPS, RESULTS AND DISCUSSIONS

A. Data

The data used in this study is obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (http://adni.loni.usc.edu/). ADNI was launched in 2003 as a public-private partnership, led by Principal Investigator Michael W. Weiner, MD. The primary goal of ADNI has been to test whether structural Magnetic Resonance Imaging (MRI), Positron Emission Tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of MCI and early AD.

Longitudinal medical records from 1458 subjects (341 CN, 255 EMCI, 529 LMCI, and 333 AD) have been incorporated in this dataset. During an 11-year study, each patient has been recalled for a follow-up visit every six months. These subjects have undergone several medical screening tests including MRI, PET, genetic tests, CSF tests, and cognitive impairment assessments. At each visit, an expert monitors the test results and updates the diagnosis for the participants. This categorical diagnosis (AD, MCI, NC) is used as the label for the multiclass classification experiment proposed in this study and the numerical Mini-Mental State Examination (MMSE) scores, an indicator of the AD cognitive impairment, with a range of 0-30 is adopted for the regression experiment. Characteristics of the dataset used in this study are summarized in Table I.

B. Longitudinal Data Preprocessing

Initially, the data is preprocessed to alleviate any inconsistencies caused by utilizing different data modalities and various protocols. Subjects who have participated at all five consecutive intervals including baseline, six months after the first visit (M_{06}), twelve months after the first visit (M_{12}), twenty-four months after the first visit (M_{24}) and thirty-six months after the first visit (M_{36}) have been considered. In the initial step of the experiments, data cleaning [17], [18], mean centering, data normalization, missing feature handling, and univariate feature analysis has been performed to discard uninformative features. Furthermore, subjects whose medical diagnosis are not reported are removed from further analysis.

TABLE II. SUMMARY OF MULTIMODAL FEATURES UTILIZED IN THIS STUDY

Source	Features
Cognitive tests	Everyday Cognition (ECog) questionnaire measurements, FAQ, MOCA, RAVLT, CDRSB
MRI	Ventricular volume, Hippocampus volume, Whole Brain volume, Entorhinal Cortical thickness, Fusiform, Middle temporal gyrus, ICV
PET	FDG, PIB amyloid, AV45 amyloid
Genetic	APOE4
Demographic	Age, Gender, Education
CSF	Amyloid Beta, Phosphorylated Tau, Total Tau

TABLE III. REGRESSION RESULTS

Algorithm	M12		M24		M36		Total
	RMSE	Corr	RMSE	Corr	RMSE	Corr	MSE
Ridge	2.07	0.58	2.66	0.62	2.99	0.63	6.82
SVR	2.14	0.59	2.86	0.61	3.17	0.58	7.68
LSTM	1.97	**0.63**	2.33	0.69	2.54	**0.72**	5.26
GRU	1.97	**0.63**	2.33	0.69	2.54	**0.72**	5.24
Ridge + FS*	2.02	0.62	2.67	0.65	2.93	0.65	6.65
SVR + FS*	2.16	0.60	2.76	0.65	3.26	0.62	7.70
LSTM + FS*	1.85	**0.63**	2.25	0.70	2.48	0.70	4.98
GRU + FS*	**1.82**	**0.63**	**2.21**	**0.71**	**2.44**	0.70	**4.77**

*Feature selection

C. *Simulation* and Results

This study evaluates the performance of two RNN variations, LSTM and GRU, on the ADNI cohort for the two tasks of classification and regression. The experiment proceeds with the selection of historical records from subjects at three intervals (baseline, M_{06}, and M_{12}) to predict the status of the subjects in three future time points of M_{12}, M_{24} and M_{36}. Estimating the MMSE scores of subjects is pursued as a regression problem and predicting the diagnosis labels is defined as multiclass classification problem. The data has been split randomly to a 75% training set, a 10% validation set, and a 15% testing set. Grid search has been utilized to select the best hyperparameters for regression and classification networks separately. In order to feed the longitudinal feature space into the RNNs, the data has been framed in the tensor form of [samples, time steps, features] which in this case is 3-time steps of the 532 samples with 34 features involving MRI, PET, Cerebrospinal fluid (CSF) and cognitive test scores as provided in Table II.

The performance of LSTM and GRU, implemented using the Keras deep learning library, are compared with state-of-the-art methods. It is worth noting that conventional methods cannot incorporate historical records of subjects for enhancing the prediction accuracy. This limitation has been compensated by concatenating all three-historical feature sets. Competing methods are then trained on this new feature space to find an individual direct map between the feature space from past intervals with the corresponding future time points.

As regression, RMSE and R-Correlation factor are used as evaluation metrics to compare Ridge and SVR from *Scikit*-learn library with LSTM and GRU, and the Results are reported in Table III.

TABLE IV. CLASSIFICATION RESULTS

Method	M12				M24				M36			
	ACC	PRE	REC	F1	ACC	PRE	REC	F1	ACC	PRE	REC	F1
SVM	0.66±0.04	0.44±0.05	0.66±0.05	0.52±0.04	0.61±0.04	0.38±0.04	0.61±0.04	0.46±0.04	0.61±0.03	0.38±0.04	0.61±0.03	0.48±0.04
LSTM	0.84±0.10	0.86±0.06	0.84±0.10	0.81±0.16	0.82±0.12	0.77±0.22	0.82±0.12	0.79±0.18	0.80±0.09	0.84±0.06	0.80±0.09	0.78±0.15
GRU	0.61±0.09	**0.95±0.00**	0.60±0.09	0.74±0.07	0.37±0.06	**0.99±0.00**	0.37±0.06	0.53±0.06	0.61±0.04	**0.98±0.00**	0.61±0.04	0.75±0.03
LSTM + FS	**0.88±0.03**	0.89±0.02	**0.90±0.02**	**0.89±0.02**	**0.87±0.01**	0.86±0.04	**0.87±0.02**	**0.86±0.02**	**0.88±0.02**	0.87±0.03	**0.88±0.02**	**0.87±0.03**
GRU + FS	0.68±0.09	**0.95±0.00**	0.68±0.09	0.79±0.07	0.28±0.11	**0.99±0.00**	0.29±0.11	0.43±0.13	0.51±0.08	**0.98±0.00**	0.51±0.08	0.67±0.04

Similarly, the classification problem is defined as the diagnosis of subjects at three future time points based on three previous intervals. For the classification task, SVM from Scikit-learn library is selected as the competitive alternative to evaluate the performance of the LSTM and GRU. F-score, precision, recall, accuracy has been utilized as the classification metrics and the results are summarized in Table IV.

From Tables III and IV, it can be observed that the LSTM and GRU on the original feature space demonstrate lower performance in comparison to the competitive methods in some cases. Incorporating *L1* has led to a noticeable improvement in the prediction accuracy, which could be associated with the overfitting of networks. Since RNNs have a high number of variables and weights, they require a larger number of samples for training. The approach investigated in this paper employs the *L1* feature selection to overcome the limited number of samples for training an effective network, which can predict the future status of AD subjects using their historical measurements.

IV. CONCLUSION

For tracking the progression of the AD at multiple future intervals and gauging the merits and gradual effects of any potential treatment plan in longitudinal AD studies, this paper aimed to apply Recurrent Neural Networks to the ADNI dataset. Three historical time points from subjects in three categories of CN, MCI and AD were selected to form a feature space. Then, the model is trained on 75% of the data to predict three future MMSE scores and diagnosis labels of the subjects with two different variations of RNN (LSTM and GRU). Employing L1 feature extraction prior to application of the RNNs lead to a higher performance in both regression and classification models in comparison to other state of the art algorithms, which can be observed from the results provided in Tables III and IV.

ACKNOWLEDGMENT

This research has been supported by the National Science Foundation (NSF) under NSF grants CNS-1532061, CNS-1429345, CNS-1551221, and CNS-1338922. We also greatly appreciate the support of the 1Florida Alzheimer's Disease Research Center (ADRC) (NIA 1P50AG047266-01A1) and the Ware Foundation. Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from other sources.

REFERENCES

[1] R. Cuingnet, E. Gerardin, J. Tessieras, G. Auzias, S. Lehéricy, MO. Habert, M. Chupin, H. Benali, O. Colliot, "Automatic classification of patients with Alzheimer's disease from structural MRI: A comparison of ten methods using the ADNI database," *Neuroimage*, vol. 56, no. 2, pp. 766–781, 2011.

[2] L. Nie, L. Zhang, L. Meng, X. Song, X. Chang, and X. Li, "Modeling Disease Progression via Multisource Multitask Learners: A Case Study with Alzheimer's Disease," *IEEE Trans. Neural Networks Learn. Syst.*, vol. 28, no. 7, pp. 1508–1519, 2017.

[3] C. Hinrichs, V. Singh, L. Mukherjee, G. Xu, M. K. Chung, and S. C. Johnson, "Spatially augmented LPboosting for AD classification with evaluations on the ADNI dataset," *Neuroimage*, vol. 48, no. 1, pp. 138–149, 2009.

[4] Y. Gao *et al.*, "MCI identification by joint learning on multiple MRI data," in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 2015.

[5] J. Zhou, J. Liu, V. A. Narayan, and J. Ye, "Modeling disease progression via multi-task learning," Neuroimage, vol. 78, pp. 233–248, 2013.

[6] L. Huang, Y. Jin, Y. Gao, K.-H. Thung, and D. Shen, "Longitudinal clinical score prediction in Alzheimer's disease with soft-split sparse regression based random forest," Neurobiol. Aging, vol. 46, pp. 180–191, 2016.

[7] G. Gavidia-Bovadilla, S. Kanaan-Izquierdo, M. Mataroa-Serrat, and A. Perera-Lluna, "Early prediction of Alzheimer's disease using null longitudinal model-based classifiers," PLoS One, 2017.

[8] S. Tabarestani, M. Aghili, M. Shojaie, C. Freytes and M. Adjouadi, "Profile-Specific Regression Model for Progression Prediction of Alzheimer's Disease Using Longitudinal Data," 2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA), Orlando, FL, 2018, pp. 1353-1357.

[9] S. Tsao, N. Gajawelli, J. Zhou, J. Shi, J. Ye, , Y. Wang & N. Leporé, "Feature selective temporal prediction of Alzheimer's disease progression using hippocampus surface morphometry," *Brain Behav.*, vol. 7, no. 7, pp. 1–11, 2017.

[10] M. Nguyen, N. Sun, D. C. Alexander, J. Feng, and B. T. Thomas Yeo, "Modeling Alzheimer's disease progression using deep recurrent neural networks," in 2018 International Workshop on Pattern Recognition in Neuroimaging, PRNI 2018, 2018.

[11] R. Cui, M. Liu, and G. Li, "Longitudinal analysis for Alzheimer's disease diagnosis using RNN," *Proc. - Int. Symp. Biomed. Imaging*, vol. 2018–April, no. Isbi, pp. 1398–1401, 2018.

[12] E. Choi, A. Schuetz, W. F. Stewart, and J. Sun, "Using recurrent neural network models for early detection of heart failure onset," *J. Am. Med. Informatics Assoc.*, vol. 24, no. 2, pp. 361–370, 2017.

[13] T. Wang, R. G. Qiu, and M. Yu, "Predictive Modeling of the Progression of Alzheimer ' s Disease with Recurrent Neural Networks," *Sci. Rep.*, pp. 1–12, 2018.

[14] M. Aghili, S. Tabarestani, M. Adjouadi, and E. Adeli, "Predictive Modeling of Longitudinal Data for Alzheimer's Disease Diagnosis Using RNNs," in *PRedictive Intelligence in MEdicine*, 2018, pp. 112–119.

[15] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," Neural Comput., 1997.

[16] K. Cho, B. van Merrienboer, Ç. Gülçehre, F. Bougares, H. Schwenk, and Y. Bengio, "Learning Phrase Representations using {RNN} Encoder-Decoder for Statistical Machine Translation," *CoRR*, vol. abs/1406.1078, 2014.

[17] M. Mafi, S. Tabarestani, M. Cabrerizo, A. Barreto and M. Adjouadi, "Denoising of ultrasound images affected by combined speckle and Gaussian noise," in IET Image Processing, vol. 12, no. 12, pp. 2346-2351, 12 2018.

[18] L. Ghorbanzadeh, et al., "Development of a synthetic adaptive neuro-fuzzy prediction model for tumor motion tracking in external radiotherapy by evaluating various data clustering algorithms," Technology in cancer research & treatment, vol. 15, pp. 334-347, 2016.