Missing Data in Longitudinal Image Retrieval for Alzheimer's Disease

Katarina Trojachanec Dineva
Faculty of Computer Science and Engineering
Skopje, Republic of North Macedonia
katarina.trojacanec@finki.ukim.mk

Ivan Kitanovski
Faculty of Computer Science and Engineering
Skopje, Republic of North Macedonia
ivan.kitanovski@finki.ukim.mk

Ivica Dimitrovski
Faculty of Computer Science and Engineering
Skopje, Republic of North Macedonia
ivica.dimitrovski@finki.ukim.mk

Suzana Loshkovska
Faculty of Computer Science and Engineering
Skopje, Republic of North Macedonia
suzana.loshkovska@finki.ukim.mk

*for the Alzheimer's Disease Neuroimaging Initiative

***Abstract*— The paper is focused on the missing scans in the context of longitudinal image retrieval for Alzheimer's Disease. Namely, we explore the influence of missing data on the retrieval results when the subjects are represented by the longitudinal changes calculated on the basis of the within-subject template generated using the available time points. To evaluate the effect of the missing scans, we defined two (most characteristic and most common) scenarios, in which missing scans at a specific time point are considered, and one scenario that is based on complete data used as a baseline to compare against. Additionally, we increased the number of patients with missing scans from 10% to 50% and evaluated its impact on the retrieval results.**

The evaluation showed that from the examined types of feature vectors, concatenated longitudinal changes of the volumes of the cortical and sub-cortical structures are superior and robust. In the case when the dimensionality of the descriptor is an important criterion, we recommend the usage of the percent change or symmetrized percent change of the volumetric measures. Additionally, the influence of the missing scans on the retrieval results is lower when incomplete data occurs in the early time points, rather than in later ones. Moreover, very little or no performance reduction was detected by increasing the number of subjects with missing scans. In general, the evaluation showed very small or no performance degradation in the retrieval process in the scenarios with missing scans, in comparison to the scenario with fully complete data.

Keywords*— *Missing Data, Longitudinal Images, Longitudinal Image Retrieval, Alzheimer's Disease, Magnetic Resonance Imaging.

I. INTRODUCTION

Alzheimer's Disease (AD) is a progressive neurodegenerative disease and the most common cause of dementia [1]. Considering this, early diagnoses of the disease, monitoring the patient's condition or the disease progression, finding powerful diagnostic or prognosis biomarkers, identifying the patients who are most probable to ultimately develop AD, as well as reaction to the therapy are active research fields.

Advances in genetics and medicine, as well as the rapid evolution of technology and neuroimaging techniques increase the amount of generated data of the medical cases for AD. Among all the data generated within the medical cases for AD, Magnetic Resonance Images (MRI) offer extremely good opportunities and are found to have a key role in early diagnosis, prognosis and monitoring of AD [2]. Considering that the information extracted from the brain images provide precise and consistent markers for diagnosis and monitoring the development of the disease [3-4], our research is towards medical case retrieval by using medical images as input queries.

The image retrieval process involves representation of the query image with a descriptor and then comparing it with the descriptors of all images in the medical data- base. The result is a list of all images sorted in the database by similarity, so that the most similar one is at the top [6]. In that sense, for a given medical case of a patient for whom an MRI was acquired, the image is given as a query to the system. The retrieval system provides a sorted list of all images of other patients according to their similarity to the query. The main challenge is to get the result that is semantically relevant [7-8].

Taking into consideration the progressive nature of AD, longitudinal data are extremely important. Longitudinal images (images acquired at multiple consecutive time points) are meant to reflect the disease progression. However, a key problem that arises considering longitudinal images is missing data, i.e. lack of scan/s for one or more time points. The reason for this might be the patient's inability to undergo scanning at the predefined period, disability to continue the examination, poor quality of the scan, etc. [9]. In addition, the occurrence of missing data where the time difference between time points is not uniform is very common [10].

One possible solution is to exclude from the research all patients with incomplete data. The main limitation of this solution is that it significantly reduces the number of patients included in the examination, excluding possibly important knowledge that might be extracted from them. One way to overcome this limitation is to use data imputation [5,9, 11-12]. This would complement those features that correspond to the missing time point, in order to maintain the same dimension of the vector for all scans and to artificially

[*] 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 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.

supplement the missing information. In [13] multivariate normal imputation (MVNI) method and fully conditional specification (FCS) method are compared and discussed. This study is the first that uses real data as part of an ongoing trial to make the validation and comparison. Detailed review of the hierarchy of missing data mechanisms and their relationship to likelihood-based methods and a series of simulation studies with designs common in longitudinal neuroimaging studies is provided in [14]. Although the challenging problem of missing data has been researched for over forty years [14], it still an active research field and a big challenge.

The aim of the paper is to investigate the influence of missing scans on the retrieval performance in the context of longitudinal image retrieval for AD. Several key questions arise here:

1. What kind of features are more powerful and robust when missing scans are present?
2. How does retrieval results change with increasing the number of patients with missing data?
3. Does the time point at which the scan is missing affect the retrieval results?
4. How does a missing scan at a different time point affect the retrieval results in comparison to fully complete dataset?

A research that answers these questions in the context of image retrieval for AD is still not performed. Thus, our research is aimed to provide insight into the problem of missing data in this context as well as answers to the aforementioned questions.

Moreover, taking into consideration the diversity of the datasets and/or the sample selection for the evaluation process, we cannot provide an objective comparison with other research involving addressing the problem of missing data.

The paper is organized as follows. We present the methods used in this study in Section 2. The experimental results and discussion are provided in Section 3. The concluding remarks and future work are summarized in Section 4.

II. Materials and Methods

A. Participants and Inclusion Criteria

The performed research is based on the scans and data obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). launched in 2003 by the National Institute on Aging (NIA), the National Institute of Biomedical Imaging and Bioengineering (NIBIB), the Food and Drug Administration (FDA), private pharmaceutical companies, and non-profit organizations as a $60 million, 5-year public–private partnership. The aim of the initiative is to enable research on whether serial magnetic resonance imaging (MRI), positron emission tomography (PET), other biological markers, such as cerebrospinal fluid (CSF) markers, APOE status and full-genome genotyping via blood sample, as well as clinical and neuro-psychological assessments can be used together to indicate and estimate the progression of mild cognitive impairment (MCI) and Alzheimer's Disease (AD). Finding appropriate markers indicating early AD progression is aimed towards development of new treatments, improving the process for monitoring treatments effectiveness, and reducing the time and cost of the clinical trials.

The Principal Investigator of this initiative is Michael W. Weiner, MD, VA Medical Center and University of California - San Francisco. Many coinvestigators from a broad range of academic institutions and private corporations have given their support and contribution to make ADNI a valuable product. Subjects have been recruited from over 50 sites across the U.S. and Canada, initially supposed to be 800 subjects in total. However, ADNI has been followed by ADNI-GO and ADNI-2, with over 1500 participants, namely in fact adults in the age group of 55 to 90 years. The protocols for ADNI-1, ADNI-2, and ADNI-GO specify the three groups (Cognitively normal individuals, adults with early or late MCI, and people with early AD) available in the dataset with different follow up duration of each group. For up-to-date information, see http://www.adni-info.org.

For this research, we used the standardized list from ADNI-1, containing images acquired at multiple time points. From this list, we selected the subjects that have available scans at baseline (TP1), and the 6-month (TP2), 12-month (TP3), and 24-month (TP4) follow-ups and belong to AD or normal control (NL) group. In this way, we selected a total of 267 subjects, 168 in AD group, and 99 in NL group. Patients' demographics information can be found in [15]. The timing of scans per time point by clinical group is provided in Table I, also available in [15], but included in this paper as well for clarity.

We chose exactly this selection criteria because of the following reasons: (1) more time points (ex. 36-month follow-up) is not available for the AD group of patients and, additionally, the total number of patients for whom all the scans are available is reduced by more than 12%; (2) a smaller number of time points would not give enough space for research and an opportunity to have a good insight into the problemч (3) with the selected time points, we have an opportunity to analyze the problem in the case of equally and unequally spaced available time pointes (depending of which time point is missing).

TABLE I. Timing of Scans per Time Point by Clinical Group

Time Point	Time from baseline (years)	
	AD	***NL***
TP1 (Baseline)	0	0
TP2 (Month 6)	0.57 (0.05)	0.58 (0.06)
TP3 (Year 1)	1.07 (0.05)	1.08 (0.07)
TP4 (Year 2)	2.09 (0.14)	2.1 (0.11)

Abbreviations: TP = Time Point.

The time from baseline is given in mean (standard deviation).

B. Longitudinal image representation

This study is based on longitudinal data for patients that undergo examination for Alzheimer's Disease. This kind of data are characterized by that the outcome variables are measured repeatedly on the same cohort of individuals at multiple time points. This enables identifying the influence of the changes in the derived measurements over time to the examined clinical, biological or experimental factor, possibly reflecting the diseases progress and/or reaction to the treatment. Longitudinal studies provide direct assessment of

within-subject changes across different time points, free of any between-subject variability [10, 20]. Several distinctive characteristics should be emphasized for longitudinal data [10]:

1. Longitudinal measurements are ordered in time. They reflect the temporal trajectory of an underlying non-stationary continuous process.
2. Typically, serial measurements obtained for a single subject are positively correlated caused by the smooth trajectory of the underlying biological process.
3. Between-subject variance is not usually constant over the duration of the study.
4. Missing data and non-uniform timing are extremely common, particularly for longitudinal studies of larger duration.

This study takes into account the aforementioned characteristics and focuses specifically on the fourth. Namely, when longitudinal images are available for the patients, the information contained in all of them should be properly extracted and utilized. This is a crucial step to be able to get the semantically more relevant retrieval result. A significant challenge that arises in this case is how to deal with the missing scan/s for some of the patients.

In fact, to deal with the problem of missing scans for Alzheimer's disease, we used the feature extraction method that we explored in our previous research [15] and evaluate the impact of the missing time point when using this approach. Namely, for each patient who has a missing scan at a specific time point, we processed the data longitudinally and generated a within-subject template only from the available time points for a particular patient and not for all possible time points. For example, if for a particular patient only the scan at baseline (TP1), the 12-month (TP3), and the 24-month (TP4) follow-ups are available, then we generate the template on the basis of only these time points, instead of on the basis of all the possible four time points (scans at baseline (TP1), the 6-month (TP2), 12-month (TP3), and 24-month (TP4) follow-ups)). The longitudinal scheme for generating the template is designed to be unbiased with respect to any time point and there was no initialization with information from a specific time point [16].

We performed this processing with the FreeSurfer's fully automated longitudinal pipeline [16]. Fig. 1 shows a schematic illustration of the main steps in the longitudinal pipeline. The full list of methods applied to the examined dataset is included in [15] and proposed and explained in more details in [16].

After this kind of processing, we generated the feature vector by using longitudinal changes of the volume of the cortical and sub-cortical regions. The following reliable estimates of the longitudinal changes on the bases of the available time points were calculated and used as features [15 -17]:

- VolumesRC - rate of change (RC) of the volumes (calculated as a difference per time unit). The statistic is derived from the volumetric temporal information per subject for each cortical (34 measures for each hemisphere) and sub-cortical region (55 measures), 123 measures in total;
- VolumesPC1fit - percent change of the volumes (calculated with respect to the value obtained from the linear fit at baseline, i.e. percent thinning/volume loss per year) The statistic is derived from the volumetric temporal information per subject for each cortical and sub-cortical region, 123 measures in total;
- VolumesSPC – symmetrized percent change (the rate with respect to the temporal average). The statistic is derived from the volumetric temporal information per subject for each cortical and sub-cortical region, 123 measurements in total.

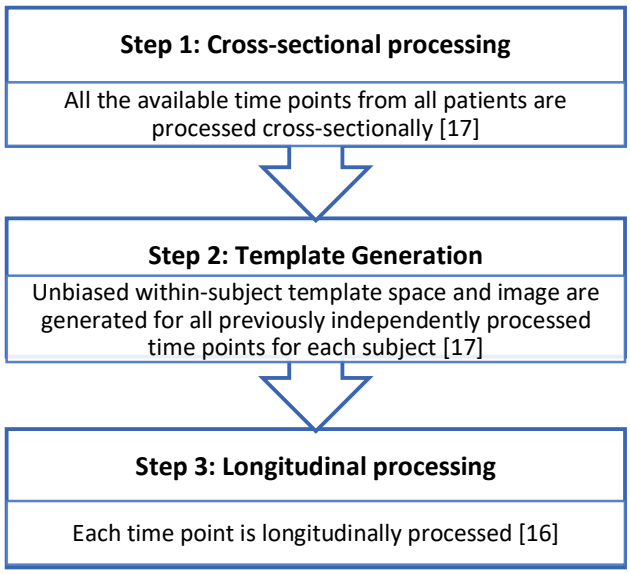

Fig. 1. Schematic illustration of the main steps in the FreeSurfer' longitudinal pipeline

These features reflect the speed and degree of degeneration caused by the disease, in fact the disease severity and progression. This way, for a patient with missing data, a representation extracted only from the available scans will be generated and will be of the same type and with the same dimension as those for the patients with complete data. In fact, the dimension of all descriptors for all patients will be the same, regardless the number of the available time points. For example, if a scan taken at the screening 6 months after the first visit to the hospital is missing, then the scans at TP1, TP3, and TP4 are longitudinally processed and a template on the basis of these scans is generated and subsequently the longitudinal changes of the volume of the cortical and sub-cortical structures as well as of the cortical thickness are calculated. For a patient with complete data for instance, all the time points are longitudinally processed, and the same statistics of the same regions are generated. However, the dimension of the feature vector is the same as it is for the patient with three available time points.

Although it is also possible to calculate longitudinal changes of the cortical thickness measures, the reason for using the volumetric changes is that in our previous research [15] they have proven to be superior over the longitudinal changes of the cortical thickness in the context of image retrieval for AD.

We chose this strategy because it has two main advantages: (1) it can operate only with the available data independently for each subject, whether or not they have a missing scan at some point in time (2) it provides the same

descriptor dimensionality for all patients, regardless of the number of the available time points.

Additionally, we applied the Correlation-based Feature Selection (CFS) method [18] with the aim to reduce the feature vector dimensionality and to select the most relevant features, because it provided superior results on the experimental basis for the longitudinal image retrieval for Alzheimer's disease that we reported in [19]. To provide relevant and unbiased results, we performed the feature selection method independently of the query subject information. In fact, we obtained a specific feature subset for each query subject using the information of all other subjects similarly like in [15].

C. *Evaluation of the influence of the missing data on the retrieval results*

To be able to evaluate the influence of the missing data on the retrieval results, we defined two most characteristic and most common scenarios with missing time points and one scenario for reference:

- M1 – from the available four time points, scans at TP2 were excluded from x% randomly selected patients, where $x \in \{10, 20, 30, 40, 50\}$, keeping at least 50% of the patients with fully complete data.
- M2 – from the available four time points, scans at TP4 were excluded from x% randomly selected patients, where $x \in \{10, 20, 30, 40, 50\}$, keeping at least 50% of the patients with fully complete data.
- M3 – fully complete dataset that we used as a baseline to compare against.

It should be noted that both M1 and M2 scenarios contain the first time point. In the clinical environment, this is most often the case because it is a screening made during the first visit to a medical institution and in most cases, it is available with good quality. We did not evaluate the scenario in which the third time point is missing, because during the longitudinal processing of the other available scans (when the third TP is excluded), there was an error that required manual interventions in the processing in some of the cases. We wanted to keep the processing fully automated, so as to be able to reduce the subjectivity of the human factor and to perform fair comparison. Another characteristic of M1 and M2 is that the time points available are at the same time distance (12 months for M1 and six months for M2). Additionally, only the M1 scenario contains the time point that is the most temporally distant from the first visit to a doctor, that is the fourth TP. This is important to be noted because according to the previous experiments it has an impact on retrieval results when all time points are available [15]

In this examination, we did not consider templates and accordingly, scenarios, with only two time points, because in order to generate the template, a voxel-wise median was used. But, when only two time points are available, median is equal to mean, hence resulting in blurry, instead of crispy edges, which is the case when more time points are available to build the template [16].

To evaluate the proposed strategy, we used the standard evaluation metric for quantitative measurement of the retrieval performance, Mean Average Precision (MAP). It was calculated as the mean of the average precision scores for each query, evaluation metric for the general retrieval performance. It is meant to favor retrieval systems that return more relevant subjects at the top of the list. The retrieved subject is assumed to be relevant if the patient has the same diagnosis as the query one. The higher the relevant subjects are in the retrieved list, the higher the value of the precision is.

Considering the small number of patients included in this study, we used leave-one-out strategy. This means that each patient's representation was used as a query against all other representations stored in the database.

Considering that a random selection of patients with missing data was performed, and to get representative results, we did 10 repetitions of the random selection and we repeated the experiment for each case accordingly. Then we calculated the mean of the MAP for each of the 10 repetitions. Subsequently we compared those results with the retrieval results obtained by the scenario based on the complete dataset (M3).

III. Experimental Results and Discussion

This section summarizes the results from the evaluation of the presented strategy for dealing with missing data in the context of longitudinal image retrieval for AD.

Table II shows the results associated with the M1 scenario for a different percentage of patients with a missing scan at TP4. According to the results, the rate of change of the volume of the cortical and sub-cortical regions provides the same value as in the case of fully complete data when the number of patients with missing data is not more than 30% of the total number of patients in the examined dataset. When 30-50% of the subjects have missing scan at TP4, the value of MAP is lower by 1%. The value of VolumesPCfit and the VolumesSPC remains constant as the number of patients with missing data increases up to 50%, which makes them stable features, suitable in the case of missing data. It should be noted that VolumesPCfit led to a 1% lower value of MAP in comparison with the case of fully complete data. The combination of all calculated longitudinal changes led to the best result in terms of MAP. In this case, the value of MAP remained the same, even though the number of patients with missing data was increased up to 50%. Moreover, it is the same as the corresponding value in the case of fully complete data.

TABLE II. Evaluation of the scenario M1 based on the value of MAP – missing scan at TP2 in x% randomly selected patients, where $x \in \{10, 20, 30, 40, 50\}$

	M1 (10%)	**M1 (20%)**	**M1 (30%)**	**M1 (40%)**	**M1 (50%)**	**M3**
VolumesRC	0.76	0.76	0.76	0.75	0.75	0.76
VolumesPCfit	0.78	0.78	0.78	0.78	0.78	0.79
VolumesSPC	0.78	0.78	0.78	0.78	0.78	0.78
All volume rates - concatenated	0.79	0.79	0.79	0.79	0.79	0.79

Regarding the second scenario M2 (Table III), the obtained results indicate reduction of MAP value by 1-4% when the number of patients with missing data increases to 50%, when the patients are represented by the rate of change of the volumetric measures. Even when 10% of the patients

are with missing data, the value of MAP is reduced. By increasing this percentage to 20, there was again a reduction in MAP by 1%, but it remained the same when the number of patients with missing data was increased to 30%. Then again, the value of MAP decreased by 1% in case of 40% of patients with missing data. This trend was also the case when the number of patients with missing data was increased to 50%. When PCfit and SPC of the volumes of cortical and sub-cortical structures were used to generate the descriptors, the value of MAP started at 0.77 and decreased by 0.01 as the number of patients with missing data increased to 20%. It remained the same in the case of 30% of patients with missing data. Then it decreased firstly by 1% and then, by 2% as the number of patients with missing data increased to 40% and 50% respectfully. When concatenation of all the longitudinal changes was used to generate the feature vector, it led to the best results regarding the M2 scenario but showed the same trend of decreasing the values of MAP.

TABLE III. EVALUATION OF THE SCENARIO M2 BASED ON THE VALUE OF MAP – MISSING SCAN AT TP4 IN X% RANDOMLY SELECTED PATIENTS, WHERE $X \in \{10, 20, 30, 40, 50\}$

	M2 (10%)	M2 (20%)	M2 (30%)	M2 (40%)	M2 (50%)	M3
VolumesRC	0.75	0.74	0.74	0.73	0.72	0.76
VolumesPCfit	0.77	0.76	0.76	0.75	0.73	0.79
VolumesSPC	0.77	0.76	0.76	0.75	0.73	0.78
All volume rates - concatenated	0.78	0.77	0.77	0.76	0.74	0.79

According to the performed evaluation, we can note that the presence of a time point that is more distant from the first visit to a medical institution, makes the features more relevant, more stable and more powerful to deal with missing scan. In fact, the ability of the algorithm to handle missing data in the context of image retrieval is closely related to the longitudinal processing of the scans, i.e. the stability and relevance of the results of the longitudinal processing using the available points. Namely, according to the previous research [15], it was concluded that the features extracted from the scans at time point 1, 3 and 4 are more promising and lead to greater accuracy compared to those extracted on the basis of the scans at time point 1, 2 and 3. This was directly reflected in the results obtained from the evaluation of the scenarios with missing data.

The main findings of our research on the bases of the evaluation of the defined scenarios are:

1. From the examined features vector types, VolumesPCfit or VolumesSPC, or concatenated feature vector of all longitudinal change, VolumesPCfit and VolumesSPC proved to be more efficient, stable and powerful to deal with missing scans when separate longitudinal changes are used.
2. Very small or no performance reduction as the number of cases with missing data increases. This means that the strategy for representing the subjects with longitudinal changes of the volumetric measures provides reliable and robust way of image representation, and the whole setup that we used for image retrieval provides promising way for facing the problem of missing data.
3. When the scan is missing in the earlier time points, then the influence to the retrieval results is smaller, rather than in the case of missing data at the later time points. This is highly expected because the changes in the brain as the time passes carry more significant information about the progression of the disease.
4. There is no or small degradation in the retrieval performance in cases with missing scans in comparison to the scenario with fully complete data.

According to the obtained results, in the case of missing data we suggest representing the patients' scans by feature vector comprised of concatenated longitudinal changes of the volumes of cortical and sub-cortical structures. In case when the feature vector should be kept as short as possible, then VolumesSPC or VolumesPCfit of the volumetric measurements are advisable as they turned out to be superior and more robust.

IV. CONCLUSION

Missing data is a serious problem when longitudinal data are considered. In this paper, we evaluated the influence of the missing data in the context of image retrieval for Alzheimer's disease. Our results showed that the retrieval performance is closely related to time points at which the scans are missing and to the reliability of the longitudinal data extracted from the available time points. The features extracted from the scans at the later time points, proved to be more efficient, stable and powerful to deal with missing scans. From the examined types of feature vectors, concatenated longitudinal changes of the volumes of cortical and sub-cortical structures provided superior results. In the case when the dimensionality of the descriptor is important factor, VolumesSPC or VolumesPCfit of the volumetric measurements are advisable to be used as more powerful and robust. In all cases, we noticed very small or no degradation in the retrieval performance when we used these features as the number of cases with missing data increased.

In the future, we plan to extend the evaluation of this strategy on a bigger cohort, to include more available time points and also to consider scenarios with more than one missing scan per patient. Moreover, we are going to examine deep neural network architectures developed in a longitudinal manner to provide a solution for dealing with missing data.

ACKNOWLEDGEMENT

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). This study was partially funded by the Faculty of Computer Science and Engineering, Skopje through the SatView project.

REFERENCES

[1] Alzheimer's Association, 2019. 2019 Alzheimer's disease facts and figures. Alzheimer's & Dementia, 15(3), pp.321-387. [Last access: 01.03.2022]

[2] Ye,J.,Wu,T.,Li,J.andChen,K.,2011.Machinelearningapproachesforthe neuroimaging study of Alzheimer's disease. Computer, 44(4), pp.99-101. [Last access: 01.03.2022]

[3] Bron, E.E., Klein, S., Reinke, A., Papma, J.M., Maier-Hein, L., Alexander, D.C. and Oxtoby, N.P., 2022. Ten years of image analysis and machine learning competitions in dementia. *NeuroImage*, *253*, p.119083. [Last access: 15.04.2022]

[4] Hedges, E.P., Dimitrov, M., Zahid, U., Vega, B.B., Si, S., Dickson, H., McGuire, P., Williams, S., Barker, G.J. and Kempton, M.J., 2022.

Reliability of structural MRI measurements: The effects of scan session, head tilt, inter-scan interval, acquisition sequence, FreeSurfer version and processing stream. *NeuroImage*, *246*, p.118751. [Last access: 15.04.2022]

[5] Peng, L., Lin, L., Lin, Y., Chen, Y.W., Mo, Z., Vlasova, R.M., Kim, S.H., Evans, A.C., Dager, S.R., Estes, A.M. and McKinstry, R.C., 2021. Longitudinal Prediction of Infant MR Images With Multi-Contrast Perceptual Adversarial Learning. *Frontiers in neuroscience*, p.1114. [Last access: 15.04.2022]

[6] Piras, L. and Giacinto, G., 2017. Information fusion in content based image retrieval: A compre-hensive overview. Information Fusion, 37, pp.50-60. [Last access: 14.03.2022]

[7] Faria, A. V., Oishi, K., Yoshida, S., Hillis, A., Miller, M. I. and Mori, S., 2015. Content-based image retrieval for brain MRI: An image-searching engine and population-based analysis to utilize past clinical data for future diagnosis. NeuroImage: Clinical, 7, pp.367-376. [Last access: 14.03.2022]

[8] Qayyum, A., Anwar, S. M., Awais, M. and Majid, M., 2017. Medical image retrieval using deep convolutional neural network. Neurocomputing. [Last access: 14.03.2022]

[9] Thung, K.H., Wee, C.Y., Yap, P.T. and Shen, D., 2016. Identification of progressive mild cognitive impairment patients using incomplete longitudinal MRI scans. Brain Structure and Function, 221(8), pp.3979-3995. [Last access: 17.03.2022]

[10] Bernal-Rusiel, J.L., Greve, D.N., Reuter, M., Fischl, B., Sabuncu, M.R. and Alzheimer's Disease Neuroimaging Initiative, 2013. Statistical analysis of longitudinal neuroimage data with linear mixed effects models. Neuroimage, 66, pp.249-260. [Last access: 17.03.2022]

[11] Lei, B., Yang, M., Yang, P., Zhou, F., Hou, W., Zou, W., Li, X., Wang, T., Xiao, X. and Wang, S., 2020. Deep and joint learning of longitudinal data for Alzheimer's disease prediction. Pattern Recognition, 102, p.107247. [Last access: 24.03.2022]

[12] Liu, X., Chen, K., Wu, T., Weidman, D., Lure, F. and Li, J., 2018. Use of multimodality imaging and artificial intelligence for diagnosis and prognosis of early stages of Alzheimer's disease. Translational Research, 194, pp.56-67. [Last access: 24.03.2022]

[13] Rosato, R., Pagano, E., Testa, S., Zola, P. and di Cuonzo, D., 2021. Missing data in longitudinal studies: Comparison of multiple imputation methods in a real clinical setting. Journal of Evaluation in Clinical Practice, 27(1), pp.34-41. [Last access: 24.03.2022]

[14] Matta, T. H., Flournoy, J. C. and Byrne, M. L., 2018. Making an unknown unknown a known unknown: Missing data in longitudinal neuroimaging studies. Developmental cognitive neuroscience, 33, pp.83-98. [Last access: 24.03.2022]

[15] Trojachanec, K., Kitanovski, I., Dimitrovski, I. and Loshkovska, S., 2017. Longitudinal brain MRI retrieval for Alzheimer's disease using different temporal information. IEEE Access, 6, pp.9703-9712. [Last access: 24.03.2022]

[16] Reuter, M., Schmansky, N.J., Rosas, H.D. and Fischl, B., 2012. Within-subject template estimation for unbiased longitudinal image analysis. Neuroimage, 61(4), pp.1402-1418. [Last access: 24.03.2022]

[17] Reuter, M., Rosas, H. D., Fischl, B., 2010. Highly Accurate Inverse Consistent Registration: A Robust Approach. Neuroimage 53 (4), pp.1181-1196. [Last access: 25.03.2022]

[18] Hall, M., and Holmes, G., 2003. Benchmarking attribute selection techniques for discrete class data mining. IEEE Transactions on Knowledge and Data Engineering. 15(6), 1437–1447. [Last access: 01.04.2022]

[19] Trojacanec, K., Kitanovski, I., Dimitrovski, I. and Loshkovska, S., 2015, October. Medical image retrieval for Alzheimer's disease using data from multiple time points. In International Conference on ICT Innovations (pp. 215-224). Springer, Cham. [Last access: 01.04.2022]

[20] Fitzmaurice, G.M., Laird, N.M. and Ware, J.H., 2012. *Applied longitudinal analysis*. John Wiley & Sons. [Last access: 01.04.2022]