Alzheimer Diagnostics using Biomarkers

1st Florin Barbisch
School of Engineering, Data Science
FHNW University of Applied Sciences and Arts Northwestern Switzerland
Windisch, Switzerland
florin.barbisch@students.fhnw.ch

2nd Yvo Keller
School of Engineering, Data Science
FHNW University of Applied Sciences and Arts Northwestern Switzerland
Windisch, Switzerland
yvo.keller@students.fhnw.ch

3rd Patrick Schürmann
School of Engineering, Data Science
FHNW University of Applied Sciences and Arts Northwestern Switzerland
Windisch, Switzerland
patrick.schuermann@students.fhnw.ch

***Abstract*—Alzheimer's disease is a disease that generally affects the elderly population and for which no cure exists. It forms a subtype of dementia. A large number of studies on prevention, prediction, and recovery have been conducted worldwide in recent decades, but there are still no established drugs and the cause of the disease remains unclear. In this study, we investigate whether changes over different time periods in gut bacteria, cerebrospinal fluid (CSF) biomarkers, sleep behavior, and MRI scans can be used to predict degeneration to diagnose AD in patients with mild cognitive impairment. To do this, we use KNN, logistic regression, multilayer perceptron, and 3D-CNN models. The data basis is the Alzheimer's Disease Neuroimaging Initiative (ADNI). Further, we investigate whether combining the three factors gut bacteria, CSF, and sleep behavior can lead to a more accurate result. Finally, we extend the data used for prediction to include MRI scans of the brain. We found that when the individual factors are used separately, only the change in sleep behavior provides meaningful results. This was across the entire study period of the patients and with a precision of 82%. Based on the changes from one doctor visit to the next, no reliable predictions could be made with any of the data sets. With the combined data set, a precision of 65% could be achieved. Again, this result only applies over the entire observation period of individual patients. With the addition of MRI scans to the combined data set, our multi-modal model achieved a precision of 72%. The results are discussed regarding limitations and implications for AD diagnostics.**

***Index Terms*—Alzheimer's disease, biomarkers, sleep, CSF, gut bacteria, magnetic resonance imaging, ADNI, classification, diagnosis, machine learning, deep learning**

I. Introduction

With this paper, we hope to create new insights and contributions to the community investigating AD and finding methods to fight it. We investigate the correlation between Alzheimer's disease (AD) and several factors including gut bacteria, sleep behavior, and blood values.

A. Alzheimer's Research on Microbiota, Sleep, and CSF

Alzheimer's disease is a disease that affects mostly the elderly population and is defined as a subgroup of dementia. Currently, 150,000 people in Switzerland live with dementia and about 60% of them suffer from Alzheimer's disease (AD) [1]. In 2017 16.5% of all deaths in Switzerland were caused by dementia [2]. Currently, there is no medication to prevent or cure Alzheimer's Disease. While science has found early indications, the cause of AD remains unknown. Available evidence suggests that gut microbiota is linked to brain connectivity and cognitive performance and that modulation of gut microbiota could be a promising strategy for enhancing cognition and emotional well-being in stressed and non-stressed situations [3], [4]. Over the past 10 years, the gut-brain-axis (GBA) and its relationship to health and disease of the brain have received considerable attention [5]. Current research suggests p-tau181, t-tau, and β-Amyloid are a predictor of Alzheimer's disease and neurodegeneration [6], [7]. The FDA recently approved a new treatment for patients with mild cognitive impairment that showed good results in reducing amyloid beta plaque [8]. Ju et al. [9] mention that disturbances in sleep hygiene impair the quality of life in Alzheimer's disease and suggest investigating whether improvement of the quality of sleep can reduce the risk of AD. On the other hand, Mecca et al. [10] contradict these findings and claim that a disturbance in sleep is not associated with a decline in memory.

B. Motivation and Research Questions

We aim to determine if gut bacteria, sleep behavior, and blood values can be used to predict changes in brain health. Our research questions include:

1) Can gut bacteria be used as a predictor for AD status and transition from mild cognitive impairment (MCI) to AD?
2) Can β-Amyloid-42, t-tau, and p-tau181 measured in cerebrospinal fluid (CSF) be used as early predictors of AD?
3) Can changes in sleep behavior be used to predict changes from MCI to AD?
4) Can a combination of these three factors be used to predict the transition from MCI to AD?

Our motivation for investigating the relationship between gut bacteria and AD is based on the observed link between gut bacteria and brain function [5]. Similarly, our interest in using CSF values as early predictors of AD stems from recent research [11]. Our investigation into the potential role of sleep in predicting AD is motivated by the importance of sleep for

Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (https://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

brain health and the potential influence of aging on sleep patterns [12]. Our aim to combine these three factors in predicting the transition from MCI to AD is driven by a desire to improve our understanding of the complex interactions between them. Furthermore, in identifying specific biomarkers important for predicting the transition to AD, we hope to get insights into what lifestyle choices affect these biomarkers, possibly leading a path to effective preventive lifestyle changes.

II. Data Source

We are working with the dataset collected by Alzheimer's Disease Neuroimaging Initiative (ADNI). ADNI was funded as a private-public partnership by 20 companies and two other foundations through the Foundation for the National Institutes of Health and the National Institute on Aging [13]. Phase 1 started in 2004 and lasted five years. Currently, phase 3 is active. ADNI has a standard procedure in each phase for data collection in subjects, for example vital samples and questionnaires. Patients from age 55 to 90 are being recruited in the United States and in Canada.

A. Tabular Data

From ADNI, we used the following tabular datasets.

- The "DXSUM_PDXCONV_ADNIALL" dataset, containing the longitudinal diagnosis data for each subject, served as our base dataset.
- For sleep, we worked with "NPI". The Neuropsychiatric Inventory (NPI) was developed by J.L. Cummings and is used to assess psychopathology in dementia patients (Cummings, 1997). It is based on a questionnaire for the patient's caregiver and gives information about the patient's sleeping behavior. NPI is what we refer to when mentioning the "sleep" dataset in our study.
- For researching gut bacteria, we used "ADMCGUTMETABOLITESLONG_12_13_21", which we refer to as "gut bacteria" dataset. This file is based on a project to measure a panel of gut microbial metabolites and contains 104 metabolites.
- The dataset used for CSF (Cerebrospinal fluid) values is called "UPENNBIOMK_MASTER" and will be referred to as "CSF" dataset. It contains β-Amyloid-42, Tau, and P-tau biomarkers.

B. Imaging Data

In addition to the tabular data on subjects, we used the collection "ADNI1:Complete 3Yr 1.5T" from ADNI [14]. This collection includes 2182 MRI scans created with 1.5 Tesla scanners from subjects who have baseline and follow-up screenings [15]. All MRI scans have gone through several steps of preprocessing: gradwarp, B1 non-uniformity, N3, and scaling [14].

III. Methods

This section describes the methods we applied on the data.

A. Data Preparation

We built a data pipeline to dynamically combine our tabular data from the different data sources. This pipeline is a crucial part in our method, as all our models rely on different feature sets, as described in section III-C. Each time data is needed, for example to train a model, we called the pipeline function with the relevant feature set as a parameter. It then extracted the features from their dataset and merged them using outer join on top of our base dataset "DXSUM_PDXCONV_ADNIALL". As merging identifiers, we used the Phase, RID and VISCODE2 keys. It is important to keep in mind that different preparation and preprocessing steps could lead to differing results.

B. Preprocessing

Using this previously described dataset, the pipeline continued by running the preprocessing functions on the extracted dataset. These functions include steps like setting the correct data types on columns and mapping the categorical shorthand ADNI names to more human-readable descriptions. They are inspired by the work done by Kobivasan et al. [16].

As not all tests and measurements are conducted on each visit of a patient, there are missing values. In observations of visits where the diagnosis is missing, we augment it by first forward- and then backward-filling the diagnosis per patient. Based on the diagnosis, we engineer our target variables PREV_DIAG_GROUP, NEXT_DIAG_GROUP, and PATIENT_DIAG_GROUP providing information about longitudinal changes in the patient's diagnosis on each observation. These target variables, having either a value of MCI-MCI or MCI-AD, are fundamental to our predictions.

As an example, PREV_DIAG_GROUP is calculated as the diagnosis at the patient's previous and current visit. Thus, a PREV_DIAG_GROUP of MCI-AD means that a patient was diagnosed as MCI on his visit 6 months ago but has now transitioned to AD. NEXT_DIAG_GROUP and PATIENT_DIAG_GROUP are calculated similarly, referring to the patient's current and next visit and overall first and last visit respectively. The target variables and their meaning are summarized in Table I.

TABLE I
Target variables for tabular predictions

Target Variable	Diag. Group Format	Predicting whether subject transitions…
PREV_DIAG_GROUP	PREV-CURRENT	…to AD on the current visit
NEXT_DIAG_GROUP	CURRENT-NEXT	…to AD on the next visit
PATIENT_DIAG_GROUP	FIRST-LAST	…to AD between first and last visit

At this point, all observations containing missing values are augmented by first forward- and then backward-filling per patient, similar to when a diagnosis is missing but limited to the range where the patient had the same diagnosis. As our work focuses on MCI-MCI vs MCI-AD classification, we drop all observations where the target variables are not consistent with this. This results in a clean dataset of patient visits where the values of all the selected features are known.

Before feeding the dataset into our models, some further preprocessing steps are needed. All categorical features are one-hot encoded. Assuming that the development of AD and therefore a change in diagnosis for a patient can best be predicted by longitudinal changes in our observed features and not their value on a single visit, we calculate the difference between visits. This calculation differs depending on the target variable that we try to predict with a specific model. For PREV_DIAG_GROUP and NEXT_DIAG_GROUP, we take the difference between the current and the previous visit of a patient, whereas the first and last visit of a patient is considered for PATIENT_DIAG_GROUP. By forward- and backward-filling missing values we created duplicate data, which renders a train-test split ineffective, so all duplicate observations are dropped. In preparation for model training, a train-test split (80% train / 20% test) is done on the data. Finally, to account for class imbalance, we perform oversampling on our training data. ADNI already applied some pre-processing steps to the 3D brain scans as described in section II-B. We further performed skull stripping, removing the osseous parts of the head using a tool called SynthStrip developed by Hoopes et al. [17]. An example of applied skull stripping can be seen in Fig. 1 and Fig. 2.

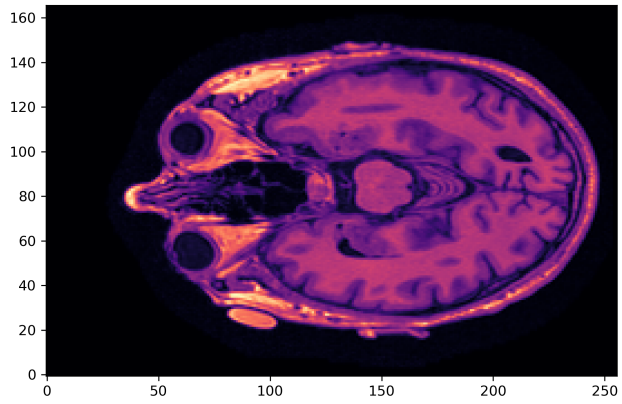

Fig. 1. Transversal Layer of an MRI Brain Scan before SynthStrip

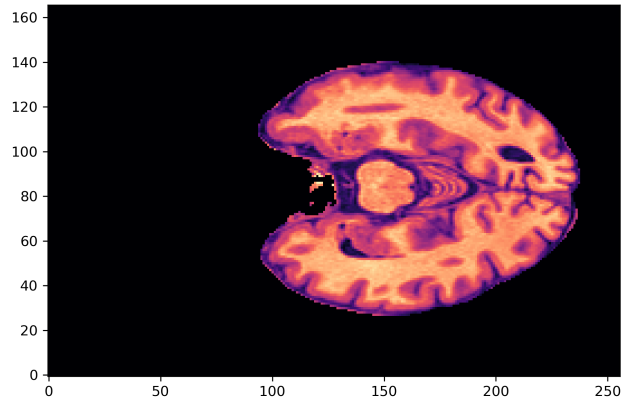

Fig. 2. Transversal Layer of an MRI Brain Scan after SynthStrip

C. Feature Selection

The datasets contain many different features. Our aim was to reduce the number of features to identify which contribute most to our predictions. We used three different methods to identify and select these from our datasets on gut bacteria, CSF, and sleep. Each of these three methods were applied on our three target variables, leading to 9 feature sets. They are implemented as follows:

1) Logistic Regression: For our first method, we trained a separate logistic regression classifier on every feature for every dataset. Then, we selected the top $\sqrt{n}$ features based on their accuracy in predicting MCI-MCI / MCI-AD groups, where n refers to the number of features in said dataset. This formula leaves us with an adequate number of features for each dataset.
2) Backward Elimination: The second method combined logistic regression and backward elimination. We first trained a logistic regression classifier on the whole dataset and then dropped the feature with the smallest contribution to the accuracy of the model. We then repeated the training, dropping further features on the new, smaller dataset following the same selection criteria. This process was stopped when the number of features reached $\sqrt{n}$.
3) Decision Tree: The third method involved training a decision tree with a maximum depth of two. Then again selecting the top $\sqrt{n}$ features by their sole accuracy in the MCI-MCI / MCI-AD classification.

All these methods result in a set of features that are most relevant in predicting the transition from MCI to AD according to the feature selection process and criteria. Thus, each list of best features for the combined dataset consists of 4 sleep, 11 gut bacteria, and 2 CSF features.

D. Model Training on Tabular Data

Using the 9 combinations of feature set and target variable obtained in our feature selection process, we loaded and pre-processed the dataset for each of the combinations and trained four different types of classifiers: logistic regression, decision tree, k-nearest neighbour (KNN), and multilayer perceptron (MLP). We also trained these four classifiers on each dataset (CSF, sleep, gut bacteria) solely using all their features.

These models were used for multiple reasons. First, our data includes quantitative and categorial variables, therefore we needed models that can be applied on both types of data. Second, these four models cover a wide range of data science methods, and we aim to find out how well these different methods work and compare them on the same dataset.

Logistic regression is strong for classifying quantitative data but does not work well with categorial variables. Since the only dataset (sleep) that contains categorial data also contains quantitative data, we are still applying this model to all data. The decision tree is strong for classifying categorial data but splits quantitative data into categories. This method can therefore still be applied to all datasets. For KNN, the datatype does

not matter, it only determines how close different observations are to each other. A multilayer perceptron can also be applied to all datatypes, but results will be difficult to explain, due to its complex nature.

Hyperparameters for each model were optimized using gridsearch. For logistic regression the penalty and the regularization strength were optimized. For decision tree the hyperparameters maximum depth, minimum samples to split, and complexity parameter for minimal cost-complexity pruning were optimized. For KNN number of neighbors, weighting of distance, and power parameter for Minkowski metric were optimized. For the multi-layer perceptron, the shape and number of the hidden layers were optimized.

In summary, we trained these models to predict the diagnosis group, specifically our three target variables PREV_DIAG_GROUP, NEXT_DIAG_GROUP and PATIENT_DIAG_GROUP. The datasets these models were trained on were compiled using the combinations of features identified as being most relevant for the prediction of the specific target variable within our CSF, gut bacteria, and sleep datasets.

E. Multi Modal Model Training

We trained a multi-modal model to combine the tabular data selected by the feature selection as described in section III-C and 3D brain scans to evaluate whether combining both can increase the predictive performance of our models. The reason for our decision to choose the feature set identified with the decision tree feature selection method and NEXT_DIAG_GROUP as target variable, even though this was not the tabular model with the best results, was the sample size. This combination gave us the biggest sample size, which we considered most important for model training. Data in all four data sources (sleep, gut bacteria, CSF, and MRI scans) needed to be available for consecutive visits, which was not given for all subjects. This specific combination gave us 516 observations to work with.

As seen in Fig. 9 in the appendix, the 3D brain scans are processed by five rounds of 3D convolution, 3D max pooling, and dropout layers, as commonly done in CNN models for image classification. The resulting feature volumes are then reduced to scalar features using a global average pooling operation. Then the number of features gets further reduced by a dense layer [18].

The tabular data is also processed by a dense layer before being combined with the output of the 3D brain scans as seen in Fig. 3. The combined features are then processed by another two dense layers with the relu activation function before being processed by a final dense layer using the softmax activation function.

F. Model Evaluation

To evaluate our models, we used the measurement precision (1). This metric tells us how many times a predicted MCI-AD transition by our model is correct. In practice this makes sure that a patient is not wrongly diagnosed with AD. To understand the overall quality of the model, we also took the F1-Score (3) into account. The F1-Score is the harmonic mean of precision and recall (2). Recall calculates the number of correctly predicted changes related to all observed changes.

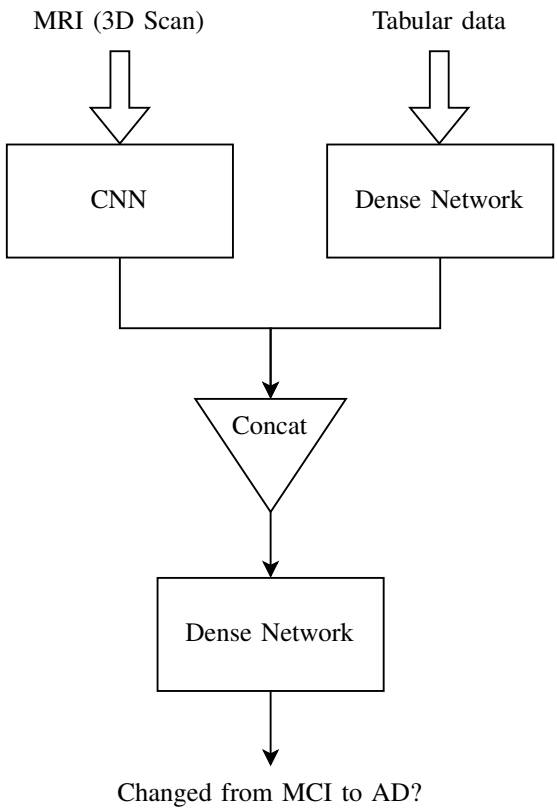

Fig. 3. Simplified view of the multi-modal classifier with 3D-image and tabular input

$$\text{precision} = \frac{TP}{TP + FP} \tag{1}$$

$$textrecall = \frac{TP}{TP + FN} \tag{2}$$

$$F_1 = 2 \cdot \frac{precision \cdot recall}{precision + recall} \tag{3}$$

To assess the plausibility of our results, we compared them to the results from other papers that also tried to predict diagnosis change based on tabular and image data [16], [18].

IV. Results and Discussion

A. Feature Selection

Feature selection generated nine sets of the most important features, one for each combination of our three feature selection methods and three target variables. Among these, there are no obvious clusters of repeating variables. Each of those nine combinations have different important features. The detailed feature selection results can be found in Table VI in the appendix.

B. Results on the Single Datasets

In the following section, we focus on answering our research questions one, two and three. Detailed results can be found in Table VII and Fig. 7 in the appendix.

1) Research Question 1: Our first research question is whether gut bacteria can be used as a predictor for AD status and transition from mild cognitive impairment (MCI) to AD. The Table II shows which F1-Score and precision was calculated per model and target group.

TABLE II
FIVE BEST RESULTS FOR GUT BACTERIA

Model	Target Group	Binary F1-Score	Precision
mlp	PATIENT_DIAG_GROUP	0.609	0.629
logistic regression	PATIENT_DIAG_GROUP	0.592	0.627
decision tree	PATIENT_DIAG_GROUP	0.467	0.610
knn	PATIENT_DIAG_GROUP	0.504	0.493
decision tree	PREV_DIAG_GROUP	0.372	0.302

The best two models have been able to predict changes from the first to last visit with a precision of 62.9% and 62.7%. Therefore, for this target group, data from gut bacteria can be used to predict a change. But only over the whole timeframe, from first to last visit. On a visit-by-visit basis, as seen in the decision tree model, the best result was achieved with a precision of 30.2%. However, this model and target group combination cannot be used, since 70% of the doctors' diagnoses would have been wrong.

2) Research Question 2: Our second research question is whether p-tau181, t-tau and β-Amyloid measured in cerebrospinal fluid (CSF) can be used as early predictors of AD. The Table III shows our results for the best model and target group combination.

TABLE III
FIVE BEST RESULTS FOR CSF

Model	Target Group	Binary F1-Score	Precision
logistic regression	PATIENT_DIAG_GROUP	0.581	0.720
knn	PATIENT_DIAG_GROUP	0.508	0.615
decision tree	PATIENT_DIAG_GROUP	0.321	0.474
mlp	PATIENT_DIAG_GROUP	0.375	0.444
logistic regression	PREV_DIAG_GROUP	0.393	0.306

Logistic regression and KNN have been able to predict changes from first to last visit with a precision of 72.0% and 61.5%. This indicates that these values can be used in combination with the two models to predict changes. But that's only for the overall diagnosis change. The visit-by-visit results are below 50% and cannot be used as an early predictor. Therefore, this research question cannot be confirmed.

3) Research Question 3: Our third research question is whether changes in sleep behavior can be used to predict changes from MCI to AD. Table IV shows our results for the best model and target group combination.

Again, the overall diagnosis change has been predicted best. The model used was a decision tree and has a precision of 81.8%. Next follows the change from the last to the current visit with a precision of 70.8% by MLP. Therefore, we conclude that changes in sleeping behavior can be used to predict diagnosis change.

TABLE IV
FIVE BEST RESULTS FOR SLEEP

Model	Target Group	Binary F1-Score	Precision
decision tree	PATIENT_DIAG_GROUP	0.818	0.818
mlp	PREV_DIAG_GROUP	0.723	0.708
knn	PREV_DIAG_GROUP	0.651	0.700
mlp	NEXT_DIAG_GROUP	0.333	0.667
mlp	PATIENT_DIAG_GROUP	0.636	0.636

The first two research questions could not be confirmed. Only predictions on changes from first to last visit had a high precision, but not on a visit-by-visit basis. Sleep does seem to generate useful predictions and 8 out of 12 trained models had a precision of at least 50%.

C. Results on Combined Dataset

To answer our fourth research question whether a combination of the factors sleep, gut bacteria and CSF can be used to predict a transition from MCI to AD, we trained models on these three. Table V shows our results for the best model, feature selection, and target group combination, with Fig. 4 visualizing them by various metrics. The detailed results can be found in Table VIII and Fig. 8 in the appendix.

TABLE V
FIVE BEST RESULTS FOR COMBINED DATASET

Model	Feature Selection	Target Group	Precision
knn	Decision Tree	PATIENT_DIAG_GROUP	0.656
logistic regression	Decision Tree	PATIENT_DIAG_GROUP	0.636
decision tree	Decision Tree	PATIENT_DIAG_GROUP	0.635
knn	Logistic Regression	PATIENT_DIAG_GROUP	0.619
mlp	Decision Tree	PATIENT_DIAG_GROUP	0.611

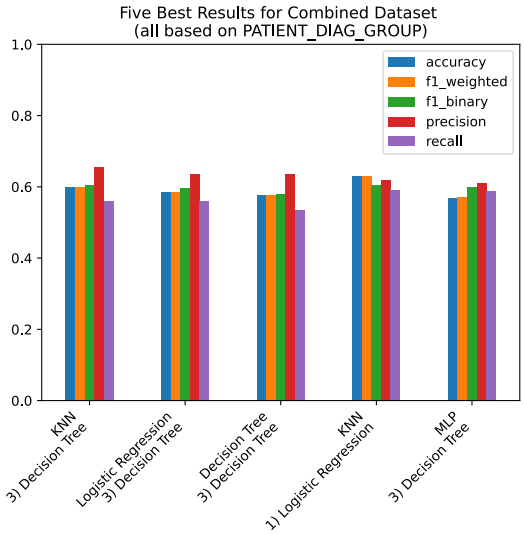

Fig. 4. Five Best Results for Combined Dataset

For predictions from first to last visit the best three results have been achieved by KNN, logistic regression and decision tree with precisions between 63.5% and 65.6%. It is noticeable that these top results are all based on feature selection by the decision tree. The best results for predictions on a visit-by-visit basis all have a precision below 50% and will therefore not be further discussed. Within each target group, it's difficult to determine why one model was better than the other. First, the precision is similar for the best models (2.1% difference for the top three models). Second, the top three results are achieved by three different models, with the fourth model being 2.4% behind the third. As noted, it is more important what model was used for feature selection.

D. Results from the Multi-Modal Model

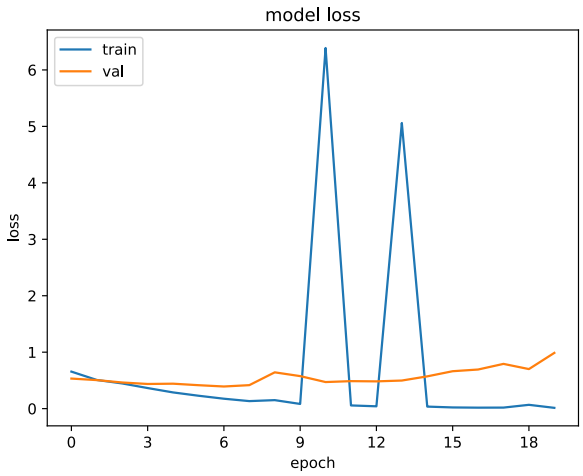

Fig. 5. Loss while training the Multi-Modal Model.

The multi-modal model was able to effectively integrate the tabular data and 3D brain scans. The model was trained on 516 observations and started over-fitting after 7 epochs as seen in Fig. 5. It achieved an accuracy of 77% and an F1-score of 82% with a precision of 72% and a recall of 97%. The model had no problem identifying AD transitions as evidenced by the absence of false negatives. However, the model predicted transitions to AD for 12 patients whose diagnosis remained MCI, shown by 12 false positives in a total of 23 actual negatives. This can be observed in Fig. 6. Judging by F1-score, these results are considerably better than the results from the tabular model approach with NEXT_DIAG_GROUP and the results generated by Kobivasan et al. [16] using only slices of brain scans, but still a bit lower than those obtained by a 3D-CNN of Payan & Montana [18].

V. Conclusion and Further Research

Based on our results, we draw the following conclusions and propose further research.

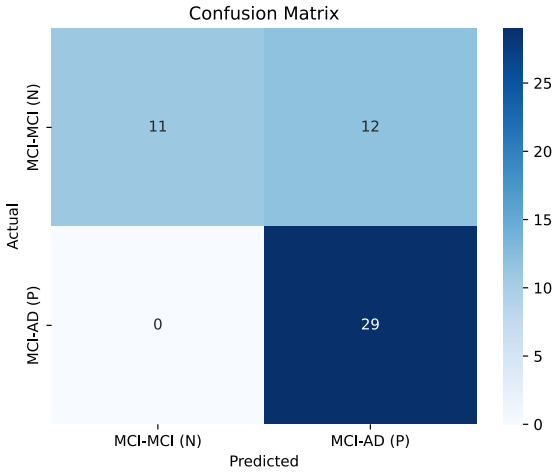

Fig. 6. Confusion Matrix showing the result of Multi-Modal Model.

A. Research Questions

As mentioned at the beginning, there are studies that showed that gut bacteria and CSF data can be used to predict Alzheimer's disease [6], [3], [7]. We therefore had expected better results. Perhaps they would have been seen on a longer timeframe, for example, not on a visit-by-visit basis, but year by year or two years. Sleep data showed strong results predicting a change. This should be further examined, for example for clusters in features.

Further, it was difficult to quantify the quality of the sleep data, as they are based on answers from caregivers of the patients. A questionnaire that focuses more on Alzheimer's disease could help improve the result.

The target group from the current visit to the next visit is somewhat unreliable because a doctor recognizes changes only at the next visit. Therefore, it could be interesting to see, if the current status (not change) would give any useful predictions.

B. Model on Combined Dataset

Our conclusion regarding the discovery of the overall target group PATIENT_DIAG_GROUP getting the best results is that a change in brain health does not happen within a few months between visits. Therefore, the timeframe between two visits is too short to predict changes with a high precision score. It was difficult to determine why one model performed better than the others within each target group, due to the complex nature of the feature selection.

C. Multi Modal Model

The use of a multi-modal model, combining tabular data with brain images, was effective in predicting the progression of Alzheimer's disease, as demonstrated by an F1-score of 0.82. This result was significantly higher than those obtained using either our tabular model or image-only models.

Given the high cost of MRI scans, it is important to evaluate the clinical utility and feasibility of such multimodal prediction models in the early diagnosis and treatment of Alzheimer's disease. Future research should aim to replicate and expand upon these findings, by visualizing the decision-making process and potentially incorporating additional sources of data, more complex models, and a larger sample size. The decision-making process could be visualized using Grad-CAM [19]. It is worth noting that our model consisted of nearly 3 million trainable parameters, but we only had 516 data points to work with, highlighting the need for a larger sample size to fully leverage the potential of the model.

Acknowledgment

We would like to express our appreciation to Prof. Dr. Arzu Çöltekin and Dr. Leticia Fernandez Moguel for their guidance and support throughout the course of this semester's challenge. The code review and constructive advice have been helpful in the development of this work and are deeply appreciated. Additionally, we would like to thank Dr. Azizi Seixas and Dr. Alberto R. Ramos for their medical insights, feedback, and new perspectives. We would like to extend our thanks to Dr. Marie-Thérèse Rudolf von Rohr and Dr. Caspar Battegay for their help and guidance in writing and reviewing of this paper.

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 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. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Therapeutic Research Institute at the University of Southern California. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

References

[1] B. f. G. BAG. "Demenz." (Sep. 16, 2022), [Online]. Available: https://www.bag.admin.ch/bag/de/home/krankheiten/krankheiten-im-ueberblick/demenz.html (visited on 11/23/2022).

[2] BASS, "Sterbefälle mit registrierter demenzerkrankung," BASS, Jan. 4, 2021. [Online]. Available: https://www.bag.admin.ch/dam/bag/de/dokumente/nat-gesundheitsstrategien/nationale-demenzstrategie/hf-daten/8_1_versorgungsmonitoring/FB1%20Sterbef%C3%A4lle%20mit%20Demenzerkrankung.pdf.download.pdf/Sterbef%C3%A4lle_mit_Demenzerkrankung_DE.pdf (visited on 01/07/2023).

[3] M. Bostancikliođlu, "The role of gut microbiota in pathogenesis of alzheimer's disease," *Journal of Applied Microbiology*, vol. 127, no. 4, pp. 954–967, 2019, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/jam.14264, ISSN: 1365-2672. DOI: 10.1111/jam.14264. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1111/jam.14264 (visited on 01/12/2023).

[4] M. B. Cooke, S. Catchlove, and K. L. Tooley, "Examining the influence of the human gut microbiota on cognition and stress: A systematic review of the literature," *Nutrients*, vol. 14, no. 21, p. 4623, Jan. 2022, Number: 21 Publisher: Multidisciplinary Digital Publishing Institute, ISSN: 2072-6643. DOI: 10.3390/nu14214623. [Online]. Available: https://www.mdpi.com/2072-6643/14/21/4623 (visited on 01/12/2023).

[5] J. F. Cryan, K. J. O'Riordan, C. S. M. Cowan, *et al.*, "The microbiota-gut-brain axis," *Physiological Reviews*, vol. 99, no. 4, pp. 1877–2013, Oct. 2019, Publisher: American Physiological Society, ISSN: 0031-9333. DOI: 10.1152/physrev.00018.2018. [Online]. Available: https://journals.physiology.org/doi/full/10.1152/physrev.00018.2018 (visited on 01/08/2023).

[6] K. Blennow, E. Vanmechelen, and H. Hampel, "CSF total tau, aβ42 and phosphorylated tau protein as biomarkers for alzheimer's disease," *Molecular Neurobiology*, vol. 24, no. 1, p. 87, Aug. 1, 2001, ISSN: 1559-1182. DOI: 10.1385/MN:24:1-3:087. [Online]. Available: https://doi.org/10.1385/MN:24:1-3:087 (visited on 01/12/2023).

[7] Y.-L. Wang, J. Chen, Z.-L. Du, *et al.*, "Plasma p-tau181 level predicts neurodegeneration and progression to alzheimer's dementia: A longitudinal study," *Frontiers in Neurology*, vol. 12, Sep. 7, 2021, ISSN: 1664-2295. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fneur.2021.695696 (visited on 01/12/2023).

[8] FDA. "FDA grants accelerated approval for alzheimer's disease treatment," FDA. Publisher: FDA. (Jun. 1, 2023), [Online]. Available: https://www.fda.gov/news-events/press-announcements/fda-grants-accelerated-

approval - alzheimers - disease - treatment (visited on 01/07/2023).

[9] Y.-E. S. Ju, B. P. Lucey, and D. M. Holtzman, "Sleep and alzheimer disease pathology—a bidirectional relationship," *Nature reviews. Neurology*, vol. 10, no. 2, pp. 115–119, Feb. 2014, ISSN: 1759-4758. DOI: 10.1038/nrneurol.2013.269. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3979317/ (visited on 11/23/2022).

[10] A. P. Mecca, H. R. Michalak, J. W. McDonald, *et al.*, "Sleep disturbance and the risk of cognitive decline or clinical conversion in the ADNI cohort," *Dementia and Geriatric Cognitive Disorders*, vol. 45, no. 3, pp. 232–242, 2018, ISSN: 1421-9824. DOI: 10.1159/000488671.

[11] D. Shea, E. Colasurdo, A. Smith, *et al.*, "SOBA: Development and testing of a soluble oligomer binding assay for detection of amyloidogenic toxic oligomers," *Proceedings of the National Academy of Sciences*, vol. 119, no. 50, e2213157119, Dec. 13, 2022, Publisher: Proceedings of the National Academy of Sciences. DOI: 10.1073/pnas.2213157119. [Online]. Available: https://www.pnas.org/doi/10.1073/pnas.2213157119 (visited on 01/12/2023).

[12] I. M. Colrain, "Sleep and the brain," *Neuropsychology Review*, vol. 21, no. 1, pp. 1–4, Mar. 1, 2011, ISSN: 1573-6660. DOI: 10.1007/s11065-011-9156-z. [Online]. Available: https://doi.org/10.1007/s11065-011-9156-z (visited on 12/05/2022).

[13] ADNI. "ADNI — about." (Jan. 1, 2017), [Online]. Available: https://adni.loni.usc.edu/about/ (visited on 11/23/2022).

[14] ADNI. "ADNI — MRI pre-processing." (2017), [Online]. Available: https://adni.loni.usc.edu/methods/mri-tool/mri-pre-processing/ (visited on 01/08/2023).

[15] ADNI. "ADNI — standardized MRI data sets." (2017), [Online]. Available: https://adni.loni.usc.edu/methods/mri-tool/standardized-mri-data-sets/ (visited on 01/08/2023).

[16] K. Kobivasan, C. Strasser, and J. Weibel, "Why healthy people stay healthy," 2022.

[17] A. Hoopes, J. S. Mora, A. V. Dalca, B. Fischl, and M. Hoffmann, "SynthStrip: Skull-stripping for any brain image," *NeuroImage*, vol. 260, p. 119 474, Oct. 15, 2022, ISSN: 1053-8119. DOI: 10.1016/j.neuroimage.2022.119474. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1053811922005900 (visited on 01/08/2023).

[18] A. Payan and G. Montana, *Predicting alzheimer's disease: A neuroimaging study with 3d convolutional neural networks*, Feb. 9, 2015. DOI: 10.48550/arXiv.1502.02506. arXiv: 1502.02506[cs,stat]. [Online]. Available: http://arxiv.org/abs/1502.02506 (visited on 01/10/2023).

[19] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, and A. Torralba, *Learning deep features for discriminative localization*, Dec. 13, 2015. DOI: 10.48550/arXiv.1512.04150. arXiv: 1512.04150[cs]. [Online]. Available: http://arxiv.org/abs/1512.04150 (visited on 01/17/2023).

APPENDIX A

The code used to obtain the results in this paper is freely available at the following link: https://gitlab.fhnw.ch/brain-health/analysis/. The code is open-source and available for use and modification under the MIT License.

APPENDIX B

TABLE VI
RESULTS OF OUR 3 FEATURE SELECTION METHODS

Method	Target Variable	Dataset	Selected Features
1) Logistic Regression	NEXT_DIAG_GROUP	gut bacteria	UCA, X3_HYDROXYISOVALERIC_ACID, ADIPIC_ACID, X7_KETOLCA, HDCA, MUROCA, X2_HYDROXYBUTYRIC_ACID, GUDCA, METHYLSUCCINIC_ACID, L_METHIONINE, C19_2_CIS_10_13
		sleep	NPIK5, NPIK8, NPIK7, NPIK6
		csf	PTAU, ABETA
	PATIENT_DIAG_GROUP	gut bacteria	C20_4_CIS_5_8_11_14, C16_1_CIS_9, C20_3_CIS_8_11_14, X3_INDOLEPROPIONIC_ACID, HDCA, L_ASPARTIC_ACID, C17_1_CIS_10, NORCA, BETA_ALANINE, CITRIC_ACID, C22_5_CIS_4_7_10_13_16
		sleep	NPIK5, NPIK2, NPIK7, NPIK9A
		csf	TAU, ABETA
	PREV_DIAG_GROUP	gut bacteria	UCA, TCA, X7_KETOLCA, MUROCA, C11_0, TDCA, GLYCINE, X7_DHCA, ADIPIC_ACID, NORDCA, C20_4_CIS_5_8_11_14
		sleep	NPIK3, NPIK4, NPIK8, NPIKTOT
		csf	TAU, PTAU
2) Backward Elimination	NEXT_DIAG_GROUP	gut bacteria	X12_KETOLCA, X3_METHYL_2_OXOVALERIC_ACID, SUBERIC_ACID, DEHYDROLCA, C20_5_CIS_5_8_11_14_17, GLYCOLIC_ACID, C1_0, CITRIC_ACID, L_METHIONINE, L_ASPARTIC_ACID, L_SERINE
		sleep	NPIK2, NPIK9C, NPIK9A, NPIK1
		csf	ABETA, PTAU
	PATIENT_DIAG_GROUP	gut bacteria	GUDCA, L_ALPHA_AMINOBUTYRIC_ACID, C9_0, GCA, C16_1_CIS_9, NORCA, C20_2_CIS_11_14, GCDCA, L_ASPARTIC_ACID, L_SERINE, L_ASPARAGINE
		sleep	NPIK9B, NPIK9C, NPIK9A
		csf	ABETA, PTAU
	PREV_DIAG_GROUP	gut bacteria	X3_METHYL_2_OXOVALERIC_ACID, HDCA, MUROCA, DCA, ISOCITRIC_ACID, C22_5_CIS_7_10_13_16_19, TCA, BETA_ALANINE, GLYCINE, TCDCA, L_ASPARAGINE
		sleep	NPIK9B, NPIK9A
		csf	TAU, PTAU
3) Decision Tree	NEXT_DIAG_GROUP	gut bacteria	L_SERINE, X_2_METHYLPENTANOIC_ACID, SUBERIC_ACID, X3_INDOLEPROPIONIC_ACID, UDCA, X3_HYDROXYISOVALERIC_ACID, L_ALANINE, INDOLEACETIC_ACID, HDCA, GUDCA, NORDCA
		sleep	NPIK5, NPIK4, NPIK7, NPIK9A
		csf	ABETA, PTAU
	PATIENT_DIAG_GROUP	gut bacteria	APOCA, L_ASPARTIC_ACID, INDOLEACETIC_ACID, X3_HYDROXYBUTYRIC_ACID, HDCA, C6_0, GLYCINE, C9_0, C18_0, L_LEUCINE, C20_3_CIS_8_11_14
		sleep	NPIK9B, NPIK3, NPIK5, NPIK2
		csf	PTAU, TAU
	PREV_DIAG_GROUP	gut bacteria	L_PROLINE, L_ALANINE, LCA, NORDCA, GCA, MUROCA, TCA, GHDCA, C18_2_CIS_9_12, ISOBUTYRIC_ACID, X_2_METHYLPENTANOIC_ACID
		sleep	NPIK9B, NPIK7, NPIK9C, NPIK4
		csf	ABETA, TAU

APPENDIX C

TABLE VII
RESULTS OF MODELS TRAINED ON SINGLE DATASETS

Method	Dataset	Target Variable	Accuracy	F1-score weighted	F1-score binary	Precision	Recall	Number of test samples
Decision tree	CSF	NEXT_DIAG_GROUP	0.779	0.796	0.118	0.100	0.143	68
		PATIENT_DIAG_GROUP	0.415	0.393	0.321	0.474	0.243	65
		PREV_DIAG_GROUP	0.551	0.585	0.286	0.222	0.400	89
	Gut bacteria	NEXT_DIAG_GROUP	0.563	0.604	0.250	0.190	0.366	206
		PATIENT_DIAG_GROUP	0.544	0.531	0.467	0.610	0.379	125
		PREV_DIAG_GROUP	0.581	0.606	0.372	0.302	0.485	258
	Sleep behavior	NEXT_DIAG_GROUP	0.593	0.650	0.294	0.200	0.556	59
		PATIENT_DIAG_GROUP	0.833	0.833	0.818	0.818	0.818	48
		PREV_DIAG_GROUP	0.779	0.787	0.691	0.594	0.826	77
K-nearest neighbors	CSF	NEXT_DIAG_GROUP	0.824	0.810	0.000	0.000	0.000	68
		PATIENT_DIAG_GROUP	0.523	0.521	0.508	0.615	0.432	65
		PREV_DIAG_GROUP	0.663	0.649	0.167	0.188	0.150	89
	Gut bacteria	NEXT_DIAG_GROUP	0.670	0.670	0.171	0.171	0.171	206
		PATIENT_DIAG_GROUP	0.464	0.463	0.504	0.493	0.515	125
		PREV_DIAG_GROUP	0.694	0.655	0.202	0.303	0.152	258
	Sleep behavior	NEXT_DIAG_GROUP	0.797	0.771	0.143	0.200	0.111	59
		PATIENT_DIAG_GROUP	0.542	0.539	0.476	0.500	0.455	48
		PREV_DIAG_GROUP	0.805	0.801	0.651	0.700	0.609	77
Logistic regression	CSF	NEXT_DIAG_GROUP	0.574	0.657	0.216	0.133	0.571	68
		PATIENT_DIAG_GROUP	0.600	0.597	0.581	0.720	0.486	65
		PREV_DIAG_GROUP	0.618	0.648	0.393	0.306	0.550	89
	Gut bacteria	NEXT_DIAG_GROUP	0.587	0.621	0.234	0.186	0.317	206
		PATIENT_DIAG_GROUP	0.592	0.592	0.592	0.627	0.561	125
		PREV_DIAG_GROUP	0.605	0.619	0.320	0.286	0.364	258
	Sleep behavior	NEXT_DIAG_GROUP	0.576	0.633	0.194	0.136	0.333	59
		PATIENT_DIAG_GROUP	0.604	0.601	0.627	0.552	0.727	48
		PREV_DIAG_GROUP	0.636	0.650	0.500	0.424	0.609	77
Multilayer perceptron	CSF	NEXT_DIAG_GROUP	0.779	0.786	0.000	0.000	0.000	68
		PATIENT_DIAG_GROUP	0.385	0.383	0.375	0.444	0.324	65
		PREV_DIAG_GROUP	0.640	0.646	0.238	0.227	0.250	89
	Gut bacteria	NEXT_DIAG_GROUP	0.743	0.733	0.293	0.324	0.268	206
		PATIENT_DIAG_GROUP	0.600	0.600	0.609	0.629	0.591	125
		PREV_DIAG_GROUP	0.694	0.655	0.202	0.303	0.152	258
	Sleep behavior	NEXT_DIAG_GROUP	0.864	0.834	0.333	0.667	0.222	59
		PATIENT_DIAG_GROUP	0.667	0.667	0.636	0.636	0.636	48
		PREV_DIAG_GROUP	0.831	0.832	0.723	0.708	0.739	77

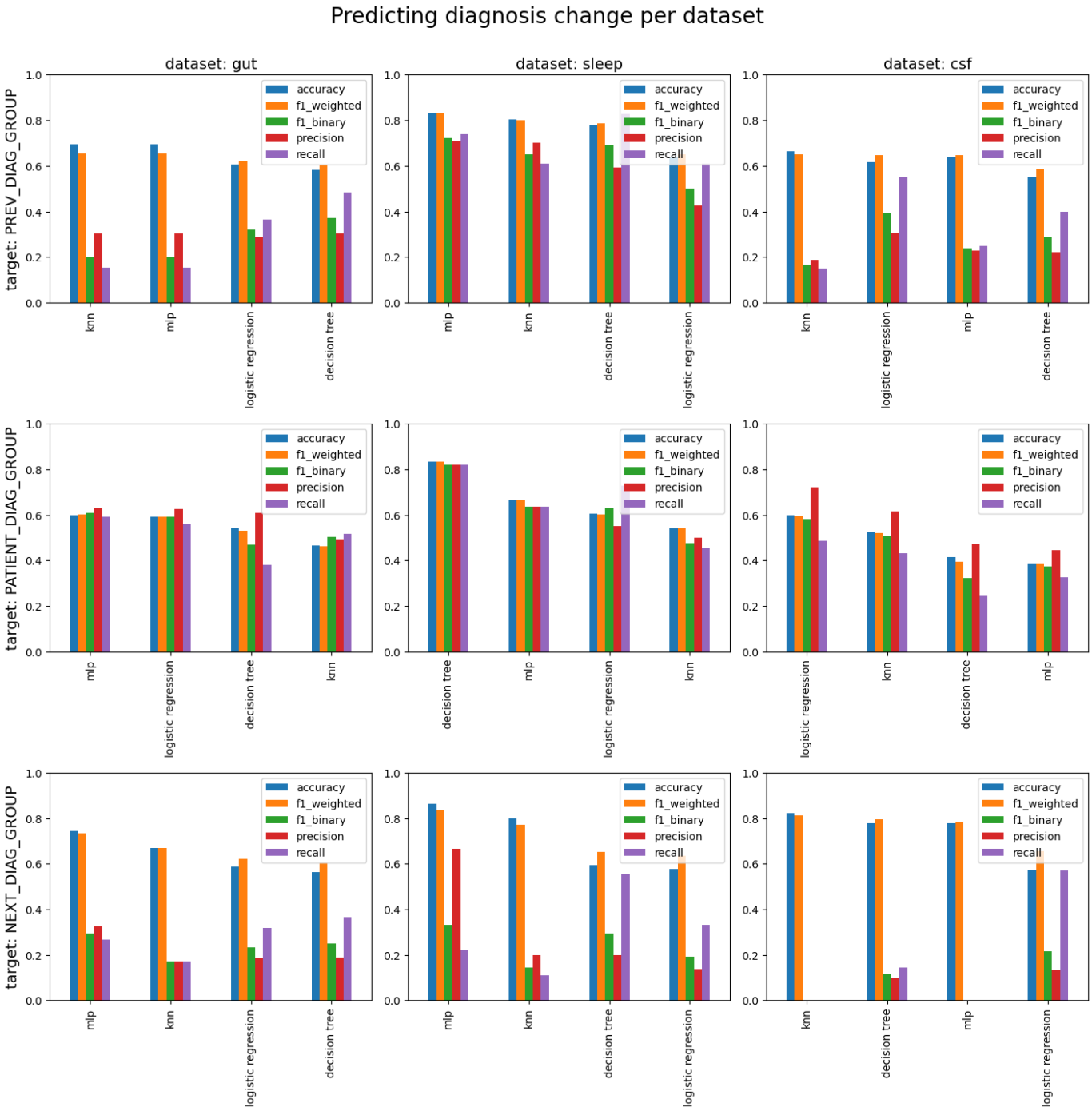

Fig. 7. Detailed results per dataset

TABLE VIII
RESULTS OF MODELS TRAINED ON COMBINED DATASETS USING FEATURE SELECTION METHODS

Method	Feature Selection Method	Target Variable	Accuracy	F1-score weighted	F1-score binary	Precision	Recall	Number of test samples
Decision tree	Decision Tree	NEXT_DIAG_GROUP	0.671	0.680	0.250	0.232	0.271	237
		PATIENT_DIAG_GROUP	0.577	0.577	0.580	0.635	0.533	137
		PREV_DIAG_GROUP	0.713	0.720	0.411	0.383	0.443	310
	Logistic Regression	NEXT_DIAG_GROUP	0.633	0.655	0.256	0.217	0.313	237
		PATIENT_DIAG_GROUP	0.522	0.515	0.441	0.500	0.394	138
		PREV_DIAG_GROUP	0.646	0.659	0.282	0.253	0.318	302
	Backward Elimination	NEXT_DIAG_GROUP	0.690	0.697	0.270	0.254	0.288	261
		PATIENT_DIAG_GROUP	0.574	0.574	0.583	0.532	0.646	141
		PREV_DIAG_GROUP	0.741	0.737	0.341	0.356	0.328	313
K-nearest neighbors	Decision Tree	NEXT_DIAG_GROUP	0.662	0.654	0.111	0.119	0.104	237
		PATIENT_DIAG_GROUP	0.599	0.599	0.604	0.656	0.560	137
		PREV_DIAG_GROUP	0.713	0.711	0.350	0.358	0.343	310
	Logistic Regression	NEXT_DIAG_GROUP	0.717	0.711	0.264	0.279	0.250	237
		PATIENT_DIAG_GROUP	0.630	0.630	0.605	0.619	0.591	138
		PREV_DIAG_GROUP	0.758	0.750	0.397	0.436	0.364	302
	Backward Elimination	NEXT_DIAG_GROUP	0.686	0.678	0.163	0.174	0.154	261
		PATIENT_DIAG_GROUP	0.553	0.554	0.540	0.514	0.569	141
		PREV_DIAG_GROUP	0.738	0.745	0.406	0.378	0.438	313
Logistic regression	Decision Tree	NEXT_DIAG_GROUP	0.561	0.603	0.297	0.220	0.458	237
		PATIENT_DIAG_GROUP	0.584	0.585	0.596	0.636	0.560	137
		PREV_DIAG_GROUP	0.713	0.720	0.411	0.383	0.443	310
	Logistic Regression	NEXT_DIAG_GROUP	0.586	0.620	0.246	0.195	0.333	237
		PATIENT_DIAG_GROUP	0.471	0.471	0.459	0.449	0.470	138
		PREV_DIAG_GROUP	0.702	0.700	0.308	0.313	0.303	302
	Backward Elimination	NEXT_DIAG_GROUP	0.563	0.607	0.360	0.254	0.615	261
		PATIENT_DIAG_GROUP	0.518	0.515	0.541	0.482	0.615	141
		PREV_DIAG_GROUP	0.671	0.690	0.335	0.286	0.406	313
Multilayer perceptron	Decision Tree	NEXT_DIAG_GROUP	0.675	0.665	0.135	0.146	0.125	237
		PATIENT_DIAG_GROUP	0.569	0.570	0.599	0.611	0.587	137
		PREV_DIAG_GROUP	0.719	0.716	0.356	0.369	0.343	310
	Logistic Regression	NEXT_DIAG_GROUP	0.709	0.691	0.169	0.200	0.146	237
		PATIENT_DIAG_GROUP	0.529	0.529	0.511	0.507	0.515	138
		PREV_DIAG_GROUP	0.762	0.760	0.446	0.453	0.439	302
	Backward Elimination	NEXT_DIAG_GROUP	0.724	0.713	0.234	0.262	0.212	261
		PATIENT_DIAG_GROUP	0.525	0.525	0.511	0.486	0.538	141
		PREV_DIAG_GROUP	0.732	0.737	0.382	0.361	0.406	313

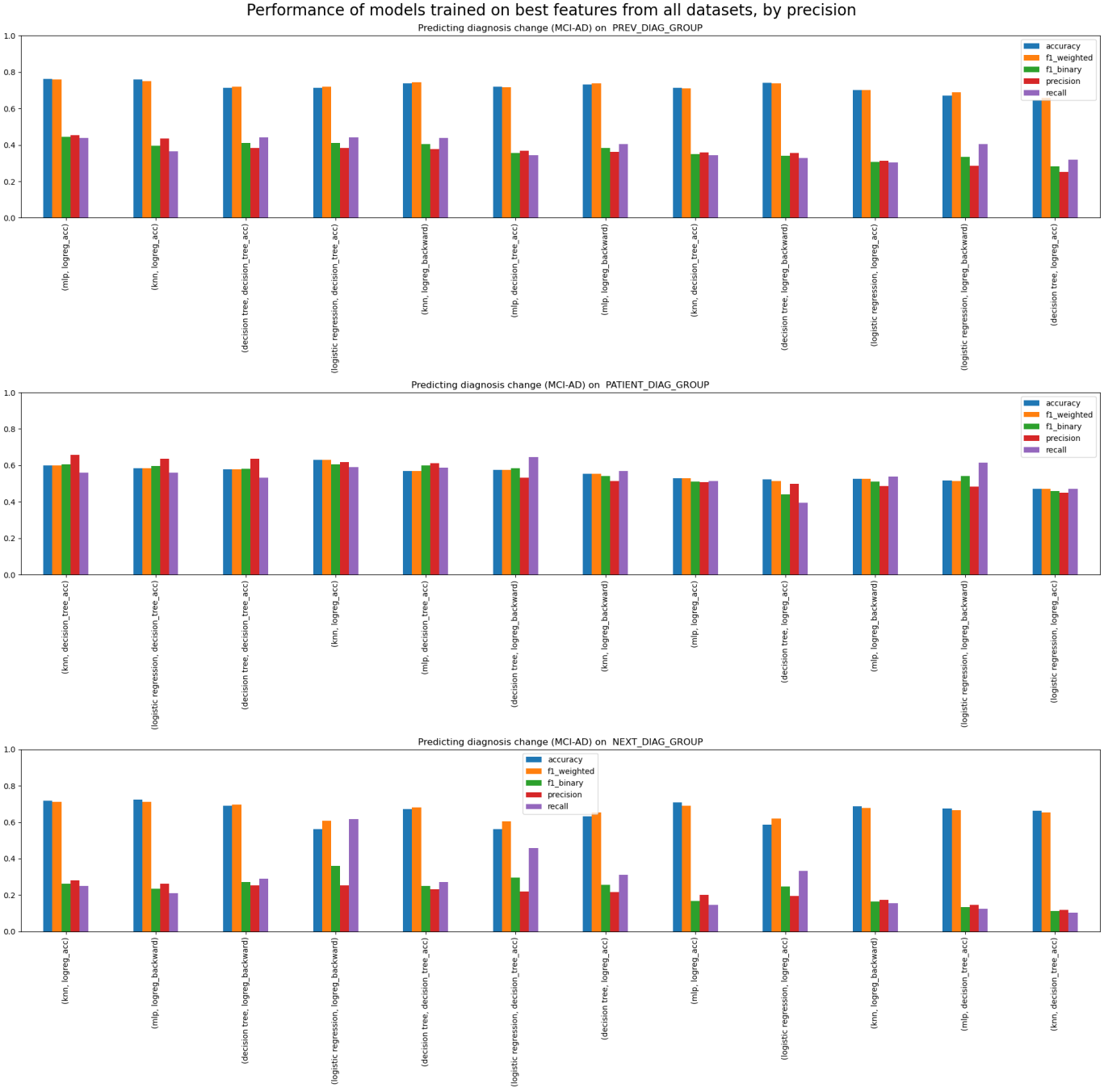

Fig. 8. Detailed results for combined dataset

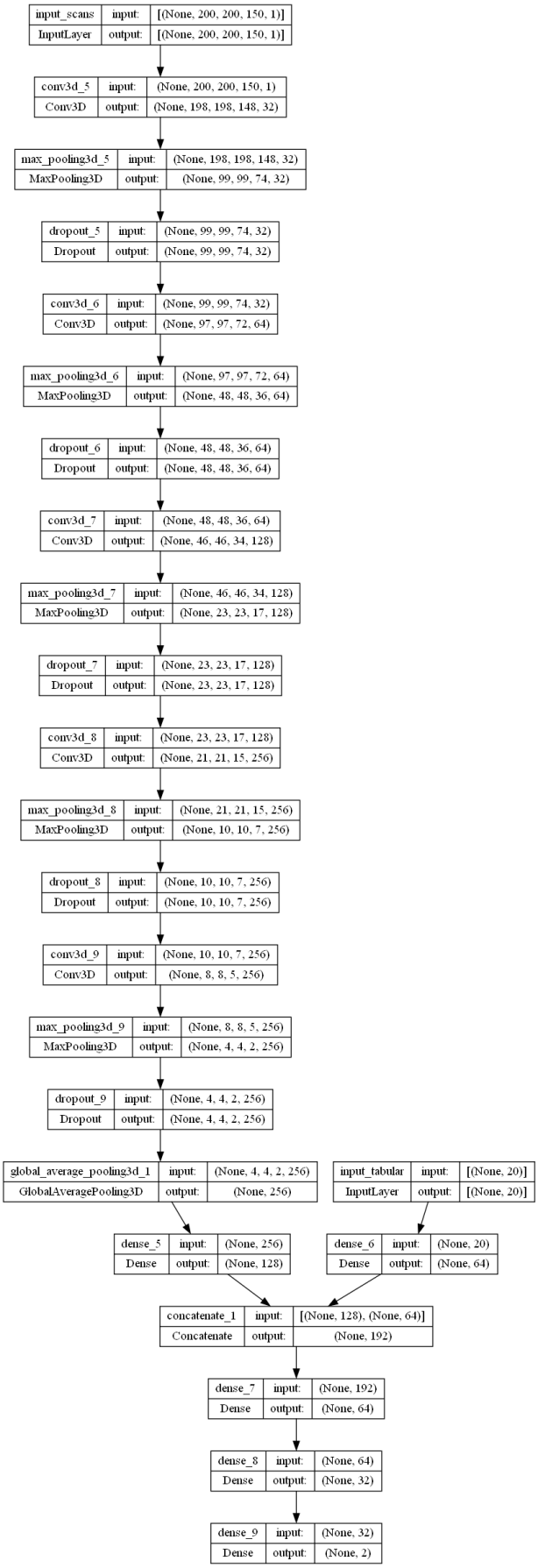

Fig. 9. Architecture of the multi-modal classifier with 3D-image and tabular input