2023 45th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC) | 979-8-3503-2447-1/23/$31.00 ©2023 IEEE | DOI: 10.1109/EMBC40787.2023.10341000

GIRUS-net: A Multimodal Deep Learning Model Identifying Imaging and Genetic Biomarkers Linked to Alzheimer's Disease Severity

Sarah Wu[1], Archana Venkataraman[1,2], Sayan Ghosal[1**]
*for the Alzheimer's Disease Neuroimaging Initiative**
1 Department of Electrical and Computer Engineering, Johns Hopkins University, Baltimore, MD, USA
2 Department of Electrical and Computer Engineering, Boston University, Boston, MA, USA
** Correspondences should be directed to sghosal3@jhu.edu

***Abstract*— We introduce an explainable deep neural architecture that combines brain structure with genetic influence to improve disease severity prediction in Alzheimer's disease. Our framework consists of an encoder, a decoder, and a rank-consistent ordinal regression module. The encoder projects neural imaging and genetics data into a low-dimensional latent space regularized by the decoder. The ordinal regression module guides the feature embedding process to find discriminative patterns representative of disease severity. We also add a learnable dropout layer that learns feature importance and extracts explainable biomarkers from the data. We evaluate our model using structural MRI (sMRI) and Single Nucleotide Polymorphism (SNP) data provided by the Alzheimer's Disease Neuroimaging Initiative (ADNI) database. In 2-class severity classification comparison, our model has a median F-score of 0.86 (baseline median F-score range: 0.57-0.81). In 3-class classification comparison, our model's median F-score is 0.50 (baseline range: 0.17 - 0.41). In 4-class classification comparison, our model's median F-score is 0.40 (baseline range: 0.14 - 0.39). We demonstrate that our model provides improved disease diagnosis alongside sparse and clinically relevant biomarkers.**

***Clinical relevance*—This study provides a deep-learning model that can predict Alzheimer's disease severity levels while identifying consistent and clinically relevant biomarkers.**

I. Introduction

Alzheimer's disease (AD) is a neurodegenerative disorder common in the elderly population [1]. Patients develop mild cognitive impairment, which progresses to dementia. AD is characterized by gradual loss of brain cells, also known as brain atrophy, which can be detected through structural magnetic resonance imaging (sMRI) [2]. Genetic factors also play a significant role in disease development [3] and progression [4]. Genetic risk factors, such as single nucleotide polymorphisms (SNPs), help pinpoint mutations in the DNA that influence the pathophysiology [5] of AD. Most research disentangles AD mechanisms by studying the neural influence and genetic factors separately. However, separating the data modalities may provide an incomplete picture of the underlying biological process [6].

Imaging-genetics studies integrate neuroimaging and genetic data to improve disease prediction [7]. Imaging features are often derived from structural and functional MRI (s/fMRI), and genetic variants are typically captured by SNPs. Data-driven imaging-genetics methods can be grouped into four main categories: simple regression, nonlinear methods, correlation methods, and deep learning approaches. The first category uses linear models like SVM [8], [9] and Logistic Regression [10] for AD classification. However, these methods typically train on single data modality and fail to discover interactions between modalities. The second category leverages gradient boosting [11], [12] and decision trees [13] for multi-class classification. These models can encode nonlinear interactions between the features, but they fail to disentangle the neuroimaging and genetic pathways linked to AD. The third category uses correlation analysis to identify associations between genetic variations and quantitative traits [14], [15], [16]. However, these models do not incorporate clinical diagnosis directly. Thus, the biomarkers obtained through such analysis may not align with the predictive group differences. The last category relies on deep learning architectures to combine high dimensional, structured data for imaging-genetic analysis [17]. Deep learning frameworks are highly complex, lacking model explainablility. Recent work such as the Genetic and Multimodal Imaging data using Neural-network Designs (G-MIND) framework can identify predictive biomarkers of a disease from imaging and genetic modalities [18]. However, G-MIND performs classification and cannot accommodate the progression of AD severity.

We introduce a novel framework to combine **G**enetic and **I**maging data using **R**ank-consistent m**U**ltimodal multicla**S**s network (GIRUS-net) that identifies neuroimaging and genetics biomarkers for AD diagnosis [18]. This work extends the G-MIND model and uses a rank-consistent ordinal regression module [19] to track disease severity from imaging and genetics data. Thus, GIRUS-net can identify biomarkers that are associated with the progressive representation of disease severity. GIRUS-net consists of an autoencoder coupled with an ordinal regression module. The encoders combine the imaging and genetics features into latent space and pass it through the ordinal regression module for disease severity prediction. We introduce a binary mask with binary concrete prior [20] as feature selection layer for biomarker detection. On a population study of AD, GIRUS-net yields sparser and more consistent biomarkers than baselines methods, while maintaining competitive classification performance.

*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 provided data but did not participate in analysis or writing of this report.

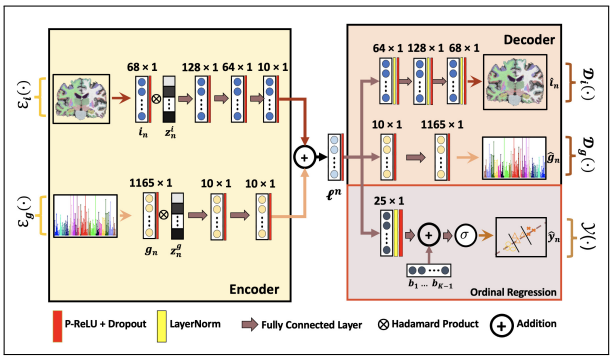

Fig. 1. GIRUS-net architecture. The inputs $\mathbf{i}_n$ and $\mathbf{g}_m$ correspond to the imaging and genetic modalities. $\mathcal{E}(\cdot)$ and $\mathcal{D}(\cdot)$ capture the encoding and decoding operations, and $\mathcal{Y}(\cdot)$ captures the ordinal regression operation. ℓ^n is the combined latent representation. $\mathbf{z}_n^i$, $\mathbf{z}_n^g$ are the learnable dropout mask for imaging and genetic features, respectively. $b_1 \cdots b_{K-1}$ are the bias terms learned during ordinal regression that ensures rank consistency.

II. Methods

Figure 1 illustrates our GIRUS-net framework. The inputs are the genotype data $\mathbf{g}_n \in \mathbf{R}^{G\times 1}$ for each subject n and the corresponding imaging features $\mathbf{i}_n \in \mathbf{R}^{I\times 1}$. The class labels $y_n \in \{1,2,3,4\}$ corresponds to different severity of AD: cognitive normal (CN), early mild cognitive impairment (EMCI), late mild cognitive impairment (LMCI), and mild Alzheimer's disease (AD). The diagnosis (phenotype) y_n is known during training but not during testing.

A. Encoder-Decoder Framework

We jointly model the imaging and genetic data using an autoencoder, coupled with an ordinal regression module.

a) Bayesian Feature Selection: The first layer of the encoder incorporates Bayesian feature selection using a learnable dropout layer. Unlike Bernoulli dropout where the underlying probability is fixed *a priori*, here we parameterize the dropout layer using Gumbell-Softmax distribution. The reparameterization trick relaxes the standard binary dropout to a continuous representation, which allows us to learn the posterior probability of the binary vectors. Mathematically, the subject specific dropout masks $\mathbf{z}_n^m$ are generated as:

$$\mathbf{z}_n^m = \sigma\Big(\frac{1}{t}\big(\log(\mathbf{p}^m)-\log(1-\mathbf{p}^m)+\log(\mathbf{u}_n^m)-\log(1-\mathbf{u}_n^m)\big)\Big). \tag{1}$$

where m indexes the data modality, $\mathbf{p}^m$ represents the underlying importance map, t captures the extent of relaxation from Bernoulli dropout, and $\mathbf{u}_n^i$ is a random vector sampled from $Uniform(0,1)$ for stochasticity. During each forward pass and for every subject n, the network randomly samples $\mathbf{z}_n^i$ for imaging and $\mathbf{z}_n^g$ for genetics, respectively. The continuous representation of the dropout masks allow us to learn the underlying importance maps during training. We also incorporate a KL divergence loss $KL(Ber(q)\|Ber(\mathbf{p}^m))$ to enforce sparsity in $\mathbf{p}^m$. As seen in Eq. (1), higher values in $\mathbf{p}^m$ are representative of the most selected features and can be identified as potential biomarkers.

b) Multimodal Latent Fusion and Decoder: The imaging and genetic features are passed through the learnable dropout layers to two separate encoders to obtain latent embeddings. These embeddings are then fused to leverage the common structure shared between both modalities. Mathematically, the fusion operation is

$$\ell_n = \frac{1}{2}\big(\mathcal{E}_i(\mathbf{i}_n, \mathbf{z}_n^i) + \mathcal{E}_g(\mathbf{g}_n, \mathbf{z}_n^g)\big) \tag{2}$$

where $\mathcal{E}_i(\cdot)$, $\mathcal{E}_g(\cdot)$ denote the encoding operations for imaging and genetics. After fusion, the latent vectors ℓ_n are passed through the decoders $\mathcal{D}_i(\cdot)$ and $\mathcal{D}_g(\cdot)$ to reconstruct the imaging and genetic data, ensuring that information is preserved during encoding. The reconstruction loss is an L_2 loss between the input and the reconstructed outputs:

$$\frac{\lambda_1}{B}\sum_{n=1}^{N}\|\mathbf{i}_n - \mathcal{D}_i(\ell^n)\|_2^2 + \frac{\lambda_2}{B}\sum_{n=1}^{N}\|\mathbf{g}_n - \mathcal{D}_g(\ell^n)\|_2^2 \tag{3}$$

where B is the batch size, and λ_1, λ_2 capture relative contribution of the loss terms.

B. Rank Consistent Ordinal Regression

Our auto-encoder is coupled with an ordinal regression module to predict the level of disease severity. The regression module ensures that the latent embeddings and the dropout masks learn discriminative information from the data.

a) Rank Consistent Prediction: Our regression module consists of a sequence of fully connected layers which predicts the disease severity level from the latent embeddings. Mathematically, the prediction probabilities are calculated as

$$P(\hat{y}_n = k) = \sigma(\mathcal{Y}(\ell_n; \mathbf{W}) + b_k) \tag{4}$$

where $\hat{y}_n$ is the predicted class label, $\mathcal{Y}(\cdot)$ is the fully connected layers parameterized by $\mathbf{W}$, and b_k are separate biases associated with each severity level. To ensure rank consistency among prediction probabilities (i.e., $P(\hat{y}_n = k) > P(\hat{y}_n = K+1)$), we need to ensure that $b_k > b_{k+1}$. Previously, the work of [19] has shown that rank consistency can be achieved using multi-label cross entropy loss:

$$\begin{aligned}\mathbf{L}_{ordinal} = &\sum_{n=1}^{B}\sum_{k=0}^{K-1} y_n^{(k)}\log\left(P(\hat{y}_n = k)\right) \\ &+ (1 - y_n^{(k)})\log\left(1 - P(\hat{y}_n = k)\right) \\ &\text{where } y_n^{(k)} = 1 \text{ if } k <= y_n\end{aligned} \tag{5}$$

where B is batch size, $y_n^{(k)}$ is a binary multi-class vector generated from y_n, and $\hat{y}_n$ is the predicted label.

Combining Eqs. (1-5), the GIRUS-net loss function is:

$$\begin{aligned}\mathcal{L} = &\frac{\lambda_1}{B}\sum_{n=1}^{B}\|\mathbf{i}_n - \mathcal{D}_i(\ell^n)\|_2^2 + \frac{\lambda_2}{B}\sum_{n=1}^{B}\|\mathbf{g}_n - \mathcal{D}_g(\ell^n)\|_2^2 \\ &+ \frac{\lambda_3}{B}\cdot\mathbf{L}_{ordinal} + \lambda_4\sum_{m\in\{i,g\}}\theta_m(KL(Ber(q)|Ber(\mathbf{p}^m)))\end{aligned} \tag{6}$$

where λ_1, λ_2 capture the contribution of the reconstruction losses, λ_3 controls the contribution of ordinal loss and θ_m captures the relative contribution of the sparsity penalties.

b) Prediction on New Data: During testing, the imaging and genetic data are multiplied by the dropout probabilities and passed through the encoder. The latent encoding then pass through the ordinal regression module for disease severity prediction. Disease severity is predicted by

$$\hat{y}_{test} = k \text{ if } P(\hat{y}_{test} = k) > 0.5 \text{ and } P(\hat{y}_{test} = k+1) < 0.5. \quad (7)$$

C. Implementation and Parameter Sweep

The model parameters, $\{\lambda_1, \lambda_2, \lambda_3, \lambda_4\}$, are selected so that individual loss terms lie within the same order of magnitude. This criterion is model agnostic and does not require us to optimize them. We weighted each sparsity penalty by θ_m to adjust for difference in number of features: $\theta_i = \frac{\text{number of Genetic Features}}{\text{number of Imaging Features}}$, and $\theta_g = 1$. The learning rate and batch size are fixed based on validation performance in a 10-fold cross validation setting. We perform grid search over learning rate $\in [0.00001, 0.001]$ and batch size $\in [8, 128]$. For all experiments, we fixed the Bernoulli probability to $q = 0.0001$, temperature $t = 0.1$, and batch size = 32. Model parameters were set to $\lambda_1 = 0.0001, \lambda_2 = 0.001, \lambda_3 = 0.5$, and $\lambda_4 = 0.0001$, where $\lambda_1, \lambda_2, \lambda_3$ were scaled to be on the same scale as sparsity penalty. Learning rate is 0.001 for 2-class and 0.0005 for 3-class and 4-class experiments. Fig. 1 shows additional architecture details.

D. Baseline Methods

We compare GIRUS-net with four standard baseline models that operate on the concatenated data modalities, i.e. $\mathbf{x} = [\mathbf{i}^T, \mathbf{g}^T]$. Hyperparameters are fixed using a grid search approach in a 10-fold cross validation setting.

a) Random Forest Classifier (RF): Random Forest is an ensemble method that constructs decision trees and average their outputs to provide a robust and accurate prediction. Feature importance is calculated as mean decrease in information gain [21] associated with each feature.

b) Support Vector Machine (SVM): Support vector machines create a hyperplane that maximally separates the data belonging to two different classes. Here, we extend the linear SVM by building $\frac{K(K-1)}{2}$ separate binary "one vs. one" classifiers to perform multi-class prediction [22]. We construct the feature importance map by taking the mean of the absolute values of weights of the linear kernels.

c) Artificial Neural Network (ANN): We train an ANN to perform classification based on input $\mathbf{x}$. ANNs can model complex patterns from the data, but usually lack feature explainability, particularly for deeper networks. Thus, the feature importance maps are calculated *post hoc* using Shapley Additive Explanations [23].

d) Ordered Logit Model (O-Logit): O-Logit is a generalized linear model that performs ordinal regression [24]. Here, we learn a linear layer of weights $\mathbf{w}$ and bias $b_1 \cdots b_{K-1}$. Similar to the implementation of GIRUS-net, we extended each label y_n to K-1 labels, $y_n^{(0)} \cdots y_n^{(K-1)}$. The predicted probability for the severity level are given by

$$P(\hat{y}_n = k) = \sigma(\mathbf{w} \cdot \mathbf{x}_n + b_k). \quad (8)$$

We train O-Logit with the multi-label cross entropy loss in Eq. (5). We use the learned weights $\mathbf{w}$ as feature importances.

E. Evaluation Strategy

GIRUS-net is compared to the baseline methods across three experiments. In the first experiment, we classify Cognitive Normal (CN) and Alzheimer's Disease (AD) patients in 2-class classification task. We set up the second experiment as a 3-class classification task, where we predict Cognitive Normal (CN), Mild Cognitive Impairment (MCI) and Alzheimer's Disease (AD). Finally, in the third experiment we aim to predict Cognitive Normal (CN), Early Mild Cognitive Impairment (EMCI), Late Mild Cognitive Impairment (LMCI), and Alzheimer's Disease (AD) patients in a 4-class classification task. In the 3-class classification task, we combine the EMCI and LMCI to a common class called MCI. We evaluate all the models for classification performance and biomarkers explainability. The biomarkers are identified by assessing the feature importance maps. For GIRUS-net, imaging and genetic feature importance maps are calculated from $\mathbf{p}^i$, $\mathbf{p}^g$ and the classification predictions are the output of the ordinal regression branch.

a) Model Prediction Evaluation: We perform 10 repeats of stratified 10 fold cross validation. We evaluate the classification performance based on accuracy, f1-score (F1), Cohen's kappa (Kappa) [25], recall, and precision.

b) Feature Importance Evaluation: The feature importance maps are evaluated based on sparsity and consistency across classification task. We re-scaled the feature importance maps to $[0, 1]$. For all baseline methods, we scale imaging and genetic features importance maps collectively because imaging and genetic data are concatenated during training. For GIRUS-net, we scale the two feature maps separately as they come from separate branches of the model.

III. Results

A. Data Credit

Data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database [26]. The 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 serial magnetic resonance imaging (MRI), positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment and early Alzheimer's disease.

B. Data Preprocessing

The subjects in this study are included from the ADNI2/GO database. The subjects are classified as cognitive normal (CN), early mild cognitive impairment (EMCI), late mild cognitive impairment (LMCI), or mild Alzheimer's disease (AD) based on ADNI2 protocol. Table 1 summarizes the demographics of the 934 subjects, which contains both MRI and genetic data. The data are pre-processed and provided as a part of the TADPOLE challenge [27].

TABLE I

SUBJECT DEMOGRAPHIC ACROSS DISEASE CATEGORIES

Category[a]	Counts	Age(years)	Gender(M/F)[b]	MMSE
CN	269	75.6 ± 5.2	146/123	29.1 ± 1.1
EMCI	153	71.6 ± 7.1	91/62	28.3 ± 1.5
LMCI	346	74.8 ± 7.2	225/121	27.2 ± 1.5
AD	166	75.7 ± 7.7	93/73	23.4 ± 2.0

Table 1. Subject demographics of the dataset, including number of subjects per category, age, gender distribution, and Mini-Mental State Examination(MMSE). The data are expressed as mean ± SD.

[a] CN: Cognitive Normal; EMCI: Early Mild Cognitive Impairment; LMCI: Late Mild Cognitive Impairment; AD: Alzheimer's Disease.

[b] M: Male; F: Female.

a) MRI Imaging Data: The T1-weighted MRI imaging data are collected using a 3T scanner. The data are processed with gradient non-linearity, B1 non-uniformity correction and peak sharpening [28]. This study uses cross-sectional cortical thickness as imaging features, which are extracted via Freesurfer [29]. The imaging features consists of 68 brain regions of interest (ROIs), with 34 features from the right hemisphere and 34 from the left, based on the Desikan-Killiany atlas [30]. As an additional preprocessing step, we normalize all testing, validation, and training imaging data with the mean and standard deviation of the training data.

b) Genetics Data: In parallel, genotyping was done with GenomeStudio v2009.1 (Illumina). Quality control was performed using PLINK, resulting in 141912 Linkage Disequilibrium (LD) independent SNPs. We subselect 1165 SNPs by thresholding the p-value $p \leq 0.001$ based on an auxiliary genome-wide associated study data (GWAS) [31].

C. Classification Performance

Fig. 2 quantifies the classification performance of all the methods across the three experimental setups. Compared to the baselines, GIRUS-net is consistently showing better, or comparable performance across all the performance metrics. The confusion matrices in Fig. 3 further show that the baseline models fail to handle class imbalance. Especially in 3-class and 4-class scenarios, the baselines tend to predict all subjects as the majority class(es). In comparison, GIRUS-net can successfully distribute its predictions across class labels. The improved performance suggests that GIRUS-net can extract discriminative patterns of disease severity.

D. Feature Importance Sparsity and Consistency

The imaging and genetics biomarkers are identified by the mean feature importance maps across 10 repeats of 10-fold cross validations. As shown in Fig. 4, the baseline models rely mainly on one modality for diagnosis: imaging for RF, SVM, and ANN; genetics for Ordered Logit. In comparison, GIRUS-net puts equal importance on both the data modalities and extract a sparse set of biomarkers.

Additionally, the low importance scores demonstrate that these baseline methods fail to jointly extract discriminative information from both the data modalities. The superior performance of GIRUS-net suggests that the learnable dropout layers can selectively find brain and genetics biomarkers that are crucial for the downstream severity prediction.

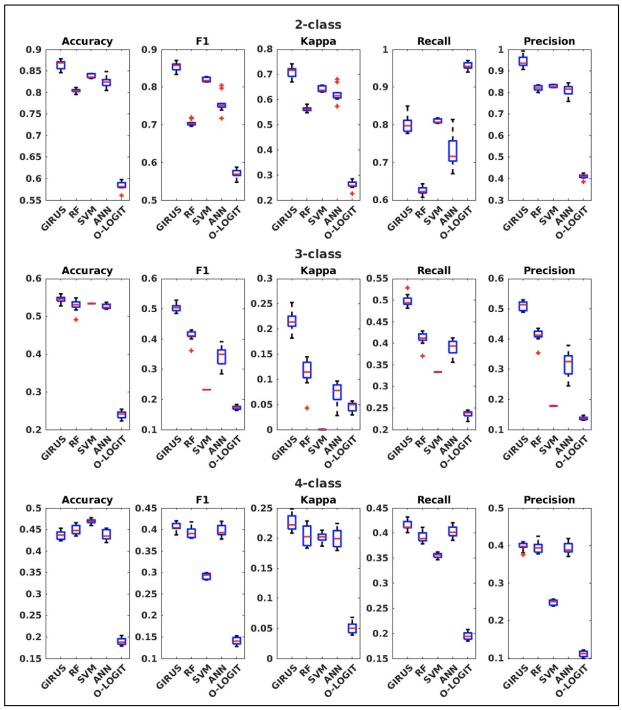

Fig. 2. Prediction performance of each method across 10 repeats of 10 fold cross validations. The box plots capture the mean and the deviation of the metrics across all the folds.

E. Analysis of Imaging Biomarkers

We average the imaging feature importance maps learned from GIRUS-net across the three classification experiments. In Fig. 5, we plot top 10 regions onto the brain with colors corresponding to the value of feature importance. Our model identifies brain regions including lateral ventricle, medial temporal lobe, inferior temporal lobe, and parahippocampal gyrus. Correspondingly, AD is characterized by enlarged ventricles [32] and loss of tissue in inferior parietal gyrus [33] and parahippocampal gyrus [34]. Functionally, the hippocampus is crucial for episodic and spatial memory [35] which is affected by AD. Overall, GIRUS-net identifies brain regions with high association to AD in the literature.

F. Analysis of Genetics Biomarkers

Fig. 6 shows the mean of the genetic feature selection maps identified by the learnable dropout layer across the three classification experiments. A higher value indicates genetic variants containing discriminative information about all the classification tasks and are potential AD risk loci. We annotate the top 10 SNPs and their overlapping or affected genes as listed in The Ensemble Variant Effect Predictor and the GWAS Catalog [36], [37]. Our model identifies well established Alzheimer's risk factors such as TOMM40 [38] and APOE [39] with high feature importance. Using the GTEx database, we identify the set of brain tissues where the set of genes show high expression levels. Aside

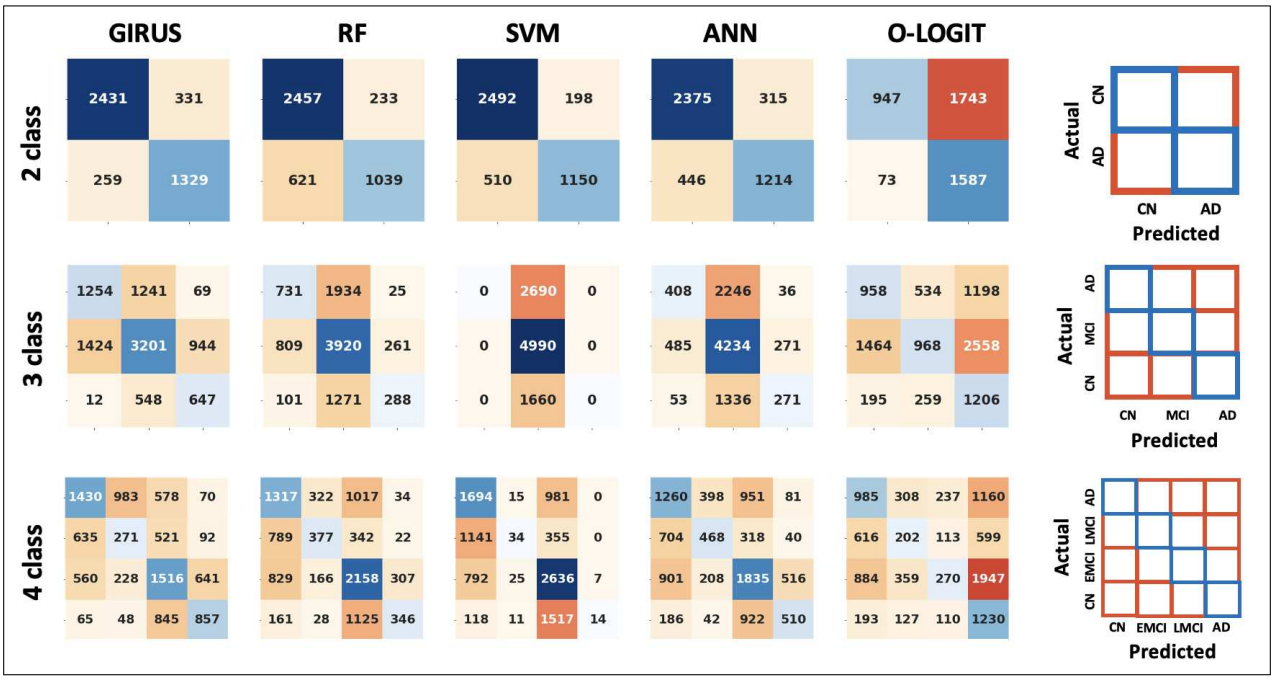

Fig. 3. The confusion matrices are constructed with all the testing data across 10 repeats of 10-fold cross validation. The diagonal boxes, colored in blue, are representative of correct predictions, and the off-diagonal orange boxes capture the inconsistency between the actual and predicted class labels.

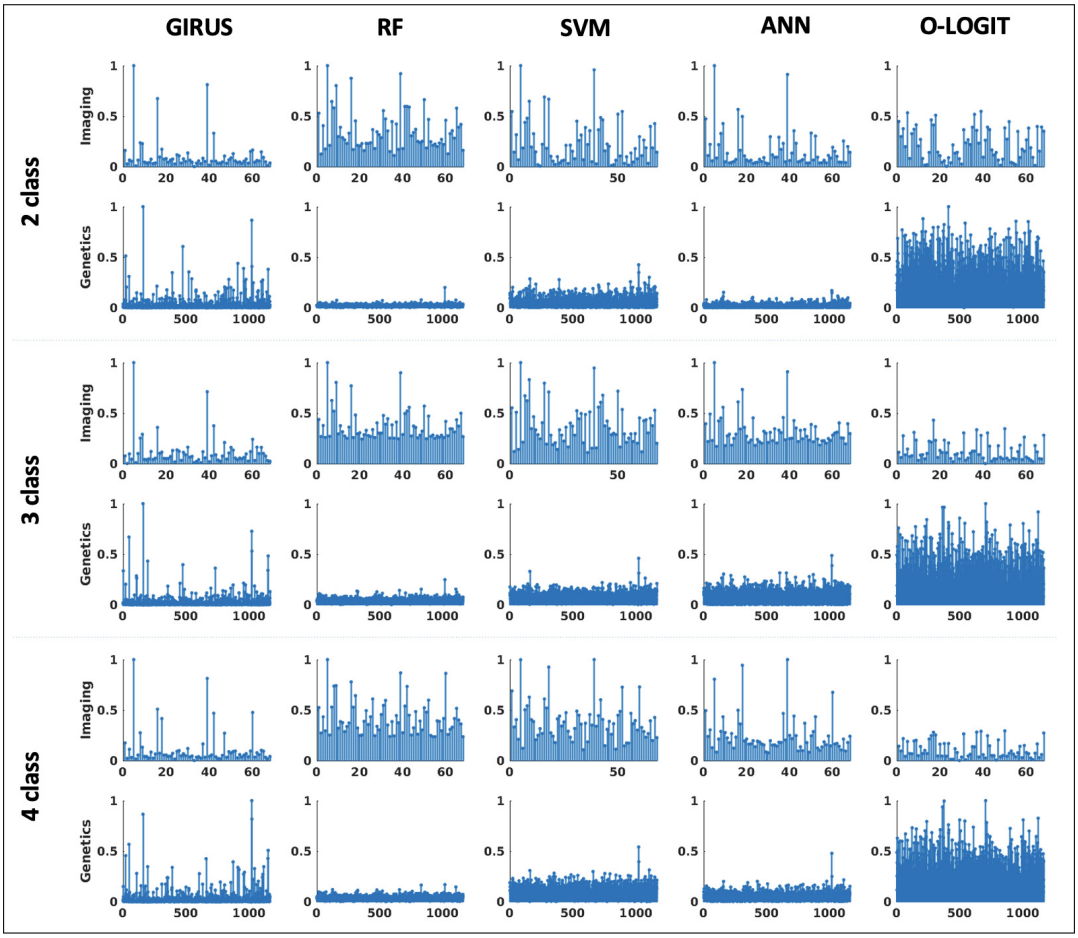

Fig. 4. The mean feature importance maps are calculated across all the cross validation folds in each experimental setting. A high feature importance value captures the location of consistent and discriminative biomarkers.

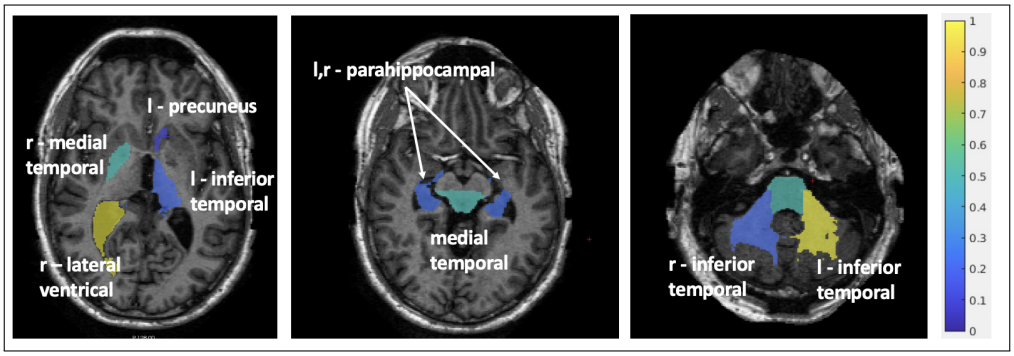

Fig. 5. The brain regions correspond to top 10 important features. The feature importances are calculated by taking the mean across all the experimental settings. The color bar corresponds to the feature importance values.

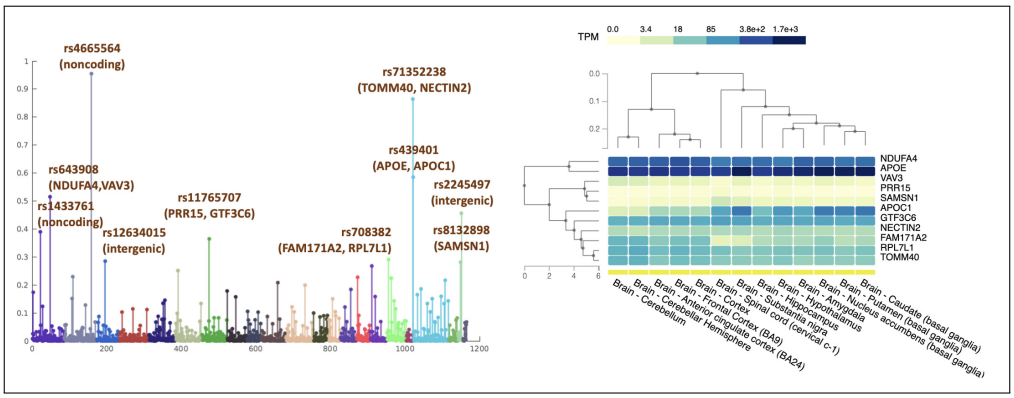

Fig. 6. Left: Mean genetic importance of all SNPs across the three classification tasks. The variants are color coded according to their location on the chromosomes. Right: The gene expression pattern of the selected set of genes in different brain tissues based on the GTEx database.

from APOE and TOMM40, we identify NDUFA4, which plays a regulatory role in the expression of synaptophysin in the hippocampus, and gene mutation could potentially lead to AD [40]. Aside from already established genes, several intergenic and non-coding SNPs are also selected with high feature importance by GIRUS-net . Their association to AD is unknown. The explainability study demonstrated that the genetic biomarkers identified by GIRUS-net aligns with research findings and may assist in future genetic analysis.

IV. Conclusions

In this paper, we introduce GIRUS-net, a deep learning framework for multi-modal fusion, biomarker extraction, and severity prediction for AD. As compared to prior work, we introduce a rank-consistent ordinal regression module that extracts discriminative features that have a progressive effect on disease severity. In a population study of AD, GIRUS-net successfully integrates imaging-genetic data for disease severity prediction. Compared to standard baselines, GIRUS-net extracts consistent, distinctive, and clinically relevant information from imaging and genetic features across various complex diagnosis tasks. In addition, GIRUS-net is not tied to any specific data modality; the flexible design allows the user to combine diverse data modalities and provide a comprehensive view of various diseases.

V. Acknowledgement

This work was supported by National Science Foundation CAREER award 1845430 (PI: Venkataraman), National Institutes of Health award R01-HD108790 (PI: Venkataraman), and a Johns Hopkins Catalyst Award (PI: Venkataraman).

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] N. Berchtold *et al.*, "Evolution in the conceptualization of dementia and alzheimer's disease: Greco-roman period to the 1960s," *Neurobiology of Aging*, vol. 19, no. 3, pp. 173–189, 1998.

[2] van der Flier *et al.*, "Medial temporal lobe atrophy and white matter hyperintensities are associated with mild cognitive deficits in non-disabled elderly people: the ladis study," *Journal of Neurology, Neurosurgery & Psychiatry*, vol. 76, no. 11, pp. 1497–1500, 2005.

[3] M. Lorenzi *et al.*, "Susceptibility of brain atrophy to TRIB3 in Alzheimer's disease, evidence from functional prioritization in imaging genetics," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 115, no. 12, pp. 3162–3167, 2018.

[4] Farrer *et al.*, "Rate of Progression of Alzheimer's Disease Is Associated With Genetic Risk," *Archives of Neurology*, vol. 52, no. 9, pp. 918–923, 09 1995.

[5] G. M. Murphy *et al.*, "Rate of cognitive decline in ad is accelerated by the interleukin-1α -889 *1 allele," *Neurology*, vol. 56, no. 11, pp. 1595–1597, 2001.

[6] Y. Xin *et al.*, "A review of imaging genetics in alzheimer's disease," *Journal of Clinical Neuroscience*, vol. 100, pp. 155–163, 2022.

[7] L. Shen *et al.*, "Brain imaging genomics: Integrated analysis and machine learning," *Proceedings of the IEEE*, vol. 108, no. 1, pp. 125–162, 2020.

[8] A. B. Rabeh *et al.*, "Diagnosis of alzheimer diseases in early step using svm (support vector machine)," in *2016 13th International Conference on Computer Graphics, Imaging and Visualization (CGiV)*, 2016, pp. 364–367.

[9] I. Illán *et al.*, "18f-fdg pet imaging analysis for computer aided alzheimer's diagnosis," *Information Sciences*, vol. 181, no. 4, pp. 903–916, 2011.

[10] P. Johnson *et al.*, "P2–266: Genetic algorithm with logistic regression for alzheimer's disease diagnosis and prognosis," *Alzheimer's & Dementia*, vol. 9, no. 4S_Part_11, pp. P455–P456, 2013.

[11] S. I. Dimitriadis *et al.*, "How random is the random forest? Random forest algorithm on the service of structural imaging biomarkers for Alzheimer's disease: from Alzheimer's disease neuroimaging initiative (ADNI) database," *Neural Regen Res*, vol. 13, no. 6, pp. 962–970, Jun 2018.

[12] M. Rangini *et al.*, "Detection of alzheimer's disease through automated hippocampal segmentation," in *2013 International Mutli-Conference on Automation, Computing, Communication, Control and Compressed Sensing (iMac4s)*, 2013, pp. 144–149.

[13] D. AL-Dlaeen *et al.*, "Using decision tree classification to assist in the prediction of alzheimer's disease," in *2014 6th International Conference on Computer Science and Information Technology (CSIT)*, 2014, pp. 122–126.

[14] S. A. Meda *et al.*, "A large scale multivariate parallel ica method reveals novel imaging–genetic relationships for alzheimer's disease in the adni cohort," *NeuroImage*, vol. 60, no. 3, pp. 1608–1621, 2012.

[15] M. Lorenzi *et al.*, "Partial least squares modelling for imaging-genetics in alzheimer's disease: Plausibility and generalization," in *2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI)*, 2016, pp. 838–841.

[16] C. Li *et al.*, "A neuroimaging feature extraction model for imaging genetics with application to alzheimer's disease," in *2017 IEEE 17th International Conference on Bioinformatics and Bioengineering (BIBE)*, 2017, pp. 15–20.

[17] S. Basaia *et al.*, "Automated classification of alzheimer's disease and mild cognitive impairment using a single mri and deep neural networks," *NeuroImage: Clinical*, vol. 21, p. 101645, 2019.

[18] S. Ghosal *et al.*, "G-MIND: an end-to-end multimodal imaging-genetics framework for biomarker identification and disease classification," in *Medical Imaging 2021: Image Processing*, vol. 11596. SPIE, 2021, p. 115960C.

[19] W. Cao *et al.*, "Rank consistent ordinal regression for neural networks with application to age estimation," *Pattern Recognition Letters*, vol. 140, pp. 325–331, 2020.

[20] Y. Gal *et al.*, "Concrete dropout," 2017. [Online]. Available: https://arxiv.org/abs/1705.07832

[21] G. Louppe, "Understanding random forests: From theory to practice," 2014.

[22] T.-f. Wu, C.-j. Lin, and R. Weng, "Probability estimates for multi-class classification by pairwise coupling," in *Advances in Neural Information Processing Systems*, vol. 16. MIT Press, 2003.

[23] Lundberg *et al.*, "A unified approach to interpreting model predictions," in *Advances in Neural Information Processing Systems*, vol. 30. Curran Associates, Inc., 2017.

[24] P. McCullagh, "Regression models for ordinal data," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 42, no. 2, pp. 109–142, 1980. [Online]. Available: http://www.jstor.org/stable/2984952

[25] J. Cohen, "A coefficient of agreement for nominal scales," *Educational and Psychological Measurement*, vol. 20, no. 1, pp. 37–46, 1960.

[26] S. G. Mueller *et al.*, "The alzheimer's disease neuroimaging initiative," *Neuroimaging Clinics of North America*, vol. 15, no. 4, pp. 869–877, 2005, alzheimer's Disease: 100 Years of Progress.

[27] R. V. Marinescu *et al.*, "TADPOLE challenge: Accurate alzheimer's disease prediction through crowdsourced forecasting of future data," in *Predictive Intelligence in Medicine*. Springer International Publishing, 2019, pp. 1–10.

[28] J. Jr. *et al.*, "The alzheimer's disease neuroimaging initiative (adni): Mri methods," *Journal of Magnetic Resonance Imaging*, vol. 27, no. 4, pp. 685–691, 2008.

[29] B. Fischl *et al.*, "Automatically Parcellating the Human Cerebral Cortex," *Cerebral Cortex*, vol. 14, no. 1, pp. 11–22, 01 2004.

[30] R. S. Desikan *et al.*, "An automated labeling system for subdividing the human cerebral cortex on mri scans into gyral based regions of interest," *NeuroImage*, vol. 31, no. 3, pp. 968–980, 2006.

[31] L. Shen *et al.*, "Whole genome association study of brain-wide imaging phenotypes for identifying quantitative trait loci in mci and ad: A study of the adni cohort," *NeuroImage*, vol. 53, no. 3, pp. 1051–1063, 2010, imaging Genetics.

[32] S. M. Nestor *et al.*, "Ventricular enlargement as a possible measure of alzheimer's disease progression validated using the alzheimer's disease neuroimaging initiative database." *Brain : a journal of neurology*, vol. 131 Pt 9, pp. 2443–54, 2008.

[33] S. W. Scheff *et al.*, "O1–01–01: Synaptic alterations in inferior temporal lobe in alzheimer's disease (ad) and mild cognitive impairment (mci)," *Alzheimer's & Dementia*, vol. 2, no. 3S_Part_1, pp. S7–S8, 2006.

[34] G. W. VAN HOESEN *et al.*, "The parahippocampal gyrus in alzheimer's disease: Clinical and preclinical neuroanatomical correlates," *Annals of the New York Academy of Sciences*, vol. 911, no. 1, pp. 254–274, 2000.

[35] V. Cutsuridis *et al.*, "Editorial: Memory processes in medial temporal lobe: Experimental, theoretical and computational approaches," *Frontiers in Systems Neuroscience*, vol. 11, 04 2017.

[36] W. McLaren *et al.*, "Deriving the consequences of genomic variants with the Ensembl API and SNP Effect Predictor," *Bioinformatics*, vol. 26, no. 16, pp. 2069–2070, 06 2010.

[37] A. Buniello *et al.*, "The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019," *Nucleic Acids Research*, vol. 47, no. D1, pp. D1005–D1012, 11 2018.

[38] A. Roses *et al.*, "A tomm40 variable-length polymorphism predicts the age of late-onset alzheimer's disease," *The pharmacogenomics journal*, vol. 10, pp. 375–84, 12 2009.

[39] M. S. Uddin *et al.*, "Apoe and alzheimer's disease: Evidence mounts that targeting apoe4 may combat alzheimer's pathogenesis," *Molecular neurobiology*, vol. 56, no. 4, p. 2450—2465, April 2019.

[40] S. K. Shil, Y. Kagawa, B. Umaru, F. Nanto-Hara, H. Miyazaki, Y. Yamamoto, S. Kobayashi, C. Suzuki, T. Abe, and Y. Owada, "Ndufs4 ablation decreases synaptophysin expression in hippocampus," *Scientific Reports*, vol. 11, 2021.