Dual-Pathway Neural Networks for Alzheimer's Diagnosis: A Comparative Study of DenseCNN and ResNet on Hippocampal MRI Segmentation

1st Noah Leuenberger
School of Engineering, Data Science
University of Applied Sciences Northwestern Switzerland
Windisch, Switzerland
noah.leuenberger@students.fhnw.ch

2nd Dominik Filliger
School of Engineering, Data Science
University of Applied Sciences Northwestern Switzerland
Windisch, Switzerland
dominik.filliger@students.fhnw.ch

3rd Nils Fahrni
School of Engineering, Data Science
University of Applied Sciences Northwestern Switzerland
Windisch, Switzerland
nils.fahrni@students.fhnw.ch

4th Etienne Roulet
School of Engineering, Data Science
University of Applied Sciences Northwestern Switzerland
Windisch, Switzerland
etienne.roulet@students.fhnw.ch

***Abstract*—This study examines the effectiveness of Dense Convolutional Neural Networks (DenseCNN) and Residual Networks (ResNet) in diagnosing Alzheimer's Disease (AD) through hippocampal MRI segmentation. This study uses the Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset to compare the performance of these two neural network models in AD detection. DenseCNN and ResNet were analyzed with and without data augmentation to evaluate their predictive capabilities. The study shows that DenseCNN performs better with augmented data, indicating its sensitivity to diverse training examples. ResNet, enhanced by pre-training on MedicalNet, also benefits from augmentation but to a lesser extent. The study also found that the choice of features affects model performance, with DenseCNN performing better with the left hippocampus under augmentation and ResNet performing better with dual-channel input. These findings indicate that a comprehensive representation of the hippocampal regions improves AD prediction accuracy. This study highlights the significance of model architecture, data augmentation, feature selection, and pre-training in the effectiveness of neural networks for medical imaging analysis in AD. Future research should continue to investigate customized approaches based on specific dataset characteristics and disease manifestations. This will expand the potential for early and accurate diagnosis of AD using advanced neural network models.**

***Index Terms*—AD Diagnosis, Convolutional Neural Networks (CNNs), DenseCNN, ResNet, MRI Segmentation, Data Augmentation, Medical Imaging, ADNI, Predictive Modeling**

I. Introduction

This paper delves into Convolutional Neural Networks (CNN) architectures, focusing on their ability to predict cognitive normal (CN) and Alzheimer's disease (AD) states based on 3D Magnetic Resonance Imaging (MRI) images, with a special emphasis on the hippocampus (left and right). The study aims to contribute to the technical domain of the AD research community by investigating different machine-learning methods for processing imaging data.

A. Current Research

The use of machine learning models in the medical field has become omnipresent in the past decade. Numerous research teams around the globe already tried to implement a new way of classifying AD together with MRI images. New methods in the general field of machine learning are invented very frequently. Therefore, it is important to consider these as they could yield better results than observed before. This paper focuses on research that was made using a Dense Convolutional Neural Network (DenseCNN) architecture, proposed by [1] and [2], and a pre-trained MedicalNet ResNet proposed by [3].

B. Hippocampus as an Alzheimer's Marker

The hippocampus, a brain region critical for memory and spatial navigation, is considered a viable marker for Alzheimer's: The hippocampus is one of the first areas affected by Alzheimer-related changes. It undergoes atrophy (tissue shrinking) and shows functional impairment early in the disease process [4]–[6].

Changes in the size, shape, and function of the hippocampus can be detected using imaging techniques like MRI [7]. These changes correlate with Alzheimer's symptoms and progression, enhancing diagnostic accuracy as shown in [8].

This makes the hippocampus region a relevant feature in predictive modeling AD as shown in [1] and [2].

C. Motivation and Research Objectives

This research is driven by previous studies results [1], [2] focusing on AD diagnostics through advanced machine learning techniques, specifically targeting the analysis of the hippocampus via 3D MRI. Recognizing the critical role of the hippocampus in early AD manifestations, the study aims to use

current methodologies, drawing upon the use of DenseCNN and MedicalNet ResNet models. The objectives are distinct yet interconnected:

a) How do different models perform in predicting Alzheimer's disease using only the Hippocampus Segments?: To rigorously compare the efficacy of different modeling approaches in accurately predicting AD from hippocampus segments.

b) How do augmentations impact the performance?: To investigate the influence of data augmentations on these models' predictive capabilities.

c) Can dual-pathway neural networks enhance the performance of prediction?: To assess the feasibility and added value of incorporating dual-pathway neural networks in this context.

II. Data Acquisition and Preprocessing

The Alzheimer's Disease Neuroimaging Initiative (ADNI) is a collaborative project that unites researchers from diverse fields with the common goal of understanding and ultimately treating AD [9]. ADNI was launched in 2004 with the primary objective of testing whether serial magnetic resonance imaging (MRI), positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessments can be combined to measure the progression of mild cognitive impairment (MCI) and early AD.

A. Dataset

This research focuses exclusively on neuroimaging data from the ADNI dataset. The approach enables an in-depth analysis of structural changes in the hippocampus, a brain region crucially affected by Alzheimer's Disease (AD).

The study utilized the 'ADNI1: Complete 3Yr 1.5T' series from the ADNI database, which includes 2182 MRI images captured using 1.5 Tesla scanning equipment. The dataset comprises both initial and subsequent screening scans. The dataset underwent several preprocessing stages, including gradwarp for geometric distortion correction, B1 field inhomogeneity correction to standardize magnetic field effects, N3 intensity normalization for uniform intensity distribution, and image scaling to enhance resolution and clarity [10].

The focus of this study lies on the CN and AD cases from the ADNI dataset. This selection aims to compare normal aging with Alzheimer's-related changes in the hippocampus. By analyzing these groups, the study aims to classify AD and CN cases, ignoring MCI cases in the process for simplicity. Filtering out the MCI cases reduces the sample size to 1201, as shown in Figure 1. This figure also illustrates the distribution between these two classes, which will be further explained in Methods.

The demographic breakdown of the participants in the study is critical for understanding the diversity and representation within the research. As illustrated in Figure 2, the sex distribution among the 1201 participants is nearly balanced, with 602 males (50.12%) and 599 females (49.88%). Additionally, the

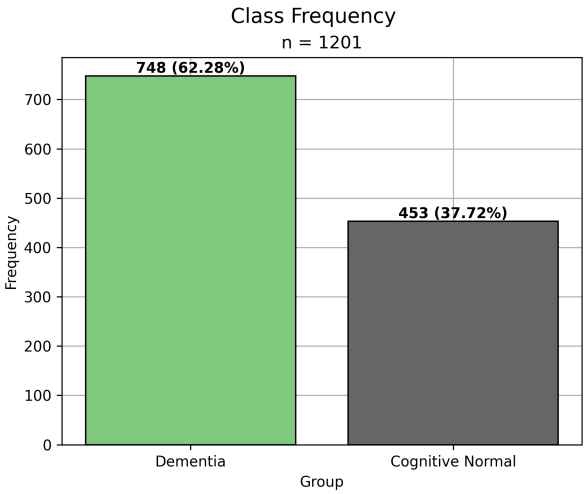

Fig. 1. Class Frequencies between CN and AD cases

age distribution is visualized, indicating a majority of participants are between 70 to 80 years old, reflecting the increased risk of AD with advancing age. Notably, the bottom-left plot in Figure 2 reveals the sex distribution across the two groups studied AD and CN, with females representing 32.4% in the CN group and 29.9% in the AD group. The bottom-right plot provides a boxplot comparison of age distribution by group, showing a similar median age for both CN and AD participants but with AD participants displaying a slightly wider age range, suggesting variability in the onset and detection of the disease.

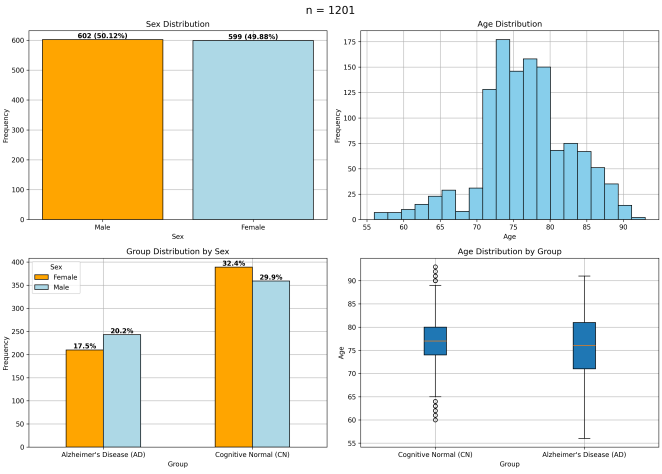

Fig. 2. Demographics of the dataset

B. Segmentation

The next step in this study's methodology is to segment the MRIs, with a specific focus on isolating the left and right Hippocampus. This segmentation is facilitated by the Hippodeep model, which is known for its rapid segmentation capabilities and its `PyTorch` implementation that maintains the three-dimensionality of the images [11], [12]. Refer to

Figure 3 for an example of a segmented left hippocampus mask. The Hippodeep model is unique in that it not only separates the hippocampal regions from the rest of the brain but also calculates the volume of each segment [12]. This volumetric data provides an additional valuable feature for the proposed models, aligning with approaches such as those of [1] who utilized global shape representation features in their research.

The dataset of the study primarily utilizes T1.5-weighted images, which differs from the T1-weighted images for which the Hippodeep-Pytorch model was specifically developed, as detailed in the GitHub repository [12]. This discrepancy is important because T1.5-weighted images present a higher resolution with better contrast properties compared to T1-weighted ones, potentially impacting the accuracy of hippocampus segmentation. Despite this, the GitHub repository's readme as well as the original study offer reassurance, stating that the model, equipped with a CNN, has been pre-trained on a vast array of images from multiple extensive cohorts. This training background suggests a high degree of robustness to variations in both subject characteristics and MR contrasts [12], [13]. Therefore, while the deviation from the model's original design parameters is noteworthy, the model's inherent versatility supports the validity of the suggested approach.

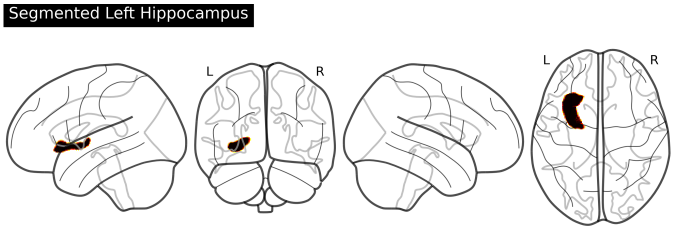

Fig. 3. Segmented left hippocampus mask

The volumetric analysis of the hippocampus, as depicted in Figure 4, reveals subtle yet important differences between AD patients and CN individuals, especially when examining sex differences and the specific role of the left hippocampus. While the boxplot indicates only slightly visible differences in left hippocampal volume between AD and CN cases, these are not statistically significant in the dataset at hand. However, the importance of the left hippocampus in the early stages of AD is well-documented, with studies showing its correlation with cognitive task performance in AD patients [14], [15]. Additionally, research highlights significant sex differences in hippocampal volume changes in AD, with a more pronounced impact on disease progression among women [16]. These distinctions underscore the need to consider both sex and specific hippocampal regions, particularly the left hippocampus, in AD research for a deeper understanding of the disease's pathophysiology.

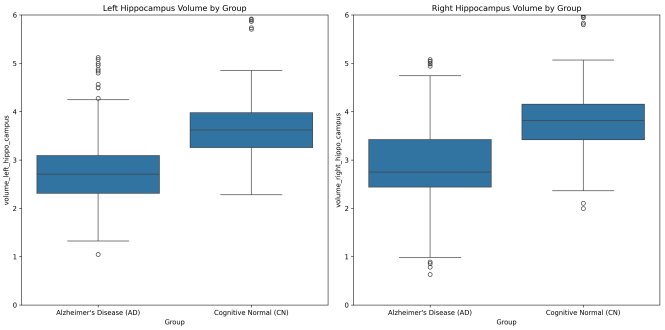

Fig. 4. Hippocampus volumes by group and sex

C. Cropping and Resizing

Post-segmentation, the identified hippocampus masks undergo a cropping process using the logic from the DeepHipp model [17]. This step ensures that only the hippocampus is retained in each image, eliminating any residual brain areas. The objective here is to create a set of images, or tensors, that exclusively represent the hippocampal regions, thereby facilitating more targeted data analysis. The cropped version of a hippocampus segmented can be seen in figure 5.

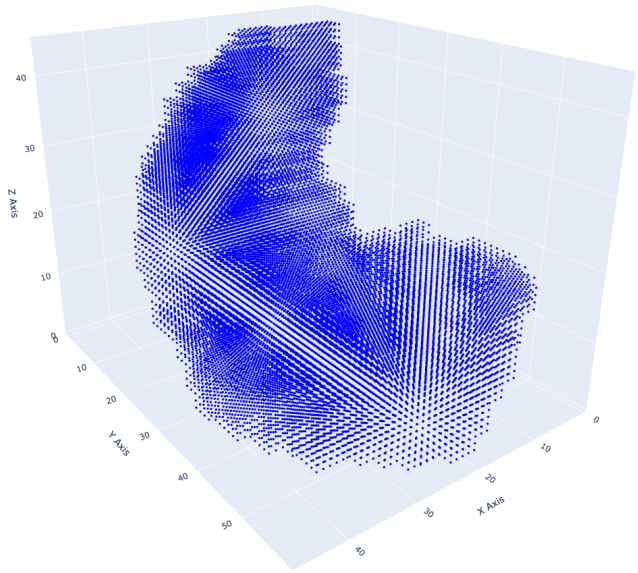

Fig. 5. Cropped and resized right hippocampus mask

An integral aspect of optimizing the image processing pipeline for computational efficiency involves resizing the hippocampus masks to reduce the demands on GPU memory. The chosen mask size of 48x60x46 is a deliberate decision, inspired by standards observed in other research [2]. They highlight that the input size to their model is comparable to standard dimensions used in 2D computer vision models, like 256x256 or 512x512, which can accommodate over 100 layers. Training deep CNNs with numerous parameters on smaller datasets can result in overfitting. By resizing the hippocampus

masks to a smaller size, the risk of bias towards false markers can be mitigated, ensuring computational feasibility and reducing overfitting.

III. Methods

This section delves into the methodologies employed in this research, focusing on the augmentation of the ADNI dataset, and the application of DenseCNN and ResNet architectures. Each subsection is dedicated to a distinct aspect of the methodology, outlining the rationale behind the choices made and the steps involved in the implementation.

A. Augmentation

The subset of ADNI data that was used to train the models subject to this paper is heavily imbalanced as explored in the section Dataset. Recognizing this issue of imbalance inherent in medical datasets, an augmentation process was implemented to counteract this disparity. Inspired by the augmentation approach of [2], which is based on a previous study [18], a similar process was followed by utilizing the `TorchIO` library. By shifting each hippocampus segment (left and right) one unit in positive and one unit in negative direction in 3-dimensional space six new samples per hippocampus were created, yielding 7206 new pairs of hippocampi.

Shifting the 3-dimensional masks by just one unit, in the context of the 48x60x46 mask size, adds enough error without manipulating or losing important information. This simple augmentation process was defined as it does not influence any of the key properties captured in the subsection Hippocampus as an Alzheimer's Marker.

The resulting dataset, including augmentations, was used for training with undersampling of the majority class to create a $50/50$ balance within the train set, meanwhile keeping the original distribution of classes in the test set.

B. DenseCNN

DenseCNNs have significantly advanced deep learning in image analysis, particularly through their unique architecture. Each layer in a DenseCNN connects in a feed-forward manner, enhancing information and gradient flow [19].

1) Applicability of DenseCNN: DenseCNNs have shown significant promise in medical imaging tasks, especially in the context of AD prediction. Their ability to handle complex image data, such as hippocampi masks and volumetric information, makes them well-suited for tasks requiring detailed structural analysis. Studies have demonstrated that DenseCNNs, through their advanced architectures, efficiently integrate both visual and global shape features of the hippocampus, leading to enhanced diagnostic capabilities in AD prediction. For instance [1] developed a lightweight 3D deep convolutional network model, DenseCNN2, tailored for AD classification using hippocampi masks and volumetric information.

2) Architecture:

a) Dense Blocks and Dual-Stream Architecture: The model of this study follows the foundational concepts introduced in the studies [2] and [1] in its use of dense blocks within a dual-stream setup. However, diverging from [2], the dense blocks implemented commence with a layer that concatenates the block's input with the output of its first layer. This approach more closely aligns with the original design introduced by [19], ensuring the integration of initial input across the block.

b) Convolution and Transition Layers: In determining convolution sizes for transition layers and the initial convolution, this study diverges from the configurations provided in [1] and [2], which did not provide detailed configurations. The transition layers used, utilize $1 \times 1 \times 1$ convolutions with a stride of 1, aiming to balance dimensionality reduction with feature retention. The initial convolution employs a $3 \times 3 \times 3$ size with padding 3, effectively capturing spatial relationships and intricate features.

c) Integration of Hippocampus Volume: In the proposed architecture of this study, the concatenation of a scalar representing hippocampus volume for both hemispheres before the fully connected layer was performed, similar to the approach in [1] of utilizing the LB spectrum. This method integrates a critical anatomical metric for AD prediction, emphasizing hippocampal atrophy common in AD patients.

d) Layer Configuration and Growth Rate: Adopting the approach in [2], the model constructed in this paper includes three dense layers per block. This configuration, as indicated by [2], effectively captures complex image data patterns crucial in medical imaging for AD prediction. Regarding the growth, a rate of 64 was selected; A divergence from both [2] and [1], which did not provide detailed configurations regarding growth rate. This growth rate strikes a balance between capturing an expansive feature set and maintaining computational efficiency.

In aligning the architecture with existing research, the goal was to harness the combined strengths of these foundational works while introducing novel elements to enhance the model's diagnostic capabilities in AD prediction.

C. ResNet

Introduced by [20], Residual Networks (ResNet) have revolutionized deep learning architectures with their unique use of residual blocks and skip connections. These innovations effectively address the vanishing gradient problem in deep neural networks, enabling efficient learning of complex patterns, an essential feature for medical imaging analysis [20].

ResNet's architecture, known for its depth and efficiency, is highly effective in AD imaging analysis for identifying subtle biomarkers. ResNet has different complexities, known as configurations. These configurations vary mainly in depth, with common versions including ResNet-18, ResNet-34, ResNet-50, and ResNet-101. These models differ in the number of layers, impacting their complexity and computational demands. As computational power and efficiency were a concern, a ResNet-18 was chosen for the AD imaging analysis due to its

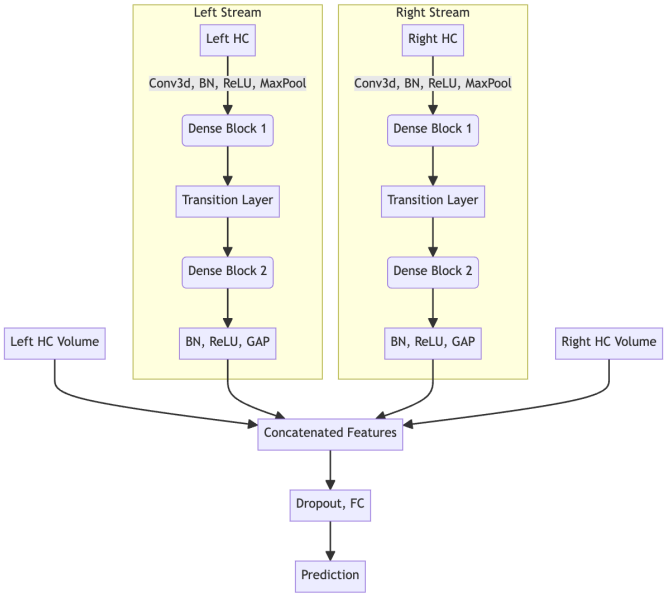

Fig. 6. Proposed adopted DenseCNN architecture

adequate depth for capturing essential features of AD biomarkers, while being computationally efficient for the dataset size [21].

1) MedicalNet: Pretrained Weights for Enhanced Learning: To further improve performance, a ResNet-18 with pre-trained weights from MedicalNet [3] was chosen. Utilizing the concept of transfer learning, these weights are applied to leverage pre-existing knowledge from diverse medical datasets [22]–[25], including brain MRI and hippocampus imaging, MedicalNet provides a comprehensive base of medical imaging knowledge. The fact that MedicalNet was not trained on the ADNI dataset ensures no data leakage emerges in the model training, allowing unbiased learning from the ADNI dataset [3].

Integrating these pre-trained weights allows the model to access pre-learned patterns in medical imaging, which should improve the overall robustness and accuracy for the classification task at hand [3].

2) Adapting ResNet for Dual Channel Hippocampal Analysis: The pre-trained ResNet model, originally designed for single-channel input, has been adapted to process dual-channel input, specifically targeting the hippocampus. This adaptation involves duplicating the input channel, allowing the network to separately analyze the left and right hippocampus regions. The dual-channel approach is essential for a comprehensive comparison with the DenseCNN approach, as it is introduced in the reference papers [1], [2].

IV. Platform and Parameters for Experiments

The segmentation and model training was executed on a robust platform comprising an NVIDIA GeForce GTX 1080 Ti GPU with 11 GB memory, an NVIDIA Quadro P6000 GPU with 24 GB memory, and an NVIDIA RTX 4000 GPU with 8GB memory.

For model development, `PyTorch Lightning` was utilized, a framework that accelerates the process of writing scalable, production-ready code, as cited in [26]. During the training phase, it was ensured that the models reached sufficient convergence for optimal performance. Defining sufficient convergence involved manual observation of the training set's performance development, with training ceasing at an epoch where a noticeable point of no significant change in the model's performance was observed. The choice to use manual observation for determining sufficient convergence was guided by the clear, consistent convergence patterns evident in all models, with epochs reliably ranging between 60 to 80. This approach was both practical and effective, negating the need for more complex methods in this context.

The learning rate for the models was managed by a Cosine Annealing Learning Rate scheduler which was optimised by an Adamax optimizer. These choices are based on common practices in the training of ResNet and DenseCNN [27], [28].

Consistent parameters were maintained across all models and final experiments: a batch size of 32, a learning rate of 10^{-3}, and a weight decay of 10^{-5}. The batch size of 32 offers an efficient compromise between memory usage and the ability to generalize the model. A learning rate of 10^{-3} is a standard choice that balances fast convergence with the risk of overshooting minima. The weight decay of 10^{-5} adds a regularization effect which is not too severe, but still crucial for preventing overfitting. Additionally, the Binary Cross Entropy loss function was employed as it is particularly effective for binary classification tasks.

For data partitioning, a $60/40$ train-test split was applied based on exploratory findings. This ratio aims to provide a valid balance while the training data is augmented, the test set remains sufficiently large to provide a reliable evaluation of the model's performance.

V. Results and Analysis

Tables I and II respectively present the performance metrics on the test set of DenseCNN and ResNet models in predicting AD. Each table is divided into two sections, one for results 'With Augmentation' and the other for 'Without Augmentation'. For both models, performance is evaluated using three metrics: F1 Score, Accuracy, and AUROC. These metrics are reported for three different feature sets: 'L' (left-sided hippocampus), 'R' (right-sided hippocampus), and 'Both' (combination of left and right-side hippocampus). Higher values in these metrics indicate better predictive performance. In the case of the DenseCNN, the volumes of the corresponding input features are always included.

A 60/40 training/test split was chosen to balance adequate training data with a substantial test set, needed for validating the models. The larger test set ensures a proportionally large validation set, essential when the training set is expanded through data augmentation.

TABLE I
DENSECNN RESULTS

With Augmentation			
Features	**F1 Score**	**Accuracy**	**AUROC**
L	0.942	0.961	0.956
R	0.920	0.960	0.937
Both	0.933	0.952	0.951
Without Augmentation			
Features	**F1 Score**	**Accuracy**	**AUROC**
L	0.800	0.827	0.850
R	0.821	0.855	0.869
Both	0.810	0.848	0.855

TABLE II
RESNET RESULTS

With Augmentation			
Features	**F1 Score**	**Accuracy**	**AUROC**
L	0.901	0.931	0.963
R	0.920	0.946	0.974
Both	0.931	0.954	0.972
Without Augmentation			
Features	**F1 Score**	**Accuracy**	**AUROC**
L	0.871	0.913	0.918
R	0.882	0.925	0.943
Both	0.873	0.913	0.931

A. DenseCNN

DenseCNN demonstrates strong performance in AD prediction, particularly with augmented data. The performance of the left hippocampus under augmentation, with an F1 Score of 0.942, Accuracy of 0.961, and AUROC of 0.956, demonstrates competitive results in line with those reported in [2] and [1], underscoring the potential of the proposed DenseCNN in AD detection. However, it is important to interpret these results with caution, as the specific dataset and segmentation methods used can significantly influence performance outcomes. In addition to that, the fact that only the left hippocampus together with its volume as input features was used to achieve the mentioned results diverges from the performance mentioned in both inspiration papers [1], [2]. As these papers [1], [2] also classify MCI, there is a noticeable difference in the dataset used for training, which explains the higher scores of this study only focusing on CN and AD.

1) Impact of Features: DenseCNN's feature-specific performance analysis under augmentation shows a clear preference for the left hippocampus (F1 Score: 0.942, Accuracy: 0.961, AUROC: 0.956), aligning with findings that suggest left hippocampal atrophy is a significant biomarker for AD [14], [15]. Interestingly, without augmentation, DenseCNN exhibits a performance shift favoring the right hippocampus. This change warrants a deeper examination of how the model interacts with different feature sets and the potential influence of dataset nuances, such as variations in hippocampal morphology or image acquisition parameters.

2) Impact of Augmentation: The introduction of data augmentation leads to a significant improvement in DenseCNN's performance across all metrics. This enhancement is most pronounced in the left hippocampus feature set, where the model's ability to detect subtle patterns associated with AD is markedly improved. The augmented data contributes to the model's learning process and improves its ability to generalize by exposing it to more diverse training examples.

B. ResNet

1) Impact of Features: The analysis of feature-specific performance in the ResNet model for AD prediction reveals some intriguing patterns. Notably, the model demonstrates superior performance when utilizing dual-channel inputs (combining left and right hippocampus) compared to either side alone.

- **Superiority of Dual-Channel Input:** The results indicate a general trend where the 'Both' feature set (left and right hippocampus combined) outperforms the individual left or right hippocampus. This suggests that a more comprehensive representation, encompassing both hemispheres of the hippocampus, is more effective for AD prediction in the dataset. The integration of features from both sides likely provides a richer and more nuanced understanding of the brain's structure and function, which appears to be beneficial to the ResNet to classify AD.
- **Contrast with Previous Studies:** Similar to the results from DenseCNN, the findings slightly deviate from other research suggesting that the left hippocampus is a more reliable indicator for AD [14], [15]. In the dataset, the right hippocampus alone, as per the F1 score, shows marginally better predictive performance than the left. This discrepancy may arise from variations in dataset characteristics, disease progression stages in subjects, or other underlying factors.

2) Impact of Augmentation: While examining the performance of ResNet in predicting AD, it becomes evident that data augmentation, although beneficial, contributes less dramatically to its performance compared to its impact on the DenseCNN model. This can be largely attributed to ResNet's use of pre-trained weights from MedicalNet.

- **Moderate Improvement with Augmentation:** While augmenting the dataset improves ResNet's performance metrics, the improvement is relatively modest. This is in contrast to the DenseCNN model, where augmentation significantly improves performance.
- **Influence of pre-trained weights:** The reason for this limited impact lies in the underlying strength of MedicalNet's pre-trained weights. Leveraging this extensive pre-training, ResNet is already equipped with a robust understanding of a wide range of medical image features. This pre-existing knowledge base diminishes the incremental benefits that augmentation typically provides in terms of exposure to diverse data patterns.
- **Greater impact on combined features:** Interestingly, the impact of augmentation is more pronounced when evaluating the "both" feature set (combining left and right hippocampus). Here a significant increase in the F1 score, about 0.6 is observed, suggesting that the model benefits

more from augmentation when dealing with a richer, combined feature set. This suggests that the integration of left and right hippocampal features provides a more comprehensive representation of AD prediction, which is further enhanced by augmentation.

- **Dataset Size and Diversity Relative to MedicalNet:** The dataset, even with augmentation, remains relatively small and less diverse compared to the massive and diverse datasets used in MedicalNet training. Consequently, the additional variation introduced by augmentation in the dataset does not significantly alter the feature landscape for ResNet as it does for a model like DenseCNN trained from scratch on this specific dataset.

C. Comparative Conclusion

The comparison between DenseCNN and ResNet models in the context of AD prediction reveals nuanced differences and similarities that are crucial for understanding the effectiveness of each approach.

DenseCNN demonstrates a significant performance boost with data augmentation, indicating its sensitivity to diverse training examples and a strong ability to generalize from augmented data. In contrast, ResNet shows a less pronounced yet positive response to augmentation, which can be attributed to the pre-existing knowledge from MedicalNet's pre-trained weights. This indicates that while ResNet benefits from augmentation, its foundational training on a diverse dataset reduces the relative impact of additional data diversity.

In terms of feature-specific insights, DenseCNN performs best with the left hippocampus as input under augmentation but shows a shift in performance favoring the right hippocampus without augmentation. This variability underscores the model's sensitivity to the data it is trained on. Conversely, ResNet exhibits superior performance with dual-channel input, combining both left and right hippocampus. This suggests that a comprehensive representation of the hippocampal regions offers a more effective approach for AD prediction in this model.

Regarding contextual interpretation and research alignment, DenseCNN aligns with existing research [14] under augmented conditions, indicating left hippocampal atrophy as a significant biomarker for AD. However, the shift in performance without augmentation challenges this alignment, suggesting a need to consider dataset-specific characteristics. ResNet deviates from conventional beliefs in this study, emphasizing the importance of holistic feature consideration and the potential differential roles of hippocampal hemispheres in AD prediction.

Model training and stability also present contrasting aspects. DenseCNN shows relative instability in training compared to ResNet as seen in 7, highlighting its dependence on dataset specifics and augmentation for optimal performance. On the other hand, ResNet benefits from the stability and robust foundational knowledge provided by MedicalNet's pre-training, making it less susceptible to training fluctuations and better equipped to handle diverse scenarios.

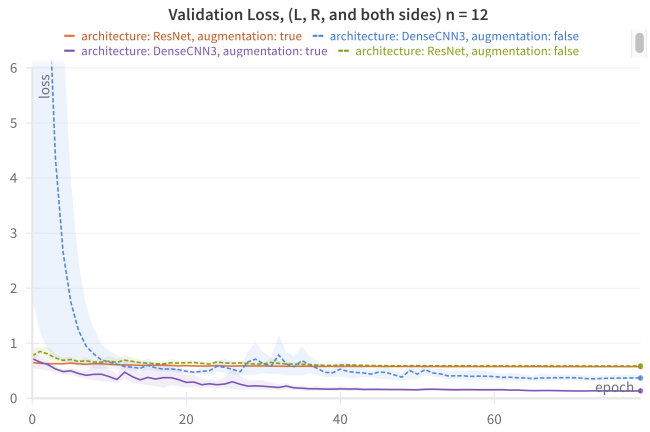

Fig. 7. Validation Loss for DenseCNN and ResNet over epochs

These findings call for more personalized approaches in AD prediction models, considering the unique characteristics of each dataset and the disease's manifestation. The differential roles of hippocampal hemispheres in AD and their representation in predictive modeling open avenues for further research. In conclusion, while DenseCNN and ResNet both exhibit strong capabilities in AD prediction, their performance characteristics highlight the importance of model architecture, data augmentation, feature selection, and pre-training in determining their effectiveness. The comparative analysis of these models offers valuable insights for future research and model development in medical imaging analysis, emphasizing the need for tailored approaches based on specific dataset characteristics and disease manifestations.

VI. Conclusion and Future Work

The study's exploration of DenseCNN and ResNet models in the diagnosis of AD via hippocampal MRI segmentation has yielded interesting insights. The research demonstrated the efficacy of both DenseCNN and ResNet models. It was observed that both DenseCNN and ResNet models improved in performance with the use of augmented data. Notably, DenseCNN exhibited a more substantial enhancement, highlighting its greater sensitivity to a diverse range of training examples and its robust ability to generalize from augmented data. Conversely, ResNet, strengthened by its pre-training on MedicalNet, exhibited a less pronounced but still positive response to data augmentation.

A key finding was the model's varying performance based on the features used. DenseCNN performed best with the left hippocampus as input under augmented conditions, aligning with existing research that suggests left hippocampal atrophy as a significant biomarker for AD. However, without augmentation, there was a notable shift favoring the right hippocampus. This variability underscores the sensitivity of the DenseCNN approach to the training data. In contrast, ResNet demonstrated superior performance with dual-channel input, combining both left and right hippocampus. This suggests

a more comprehensive representation of the hippocampal regions is beneficial for AD prediction in the ResNet model.

The study's results should be interpreted with caution due to limitations such as the small test set size, which may not fully represent the diversity of hippocampal structures in a broader population. Future work should aim to validate these findings across larger and more diverse and therefore larger datasets, enhancing the generalizability of the models. Additionally, further research could explore the differential roles of hippocampal hemispheres in AD and their representation in predictive modeling as well as the inclusion of MCI cases. This inclusion would further align the work with the inspirational papers [1] and [2] as they include the MCI cases. Based on the framework provided by this study, the challenges to implementing these additional changes are primarily time-based. This involves a change in the augmentation concept, as the MCI cases skew the distribution of cases and therefore require a different balancing strategy.

In conclusion, while DenseCNN and ResNet both exhibit strong capabilities in AD prediction, their performance characteristics highlight the importance of model architecture, data augmentation, feature selection, and pre-training in determining their effectiveness. The comparative analysis of these models offers valuable insights for future research and model development in medical imaging analysis, emphasizing the need for tailored approaches based on specific dataset characteristics and disease manifestations.

VII. Acknowledgment

We thank Prof. Dr. Arzu Çoltekin and Dr. Leticia Fernandez Moguel for their guidance and support during this semester's challenge. Their suggestions have significantly contributed to our work. We also thank Dr. Azizi Seixas and Dr. Alberto R. Ramos for their medical expertise and feedback. We would like to express our gratitude to Dr. Marie-Thérèse Rudolf von Rohr for their assistance and guidance in the writing and review process of this paper.

The funding for data collection and sharing in this project was provided by the Alzheimer's Disease Neuroimaging Initiative (ADNI), which is supported by the National Institutes of Health Grant U01 AG024904 and the Department of Defense ADNI (award number W81XWH-12-2-0012). ADNI receives funding from the National Institute on Aging and the National Institute of Biomedical Imaging and Bioengineering. This initiative also benefits from substantial contributions from various organizations including 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 along with 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 also contributes by funding ADNI clinical sites in Canada. Contributions from the private sector are coordinated through the Foundation for the National Institutes of Health (www.fnih.org). The Northern California Institute for Research and Education is the grantee organization, with the study being coordinated by the Alzheimer's Therapeutic Research Institute at the University of Southern California. Distribution of ADNI data is managed by the Laboratory for Neuro Imaging at the University of Southern California.

References

[1] S. Katabathula, Q. Wang, and R. Xu, "Predict Alzheimer's disease using hippocampus MRI data: a lightweight 3D deep convolutional network model with visual and global shape representations," *Alzheimer's Research & Therapy*, vol. 13, p. 104, May 2021. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8147046/

[2] Q. Wang, Y. Li, C. Zheng, and R. Xu, "DenseCNN: A Densely Connected CNN Model for Alzheimer's Disease Classification Based on Hippocampus MRI Data," *AMIA Annual Symposium Proceedings*, vol. 2020, pp. 1277–1286, Jan. 2021. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8075423/

[3] S. Chen, K. Ma, and Y. Zheng, "Med3D: Transfer Learning for 3D Medical Image Analysis," Jul. 2019, arXiv:1904.00625 [cs]. [Online]. Available: http://arxiv.org/abs/1904.00625

[4] V. Dhikav and K. Anand, "Potential Predictors of Hippocampal Atrophy in Alzheimer's Disease," *Drugs & Aging*, vol. 28, no. 1, pp. 1–11, Jan. 2011. [Online]. Available: https://doi.org/10.2165/11586390-000000000-00000

[5] N. C. Fox, E. K. Warrington, P. A. Freeborough, P. Hartikainen, A. M. Kennedy, J. M. Stevens, and M. N. Rossor, "Presymptomatic hippocampal atrophy in Alzheimer's disease: A longitudinal MRI study," *Brain*, vol. 119, no. 6, pp. 2001–2007, Dec. 1996. [Online]. Available: https://doi.org/10.1093/brain/119.6.2001

[6] W. Henneman, J. D. Sluimer, J. Barnes, W. M. Van Der Flier, I. C. Sluimer, N. C. Fox, P. Scheltens, H. Vrenken, and F. Barkhof, "Hippocampal atrophy rates in Alzheimer disease: Added value over whole brain volume measures," *Neurology*, vol. 72, no. 11, pp. 999–1007, Mar. 2009. [Online]. Available: https://www.neurology.org/doi/10.1212/01.wnl.0000344568.09360.31

[7] J. P. Kesslak, O. Nalcioglu, and C. W. Cotman, "Quantification of magnetic resonance scans for hippocampal and parahippocampal atrophy in Alzheimer's disease," *Neurology*, vol. 41, no. 1, pp. 51–51, Jan. 1991. [Online]. Available: https://www.neurology.org/doi/10.1212/WNL.41.1.51

[8] S. F. Eskildsen, P. Coupé, V. S. Fonov, J. C. Pruessner, and D. L. Collins, "Structural imaging biomarkers of Alzheimer's disease: predicting disease progression," *Neurobiology of Aging*, vol. 36, pp. S23–S31, Jan. 2015. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S019745801400548X

[9] C. R. Jack Jr., M. A. Bernstein, N. C. Fox, P. Thompson, G. Alexander, D. Harvey, B. Borowski, P. J. Britson, J. L. Whitwell, C. Ward, A. M. Dale, J. P. Felmlee, J. L. Gunter, D. L. Hill, R. Killiany, N. Schuff, S. Fox-Bosetti, C. Lin, C. Studholme, C. S. DeCarli, G. Krueger, H. A. Ward, G. J. Metzger, K. T. Scott, R. Mallozzi, D. Blezek, J. Levy, J. P. Debbins, A. S. Fleisher, M. Albert, R. Green, G. Bartzokis, G. Glover, J. Mugler, and M. W. Weiner, "The Alzheimer's disease neuroimaging initiative (ADNI): MRI methods," *Journal of Magnetic Resonance Imaging*, vol. 27, no. 4, pp. 685–691, 2008, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/jmri.21049. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/jmri.21049

[10] "ADNI." [Online]. Available: https://adni.loni.usc.edu/data-samples/access-data/

[11] B. Thyreau, "hippodeep," Dec. 2023, original-date: 2016-12-06T05:55:53Z. [Online]. Available: https://github.com/bthyreau/hippodeep

[12] ——, "hippodeep_pytorch," Oct. 2023, original-date: 2020-02-26T03:59:08Z. [Online]. Available: https://github.com/bthyreau/hippodeep_pytorch

[13] B. Thyreau, K. Sato, H. Fukuda, and Y. Taki, "Segmentation of the hippocampus by transferring algorithmic knowledge for large cohort processing," *Medical Image Analysis*, vol. 43, pp. 214–228, Jan. 2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1361841517301597

[14] A. Scher, Y. Xu, E. Korf, L. White, P. Scheltens, A. Toga, P. Thompson, S. Hartley, M. Witter, D. Valentino, and L. Launer, "Hippocampal shape analysis in Alzheimer's disease: A population-based study," *NeuroImage*, vol. 36, no. 1, pp. 8–18, May 2007. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S105381190601216X

[15] C. Murphy, T. L. Jernigan, and C. Fennema-Notestine, "Left hippocampal volume loss in Alzheimer's disease is reflected in performance on odor identification: A structural MRI study," *Journal of the International Neuropsychological Society*, vol. 9, no. 3, pp. 459–471, Mar. 2003. [Online]. Available: https://www.cambridge.org/core/product/identifier/S1355617703930116/type/journal_article

[16] B. A. Ardekani, A. Convit, and A. H. Bachman, "Analysis of the MIRIAD Data Shows Sex Differences in Hippocampal Atrophy Progression," *Journal of Alzheimer's Disease*, vol. 50, no. 3, pp. 847–857, Feb. 2016. [Online]. Available: https://www.medra.org/servlet/aliasResolver?alias=iospress&doi=10.3233/JAD-150780

[17] CSuperlei, "DeepHipp," Dec. 2023, original-date: 2019-07-21T09:27:11Z. [Online]. Available: https://github.com/CSuperlei/DeepHipp

[18] A. Azulay and Y. Weiss, "Why do deep convolutional networks generalize so poorly to small image transformations?" Dec. 2019, arXiv:1805.12177 [cs]. [Online]. Available: http://arxiv.org/abs/1805.12177

[19] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, "Densely Connected Convolutional Networks," Jan. 2018, arXiv:1608.06993 [cs]. [Online]. Available: http://arxiv.org/abs/1608.06993

[20] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," Dec. 2015, arXiv:1512.03385 [cs]. [Online]. Available: http://arxiv.org/abs/1512.03385

[21] M. W. Oktavian, N. Yudistira, and A. Ridok, "Classification of Alzheimer's Disease Using the Convolutional Neural Network (CNN) with Transfer Learning and Weighted Loss," Jul. 2022, arXiv:2207.01584 [cs, eess]. [Online]. Available: http://arxiv.org/abs/2207.01584

[22] B. H. Menze, A. Jakab, S. Bauer, J. Kalpathy-Cramer, K. Farahani, J. Kirby, Y. Burren, N. Porz, J. Slotboom, R. Wiest, L. Lanczi, E. Gerstner, M.-A. Weber, T. Arbel, B. B. Avants, N. Ayache, P. Buendia, D. L. Collins, N. Cordier, J. J. Corso, A. Criminisi, T. Das, H. Delingette, Demiralp, C. R. Durst, M. Dojat, S. Doyle, J. Festa, F. Forbes, E. Geremia, B. Glocker, P. Golland, X. Guo, A. Hamamci, K. M. Iftekharuddin, R. Jena, N. M. John, E. Konukoglu, D. Lashkari, J. A. Mariz, R. Meier, S. Pereira, D. Precup, S. J. Price, T. R. Raviv, S. M. S. Reza, M. Ryan, D. Sarikaya, L. Schwartz, H.-C. Shin, J. Shotton, C. A. Silva, N. Sousa, N. K. Subbanna, G. Szekely, T. J. Taylor, O. M. Thomas, N. J. Tustison, G. Unal, F. Vasseur, M. Wintermark, D. H. Ye, L. Zhao, B. Zhao, D. Zikic, M. Prastawa, M. Reyes, and K. Van Leemput, "The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS)," *IEEE transactions on medical imaging*, vol. 34, no. 10, pp. 1993–2024, Oct. 2015.

[23] M. Antonelli, A. Reinke, S. Bakas, K. Farahani, A. Kopp-Schneider, B. A. Landman, G. Litjens, B. Menze, O. Ronneberger, R. M. Summers, B. van Ginneken, M. Bilello, P. Bilic, P. F. Christ, R. K. G. Do, M. J. Gollub, S. H. Heckers, H. Huisman, W. R. Jarnagin, M. K. McHugo, S. Napel, J. S. G. Pernicka, K. Rhode, C. Tobon-Gomez, E. Vorontsov, J. A. Meakin, S. Ourselin, M. Wiesenfarth, P. Arbeláez, B. Bae, S. Chen, L. Daza, J. Feng, B. He, F. Isensee, Y. Ji, F. Jia, I. Kim, K. Maier-Hein, D. Merhof, A. Pai, B. Park, M. Perslev, R. Rezaiifar, O. Rippel, I. Sarasua, W. Shen, J. Son, C. Wachinger, L. Wang, Y. Wang, Y. Xia, D. Xu, Z. Xu, Y. Zheng, A. L. Simpson, L. Maier-Hein, and M. J. Cardoso, "The Medical Segmentation Decathlon," *Nature Communications*, vol. 13, no. 1, p. 4128, Jul. 2022, number: 1 Publisher: Nature Publishing Group. [Online]. Available: https://www.nature.com/articles/s41467-022-30695-9

[24] C. Tobon-Gomez, A. J. Geers, J. Peters, J. Weese, K. Pinto, R. Karim, M. Ammar, A. Daoudi, J. Margeta, Z. Sandoval, B. Stender, n. Yefeng Zheng, M. A. Zuluaga, J. Betancur, N. Ayache, M. Amine Chikh, J.-L. Dillenseger, B. M. Kelm, S. Mahmoudi, S. Ourselin, A. Schlaefer, T. Schaeffter, R. Razavi, and K. S. Rhode, "Benchmark for Algorithms Segmenting the Left Atrium From 3D CT and MRI Datasets," *IEEE transactions on medical imaging*, vol. 34, no. 7, pp. 1460–1473, Jul. 2015.

[25] "Liver tumor segmentation challenge." [Online]. Available: https://competitions.codalab.org/competitions/17094#results

[26] W. Falcon and The PyTorch Lightning team, "PyTorch Lightning," Mar. 2019. [Online]. Available: https://github.com/Lightning-AI/lightning

[27] M. K. Bohmrah and H. Kaur, "Classification of Covid-19 patients using efficient fine-tuned deep learning DenseNet model," *Global Transitions Proceedings*, vol. 2, no. 2, pp. 476–483, Nov. 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2666285X21000315

[28] K. Ma, L. Sun, Y. Wang, and J. Wang, "Classification of blood cancer images using a convolutional neural networks ensemble," in *Eleventh International Conference on Digital Image Processing (ICDIP 2019)*, vol. 11179. SPIE, Aug. 2019, pp. 9–14. [Online]. Available: https://www.spiedigitallibrary.org/conference-proceedings-of-spie/11179/1117903/Classification-of-blood-cancer-images-using-a-convolutional-neural-networks/10.1117/12.2539605.full

APPENDIX A

The code used to obtain the results in this paper is freely available at the following link: https://gitlab.fhnw.ch/cml1-3da-team-denn/pipeline.

Access the data needed to quickly run the pipeline and model training, if permission is granted, at: https://drive.switch.ch/index.php/f/7135136976.

The code is open-source and available for use and modification under the MIT License.

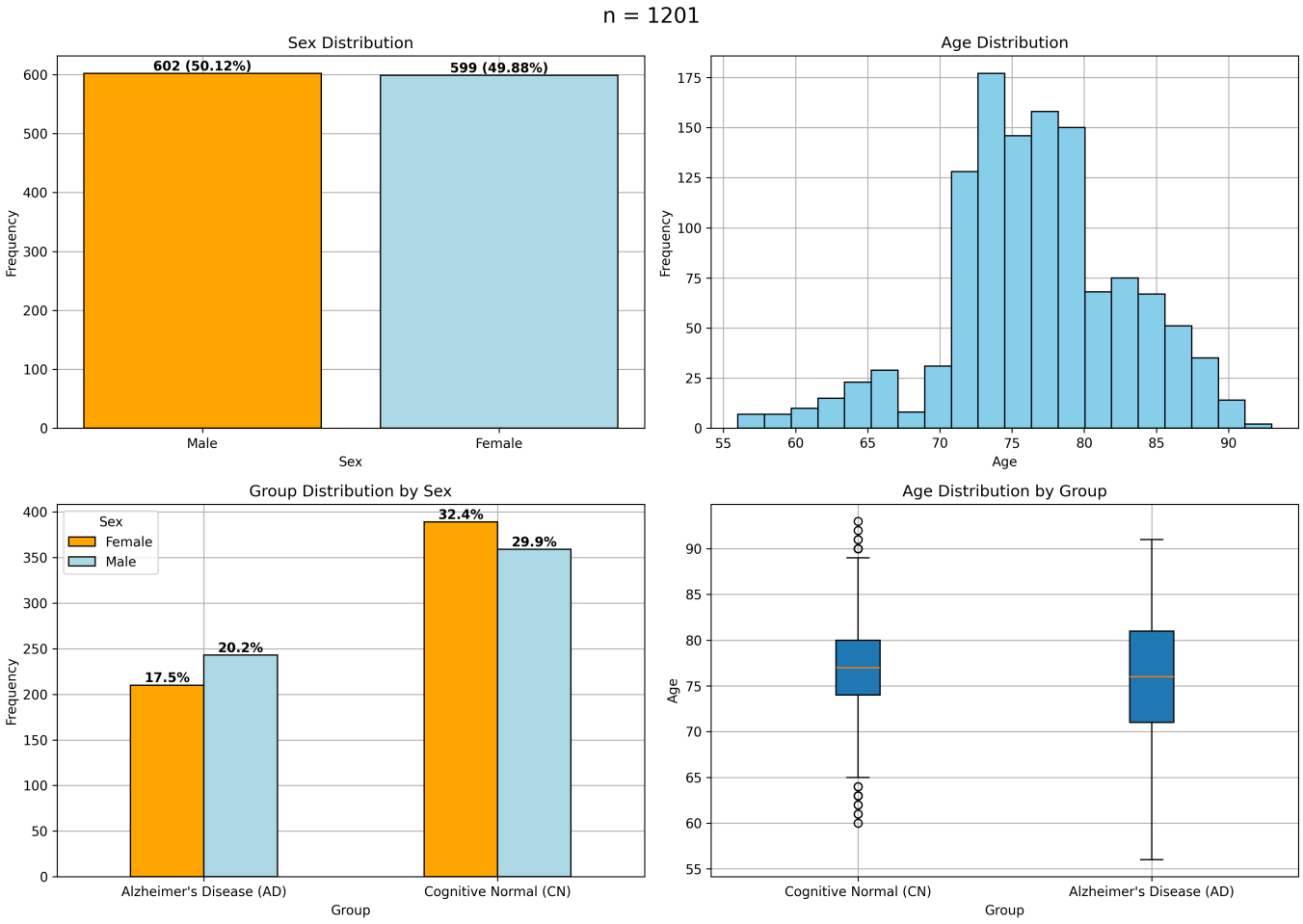

Fig. 8. Larger version of Figure 2

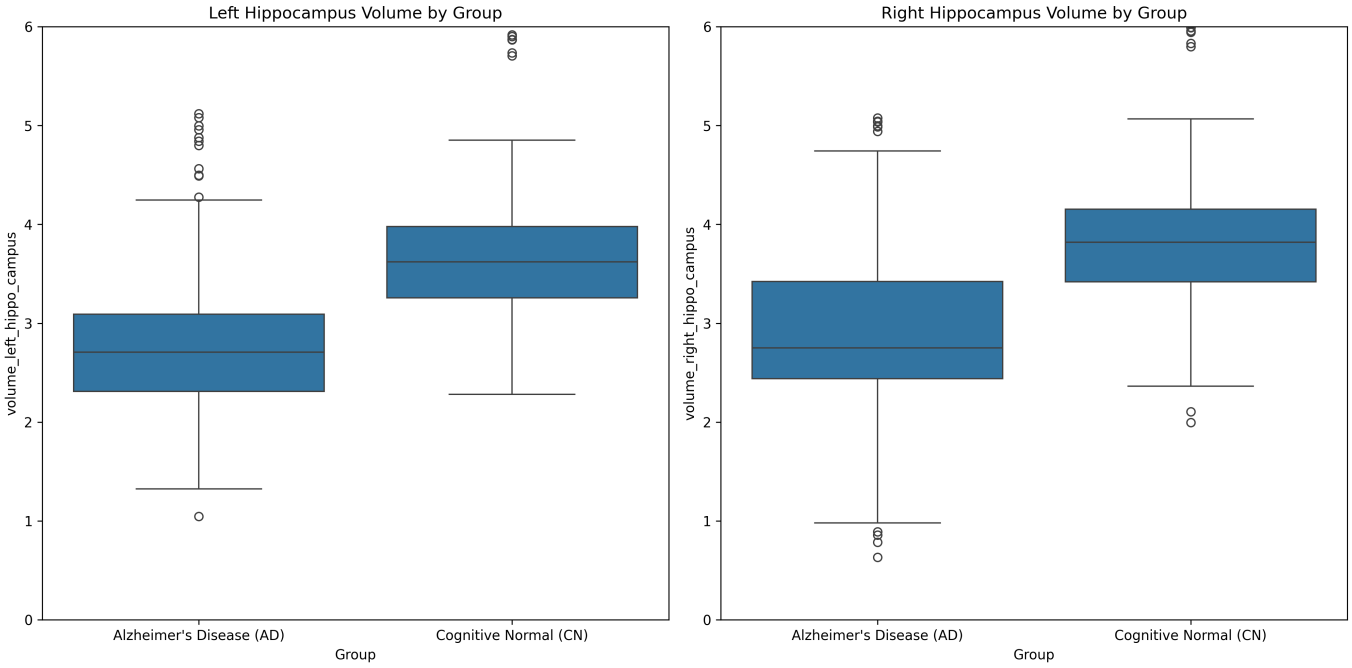

Fig. 9. Larger version of Figure 4

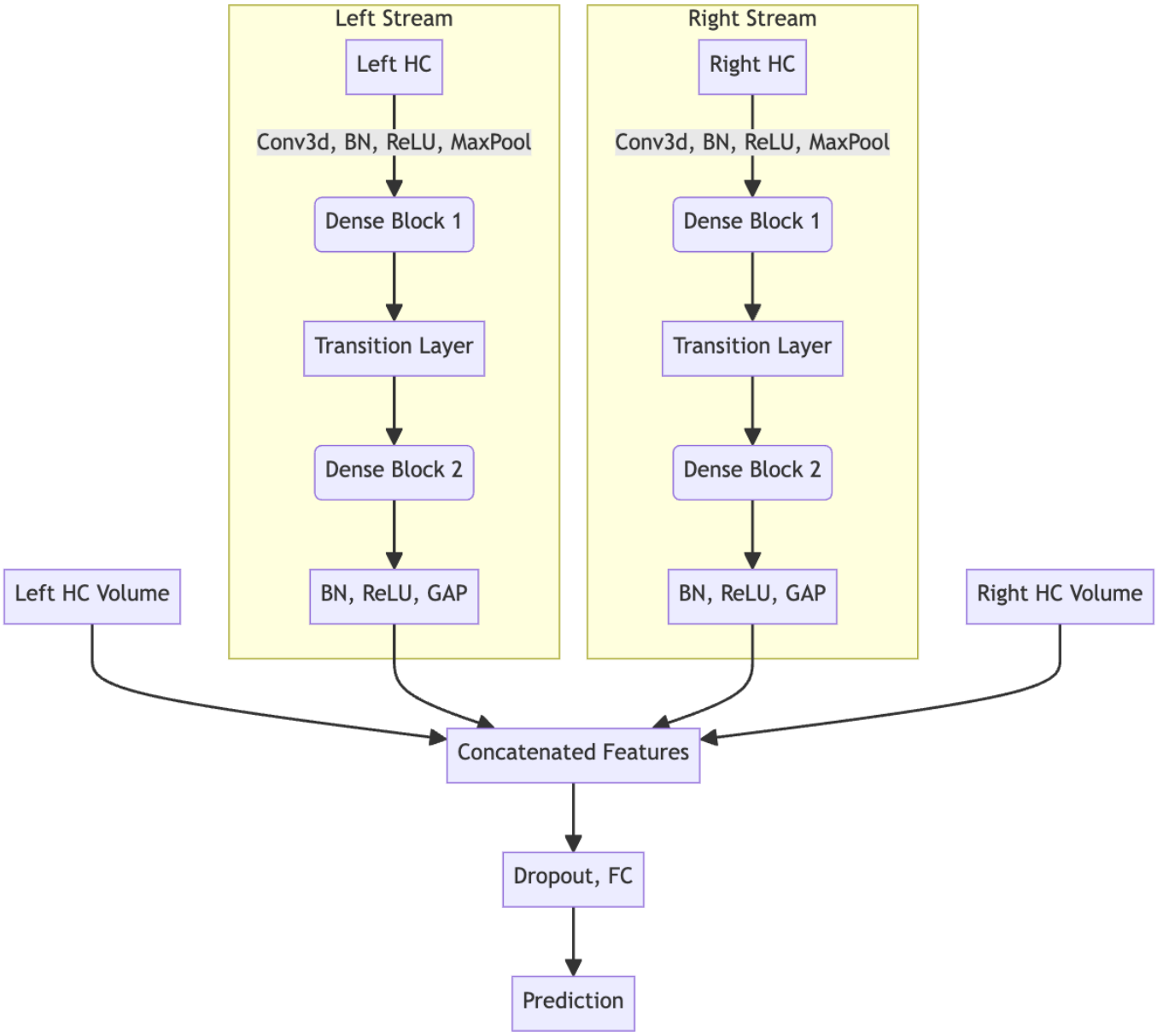

Fig. 10. Larger version of Figure 6

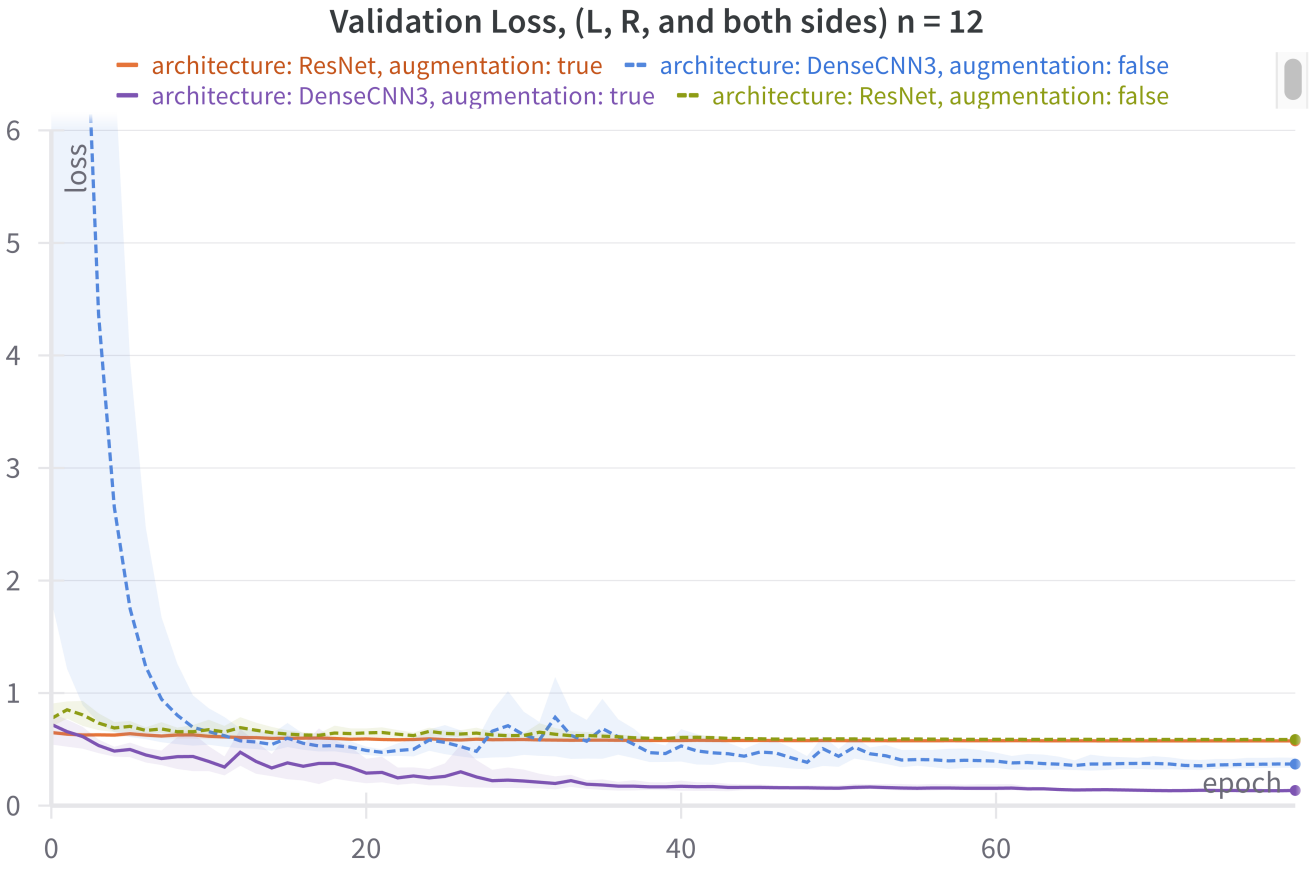

Fig. 11. Larger version of Figure 7

APPENDIX B

A. Author contribution statements

1) General: Across the board did all members of this group extensively proofread sections of others to ensure understanding and coherence across the board. In addition, the following contribution statements are not to be viewed as absolute stand-alone work; each member of the team was involved in the development process of all components to some degree: Either as a contributor, peer reviewer, or with meaningful advice that led to the results achieved. However, focal points for each member were clearly established to streamline and accelerate productivity among all resources wherever possible.

Noah Leuenberger was the driving factor in organizing and configuring the study's data structure, ensuring it was dynamic and adaptable to the team's needs. He co-developed the augmentation concept, applying it effectively in the study's context. Noah took the lead in hippocampus segmentation, involving the selection of suitable segmentation models, coding for automatic segmentation, and the preprocessing of hippocampus segments to prepare them for model training. His work in standardizing the data preparation for the study's models through the self-implemented data pipeline was crucial. He was primarily responsible for implementing the DenseCNN model across all iterations, utilizing PyTorch Lightning, and adapting insights from inspirational papers and his explorations. This process included extensive hyperparameter tuning and experimentation with different model configurations. Additionally, Noah's contribution to the writing of the paper was significant across all sections, especially in the description of methods (focusing on the Segmentation and DenseCNN), and the results.

Dominik Filliger played a pivotal role in conceptualizing the training strategy, including finding a suitable environment and tools to train and evaluate the models of the study. He created the full implementation and adaption of the ResNet and MedicalNet, alongside extensive hyperparameter tuning and experimenting with model configurations. Dominik also did the dual-channel logic of the ResNet, necessary for the comparison with the DenseCNN in the context of the study (single and dual-channel comparison). In addition, the PyTorch Lightning setup, used to train both ResNet and DenseCNN was created by him. This included the definition of all metrics and a general approach to measuring model performance, logging of all this information (Weights & Biases), and the research behind it. Dominik's contributions extend to a multitude of sections. In particular writing sections of the paper in the aspect of models (especially ResNet), methodology, overall results as well as conclusions.

Nils Fahrni went out of his way to research foundational and inspirational literature, which laid the foundation for our paper (finding first inspiration paper [1]). His development and implementation of the augmentation framework not only served this study but also set the stage for future research building upon this paper. In the early development, Nils also contributed to the first version of the DenseCNN model. On top of that he also looked into the implementation of a more common CNN model which was insightful for future work done with the final models implemented. Nils additionally contributed by summarizing the study, setting up the base layout of the paper and by authoring multiple sections of the paper, focusing on the preprocessing sections as well as the introduction.

Etienne Roulet explored various potential approaches for prediction using the entire brain, including techniques like skull stripping, although these explorations did not make it to the final paper they provided valuable insights for the work that followed it. His efforts in data exploration were integral to understanding and framing our research context and objectives. During the initial phases, he developed a CNN model including skull stripping to further explore the capabilities of the CNN approach in general. Additionally, the skeleton of the data- and the deep learning pipeline, set an important start to the final version implemented to streamline the whole training process of the models. Etienne also contributed to the project through Exploratory Data Analysis (EDA), notably in assessing sex, age and hippocampi volume distribution between cognitively normal individuals and Alzheimer's disease patients.

B. Project Environment

In our project management framework, we implemented a comprehensive and collaborative approach using GitLab's Kanban board to streamline our development workflow. Our methodology was deeply integrated with Git, where we meticulously organized our tasks through issues and branches, ensuring every feature and bug fix was tracked and version-controlled. This structure facilitated peer programming sessions, where team members paired up to tackle Kanban tickets and corresponding issues together, fostering a robust learning environment and promoting high-quality code outputs. Following the completion of these sessions, we conducted thorough code reviews, a practice that not only enhanced our codebase's integrity but also reinforced our team's coding standards and practices.

The rhythm of our project was marked by weekly meetings every Friday, notable for their perfect attendance record. These sessions were not just administrative check-ins but vital opportunities for the team to discuss outcomes, share insights, and align on the upcoming week's objectives. Our Kanban approach underpinned our project management strategy, enabling us to maintain flexibility and adaptability in our workflow while keeping a clear focus on our progress and priorities.

Communication played a pivotal role in our project's success, with Discord serving as our main channel for detailed discussions and collaborative decision-making. For more spontaneous or urgent matters, we turned to WhatsApp, ensuring that no query or update fell through the cracks. Additionally, we leveraged Weights & Biases for meticulous documentation and performance tracking of our models. This tool was instrumental in providing us with the insights needed to evaluate our models' improvements over time, serving as our score documenter.

Initially, our process was heavily reliant on pair programming, with each pair taking responsibility for a single Kanban ticket and issue. This collaborative effort concluded with a peer-led code review, ensuring a thorough scrutiny of the work before it was considered complete. However, as our project evolved and our team became more proficient, we transitioned to a more independent workflow, with each member taking on individual coding tasks. This shift allowed us to accelerate our development pace and tackle a broader range of issues simultaneously, demonstrating our team's growth and adaptability.

C. German Abstract

In dieser Studie wird die Effektivität von Dense Convolutional Neural Networks (DenseCNN) und Residual Networks (ResNet) bei der Diagnose der Alzheimer-Krankheit durch MRT-Segmentierung des Hippocampus untersucht. Es wird der Datensatz der Alzheimer's Disease Neuroimaging Initiative (ADNI) verwendet, um die Leistung dieser beiden neuronalen Netzwerkmodelle bei der Erkennung von Alzheimer zu vergleichen. DenseCNN und ResNet wurden hinsichtlich ihrer Vorhersagefähigkeiten analysiert, sowohl mit als auch ohne Datenaugmentation. Die Studie zeigt, dass DenseCNN mit erweiterten Daten besser abschneidet, was auf seine Empfindlichkeit gegenüber verschiedenen Trainingsbeispielen hinweist. ResNet, das durch Vortraining auf MedicalNet verbessert wurde, profitiert ebenfalls von der Datenerweiterung, allerdings in geringerem Masse. Die Studie zeigt, dass die Wahl der Merkmale die Leistung des Modells beeinflusst. DenseCNN mit dem linken Hippocampus unter Augmentation und ResNet mit Zweikanal-Input schneiden besser ab. Die Ergebnisse deuten darauf hin, dass eine umfassende Darstellung der Hippocampus-Regionen die Vorhersagegenauigkeit von Alzheimer verbessert. Diese Studie betont die Wichtigkeit der Modellarchitektur, der Datenaugmentation, der Merkmalsauswahl und des Vortrainings für die Effektivität neuronaler Netze bei der medizinischen Bildanalyse von Alzheimer. Zukünftige Forschungsarbeiten sollten massgeschneiderte Ansätze auf der Grundlage spezifischer Datensatzmerkmale und Krankheitsmanifestationen untersuchen. Dies wird das Potenzial erweitern, Alzheimer frühzeitig und genau zu diagnostizieren, mithilfe fortschrittlicher neuronaler Netzwerkmodelle.