



**INTELIGÊNCIA COMPUTACIONAL NO
APOIO AO DIAGNÓSTICO DA DOENÇA DE
ALZHEIMER**

RAIMUNDO GUIMARÃES SARAIVA JUNIOR

**TESE DE DOUTORADO
EM ENGENHARIA ELÉTRICA**

DEPARTAMENTO DE ENGENHARIA ELÉTRICA

**FACULDADE DE TECNOLOGIA
UNIVERSIDADE DE BRASÍLIA**

Universidade de Brasília
Faculdade de Tecnologia
Departamento de Engenharia Elétrica

Inteligência computacional no apoio ao diagnóstico da doença de
Alzheimer

Raimundo Guimarães Saraiva Junior

TESE DE DOUTORADO SUBMETIDA AO PROGRAMA
DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA DA UNIVERSIDADE
DE BRASÍLIA COMO PARTE DOS REQUISITOS NECESSÁRIOS PARA A
OBTENÇÃO DO GRAU DE DOUTOR.

APROVADA POR:

Prof. Francisco A. O. Nascimento, D.Sc. (ENE-UnB)
(Orientador)

Prof. Henrique Sarmiento Malvar, Ph.D. (University of Washington)
(Examinador Externo)

Prof. José Antonio Apolinário Jr., D.Sc. (IME)
(Examinador Externo)

Prof. Daniel Guerreiro e Silva, D.Sc. (ENE-UnB)
(Examinador Interno)

Brasília/DF, dezembro de 2023.

FICHA CATALOGRÁFICA

SARAIVA JUNIOR, RAIMUNDO GUIMARÃES

Inteligência computacional no apoio ao diagnóstico da doença de Alzheimer. [Brasília/DF] 2023.

xxx, nnp., 210 x 297 mm (ENE/FT/UnB, Doutor, Tese de Doutorado, 2023).

Universidade de Brasília, Faculdade de Tecnologia, Departamento de Engenharia Elétrica.

Departamento de Engenharia Elétrica

- | | |
|---------------------------|----------------------------|
| 1. Alzheimer | 2. Demência |
| 3. Diagnóstico | 4. Ressonância magnética |
| 5. Aprendizado profundo | 6. Inteligência Artificial |
| 7. Aprendizado de máquina | |
| I. ENE/FT/UnB | II. Título (série) |

REFERÊNCIA BIBLIOGRÁFICA

SARAIVA JUNIOR, RAIMUNDO GUIMARÃES (2023). Inteligência computacional no apoio ao diagnóstico da doença de Alzheimer. Tese de Doutorado, Publicação PPGE.202/2023, Departamento de Engenharia Elétrica, Universidade de Brasília, Brasília, DF, xxxxp.

CESSÃO DE DIREITOS

AUTOR: Raimundo Guimarães Saraiva Junior

TÍTULO: Inteligência computacional no apoio ao diagnóstico da doença de Alzheimer.

GRAU: Doutor ANO: 2023

É concedida à Universidade de Brasília permissão para reproduzir cópias desta Tese de Doutorado e para emprestar ou vender tais cópias somente para propósitos acadêmicos e científicos. O autor reserva outros direitos de publicação e nenhuma parte desta dissertação de mestrado pode ser reproduzida sem autorização por escrito do autor.

Raimundo Guimarães Saraiva Junior

Universidade de Brasília (UnB)

Campus Darcy Ribeiro

Faculdade de Tecnologia - FT

Departamento de Engenharia Elétrica(ENE)

Brasília - DF CEP 70919-970

This thesis is dedicated especially to my brother José Flávio Sombra Saraiva, a retired professor at UnB, who was diagnosed with Alzheimer's at such an early age. To my son, Eduardo, to my siblings Gilson and Ana Gláucia, to my mother Maria Alice, to my father Raimundo Saraiva (in memoriam), and to Kaline Kelle, a special person

ACKNOWLEDGEMENTS

I would like to thank my supervisor, Francisco Assis de Oliveira Nascimento, for his advice, support and trust. To the coordinators and other colleagues at PPGEE/UnB for small and large contributions, and particularly to my colleague Tony Alexandre Medeiros.

Regarding medical topics, this thesis relied on the invaluable advice of the physician and researcher Yuri Almeida de Oliveira, whom I thank for his time and for all the information he provided.

To professors Geovany Araújo Borges and Lucas Vieira Barros, for encouraging and supporting my admission process to PPGEE/UnB.

To my colleagues in the Caríssimos brotherhood, for always following our intentions closely, with great attention and care.

Special thanks to Kaline Kelle Lima de Oliveira, for her invaluable support provided throughout the preparation of this thesis.

Finally, I would like to express my gratitude for the primary data supplied by the Australian Imaging, Biomarker & Lifestyle Flagship Study of Ageing (AIBL), the Alzheimers Disease Neuroimaging Initiative (ADNI), the Open Access Series of Imaging Studies (OASIS) and the University of Southern California for authorization to use the Image Data Archive (IDA) platform.

ABSTRACT

Accurate diagnosis of Alzheimer's disease remains difficult because of the complexity of the disease and its similarity to other forms of dementia. The prevalence of dementia, and its most common form Alzheimer's, is increasing worldwide, partly due to increasing longevity. While neurological deterioration is a normal part of aging, Alzheimer's accelerates this process, straining public health services, leading to early mortality, and generating tremendous challenges for patients and their families.

We used clinical data and magnetic resonance images to investigate artificial intelligence techniques that could assist in Alzheimer's disease diagnosis. Primary information was obtained from the Australian Imaging, the Biomarker & Lifestyle Flagship Study of Ageing (AIBL), the Alzheimer's Disease Neuroimaging Initiative (ADNI), and the Open Access Series of Imaging Studies (OASIS). To process these clinical data, we developed a new feature selection method, created reduced databases, evaluated the data via machine learning classification models and achieved 99.81% accuracy in diagnosing Alzheimer's for ADNI data. MRI images were also investigated. Different strategies were proposed for building an image dataset suitable for deep learning techniques and achieved 99.3% accuracy in diagnosing Alzheimer's for ADNI data. Evaluation metrics were computed for the proposed strategies, and diagnostic results were generated for normal and established Alzheimer's cases. Individuals with mild cognitive impairment were also evaluated. Finally, the results of the current study were compared to those from other studies.

Keywords: Alzheimer's disease, Machine learning, Feature selection, Deep learning, Multiple instances.

RESUMO

O diagnóstico preciso da doença de Alzheimer ainda é um campo aberto para a ciência, tendo em vista se tratar de uma doença multifatorial e com difícil identificação das causas. As demências estão cada vez mais presentes na população mundial, que experimenta um aumento significativo na expectativa de vida, o que explica, em parte, o aumento do diagnóstico de Alzheimer, que é o tipo de demência mais frequente. A decadência neurológica faz parte do processo natural do envelhecimento, mas a doença de Alzheimer acelera o processo de degeneração neurológica, trazendo grandes transtornos ao paciente e seus familiares, onerando os serviços públicos de saúde e causando óbitos precoces.

Nesse contexto, este trabalho investiga técnicas de inteligência artificial que busquem auxiliar no diagnóstico da doença de Alzheimer, utilizando dados de natureza clínica e imagens de ressonância magnética. Mediante licenciamento para utilização, foram buscadas informações primárias nos estudos científicos Australian Imaging, Biomarker & Lifestyle Flagship Study of Ageing (AIBL), Alzheimers Disease Neuroimaging Initiative (ADNI) e no Open Access Series of Imaging Studies (OASIS). Para tratamento dos dados clínicos, foi proposto um novo método de seleção de atributos, foram criadas bases de dados reduzidas e fez-se a avaliação por modelos de classificação de aprendizado de máquina, tendo atingido a acurácia de 99,81% no diagnóstico do Alzheimer para os dados ADNI. Também foi investigado o uso de imagens obtidas em exames de ressonância magnética. Foram propostas diferentes estratégias de construção de um banco de imagens adequadas para a aplicação de técnicas de aprendizado profundo, tendo atingido a acurácia de 99,3% no diagnóstico do Alzheimer para os dados ADNI. Métricas de avaliação foram computadas sobre as estratégias propostas, e resultados de diagnóstico foram obtidos para os casos de indivíduo normal e com Alzheimer estabelecido. Também foi avaliado o caso de indivíduos com comprometimento cognitivo leve. Como parte da avaliação de desempenho, os resultados alcançados foram comparados com os resultados obtidos em outros trabalhos de pesquisa disponíveis na literatura.

Palavras-chave: Alzheimer; Aprendizado de máquina; Seleção de atributos; Aprendizado profundo; Múltiplas instâncias.

CONTENTS

Table of contents	ii
List of figures	v
List of tables	x
Glossary	xii
Chapter 1 – Introduction	1
1.1 Alzheimer’s disease	1
1.1.1 Stages	3
1.1.2 A short history	5
1.1.3 Diagnosis	7
1.1.4 Alzheimer’s in Brazil	9
1.1.5 Alzheimer’s and Covid-19	9
1.2 Rationale for the present study	9
1.3 Objectives	10
1.4 Thesis organization	11
Chapter 2 – Theoretical foundation	13
2.1 Artificial intelligence in diagnosing neurodegenerative diseases	13
2.2 Computational support in diagnosing Alzheimer’s disease	15
2.3 Literature review	16
Chapter 3 – Analysis based on clinical data	20
3.1 Database descriptions	20
3.1.1 OASIS	20
3.1.2 AIBL	21
3.1.3 ADNI	22
3.1.4 Preprocessing	23
3.2 Data analysis	25

3.3	Feature selection	27
3.3.1	Filter methods	27
3.3.2	Wrapper methods	28
3.3.3	Embedded methods	28
3.3.4	The proposed algorithm	29
3.3.4.1	The FMFS algorithm	30
3.3.4.2	Applying FMFS to the databases	34
3.4	Performance metrics	38
3.4.1	Accuracy - A_{cy}	38
3.4.2	Precision - P_{rn}	38
3.4.3	Sensitivity - S_{ny}	39
3.4.4	Specificity - S_{py}	39
3.4.5	Balanced accuracy - B_{ca}	39
3.4.6	F1 score - F1	40
3.4.7	Area under the ROC curve - AUC	40
3.5	Nested cross-validation	41
3.6	ML models used	43
3.6.1	SVM: support vector machine	44
3.6.2	kNN: k-nearest neighbors	47
3.6.3	Other ML models	48
Chapter 4 – Image based analysis		49
4.1	AD diagnosis based on MRI	49
4.2	Magnetic resonance imaging	53
4.2.1	Digital format of magnetic resonance images	54
4.2.1.1	NIFTI format	55
4.3	Database description	56
4.3.1	ADNI	56
4.3.1.1	MRI in ADNI	57
4.3.1.2	Slice selection	58
4.4	Axial plane evaluation	59
4.4.1	Proposed slice selection algorithm	59
4.4.1.1	SMR algorithm	60
4.4.1.2	Applying SMR to the data base	62
4.4.2	Resulting image dataset	63
4.5	Evaluation in the coronal plane	64
4.5.1	Strategy for building the image dataset	64

4.5.2	Multiple instance evaluation	66
4.5.3	Resulting image dataset	67
4.6	DL models	69
4.6.1	Classical CNN models	70
4.6.2	Proposed CNN models	71
Chapter 5 – Results and discussion		79
5.1	Clinical data analysis results	79
5.1.1	Feature selection in the AIBL database	79
5.1.2	Feature selection in the ADNI database	80
5.1.3	Feature selection in the OASIS database	80
5.1.4	Performance evaluation for three classes	81
5.1.4.1	ADNI evaluation	81
5.1.4.2	AIBL evaluation	82
5.1.4.3	OASIS evaluation	84
5.1.5	Performance evaluation for two classes	85
5.1.5.1	ADNI evaluation	85
5.1.5.2	AIBL evaluation	87
5.1.5.3	OASIS evaluation	88
5.1.6	Performance evaluation with other classifiers	89
5.1.7	Evaluation with other feature selection methods	90
5.2	Image analysis results	92
5.2.1	Analysis of the coronal image dataset	92
5.2.1.1	Partition setup alternatives	93
5.2.1.2	Evaluation of the S1 partition	94
5.2.1.3	Evaluation of the S2 partition	96
5.2.1.4	S2 reevaluation with augmented dataset	97
5.2.1.5	S1 reevaluation with majority voting classifier	98
5.2.2	Analysis of the axial image dataset	100
5.3	Discussion of the results of the clinical data analysis	100
5.3.1	Comparisons with other studies	102
5.3.2	Other considerations regarding clinical data analysis	105
5.4	Discussion of the image analysis results	106
5.4.1	Coronal plane	106
5.4.2	Axial plane	109
5.4.3	Comparison with other studies	110
5.5	Suggested application for new technologies	111

5.5.1	Brazilian Unified Health System - SUS	111
5.5.2	Biomedical Engineering Innovation Award for SUS 2023	112
5.5.3	Winning proposal	112
Chapter 6 – Conclusions		115
References		117
Appendix A – Portuguese thesis		125

LIST OF FIGURES

1.1	Types of dementia.	1
1.2	Images with several medial temporal atrophy scores (MTA).	3
1.3	Evolution of AD diagnosis in Olmsted County, Minnesota - USA.	6
1.4	Life expectancy at birth: country estimates, 1950-2021, and projections, 2022-2050	7
1.5	PubMed portal search results with the word "Alzheimer's" from 1980 to September 2022.	8
2.1	A Venn diagram showing how deep learning is a kind of representation learning, which is in turn a kind of machine learning that is used for many but not all approaches to AI.	14
2.2	Article distributions with respect to ML / DL.	15
2.3	Databases used in studies on AD diagnosis.	16
3.1	Feature selection methods.	27
3.2	Class histograms for the LIMMTOTAL(a) and AXT117(b) features in the AIBL dataset.	30
3.3	Nomenclature definition.	31
3.4	Features with different class means.	32
3.5	FMFS flowchart.	33
3.6	AIBL feature ranking - <i>dca</i> values.	35

3.7	AIBL performance measurement. a) The orange and blue curves represent the evolution of SVM and kNN accuracy, respectively, with indicated maximums. b) The green curve represents the evolution of the number of selected features. The orange and blue vertical lines represent the point of maximum accuracy for SVM and kNN respectively.	35
3.8	ADNI feature ranking - <i>dca</i> values.	36
3.9	ADNI performance measurement. a) The orange and blue curves represent the evolution of SVM and kNN accuracy, respectively, with indicated maximums. b) The green curve represents the evolution of the number of selected features. The orange and blue vertical lines represent the point of maximum accuracy for SVM and kNN respectively.	36
3.10	OASIS feature ranking - <i>dca</i> values.	37
3.11	OASIS performance measurement. a) The orange and blue curves represent the evolution of SVM and kNN accuracy, respectively, with indicated maximums. b) The green curve represents the evolution of the number of selected features. The orange and blue vertical lines represent the point of maximum accuracy for SVM and kNN respectively.	37
3.12	Confusion matrix.	38
3.13	Receiver Operating Characteristics (ROC) curve.	40
3.14	Train, validation and test data split.	41
3.15	Nested cross validation.	43
3.16	Generic machine learning process.	44
3.17	SVM algorithm - the support vectors, marked with grey squares, define the margin of largest separation between the two classes.	45
3.18	kNN algorithm - example with three classes and $k = 5$	47
4.1	Illustration of the hippocampus in the brain.	50
4.2	Three examples of MRI images (brain-extracted) of subjects from the ADNI (coronal slices).	50

4.3	Proposed hippocampus extraction scheme.	51
4.4	Extracted hippocampus.	51
4.5	Visualization of the hippocampus from various angles.. . . .	52
4.6	Maximum dimensions of the hippocampus in the 3 planes. Right hippocampus rotated 300 degrees.	53
4.7	MRI equipment (creative commons license BY-ND).	54
4.8	Orange lines representing axial MRI slices.	54
4.9	Axial (red bar), coronal (green bar) and sagittal (yellow bar) slice views of an MRI in NIFTI format.	56
4.10	Anatomy of the hippocampal formation: 1 = hippocampal head, 2 = hippocampal body, 3 = hippocampal tail, 4 = mesencephalon.	58
4.11	Axial slice showing the hippocampus.	59
4.12	Image matching.	60
4.13	Example SMR result. The red line in (a) shows the position of the selected slice with respect to the patient’s head. The range of interest on the mt curve (b) for the slice search is outlined in yellow. The image of the selected slice is shown in (c).	62
4.14	Template.	63
4.15	CN, MCI and AD image examples.	63
4.16	Red circles identify the left and right structures of the hippocampus in the coronal plane.	64
4.17	Preprocessing scheme	65
4.18	Coronal slices with 0.9442 mm spacing. Central, anterior, and posterior slices are highlighted.	66
4.19	Multiple instance decision scheme. In this case, a patient was evaluated and 11 instances indicated AD and 08 instances indicated CN. The majority decision is that the patient’s diagnosis is AD.	67

4.20	Coronal image dataset of central slices.	68
4.21	Coronal dataset with a center slice, 9 anterior slices, and 9 posterior slices. The slice ending in -10 is the central slice chosen.	69
4.22	Generic deep learning process.	70
4.23	CNN Evolution - examples.	70
4.24	CNN 1 model - total parameters: 4,534,866. Initial model.	72
4.25	CNN 2 model - total parameters: 717,034. Reduced dimensionality with respect to CNN-1.	73
4.26	CNN 3 model - total parameters: 18,421,666. Increased dimensionality with respect to CNN-1.	74
4.27	CNN 4 model - total parameters: 7,426,642. Increased depth compared to CNN-1.	75
4.28	CNN 5 model - total parameters: 1,786,090. Reduced dimensionality with respect to CNN-4.	76
4.29	CNNs used.	77
5.1	ADNI-11 avaliated with the SVM classifier - average confusion matrix.	82
5.2	AIBL-4 avaliated with the SVM classifier - average confusion matrix.	83
5.3	OASIS-5 avaliated with the SVM classifier - average confusion matrix.	84
5.4	ADNI-11 avaliated with the SVM classifier over CN v/s AD classes - ROC curves.	86
5.5	ADNI-11 avaliated with the SVM classifier - CN v/s AD - average confusion matrix.	86
5.6	AIBL-4 avaliated with the SVM classifier - CN v/s AD - average confusion matrix.	87
5.7	OASIS-5 avaliated with the kNN classifier - C00 v/s C12 - average confusion matrix.	88
5.8	Split fold S1 and S2 description.	93
5.9	CNN-5 model learning curves. a) Accuracy model. b) AUC model. c) Loss model	95
5.10	CNN-5 confusion matrix.	96

5.11 CNN-4 confusion matrix.	97
5.12 AIBL - kNN enhancement.	101
5.13 ADNI - kNN enhancement.	101
5.14 CN ADNI 002_S_1261 patient heat map and AD ADNI 007_S_1248 patient heat map. Red boxes highlight the region of interest.	108
5.15 Heat map of the 19 images of patient ADNI 007_S_1248 diagnosed with AD. .	109
5.16 Biomedical Engineering Innovation Award for SUS 2023.	112

LIST OF TABLES

1.1	Clinical Dementia Rating (CDR) Scale	5
3.1	OASIS: Features descriptions	21
3.2	AIBL: Features descriptions	22
3.3	ADNI: Features descriptions	23
3.4	AIBL-18: dataset composition	24
3.5	ADNI-21: dataset composition	24
3.6	OASIS-8: dataset composition	24
3.7	AIBL-18: instance distribution by class	25
3.8	ADNI-21: instance distribution by class	26
3.9	OASIS-7: instance distribution by class	26
4.1	ADNI1:Screening 1.5T - Number of patients per class	57
5.1	Reduced AIBL dataset: selected features	80
5.2	Reduced ADNI dataset: selected features	80
5.3	Reduced OASIS dataset: selected features	81
5.4	ADNI dataset results - three classes	82
5.5	AIBL dataset results - three classes	83
5.6	OASIS dataset results - three classes	84
5.7	ADNI dataset results - two classes	85
5.8	AIBL dataset results - two classes	87

5.9	OASIS dataset results - two classes	88
5.10	AIBL-4 - performance evaluation	89
5.11	ADNI-11 - performance evaluation	90
5.12	OASIS-5 - performance evaluation	90
5.13	ADNI - feature selection method evaluation	91
5.14	AIBL - feature selection method evaluation	91
5.15	OASIS - feature selection method evaluation	92
5.16	Split fold composition S1 and S2	94
5.17	S1 split fold - performance evaluation	94
5.18	S2 split fold - performance evaluation	96
5.19	Data augmentation in S2 training dataset - performance evaluation	97
5.20	Customized cnn results	98
5.21	Pretrained models results	98
5.22	CNN-4 training results	99
5.23	Split axial dataset	100
5.24	Axial plan - performance evaluation	100
5.25	Comparison with previous study (AIBL dataset)	103
5.26	Comparison with previous studies (ADNI dataset)	104
5.27	Comparison with previous study (OASIS dataset)	105
5.28	Comparison with previous DL studies	110

GLOSSARY

AD	Alzheimer's disease
AI	Artificial intelligence
AIBL	Australian Imaging, Biomarker & Lifestyle Flagship Study of Ageing
ADNI	Alzheimers Disease Neuroimaging Initiative
ANALYZE	Analyze 7.5
BCA	Balanced classification accuracy
CN	Cognitively normal
CSV	Comma-separated values
CDR	Clinical dementia rating
CNN	Convolutional Neural Networks
DL	Deep learning
DICOM	Digital Imaging and Communications in Medicine
GPU	Graphics Processing Unit
kNN	K-nearest neighbors

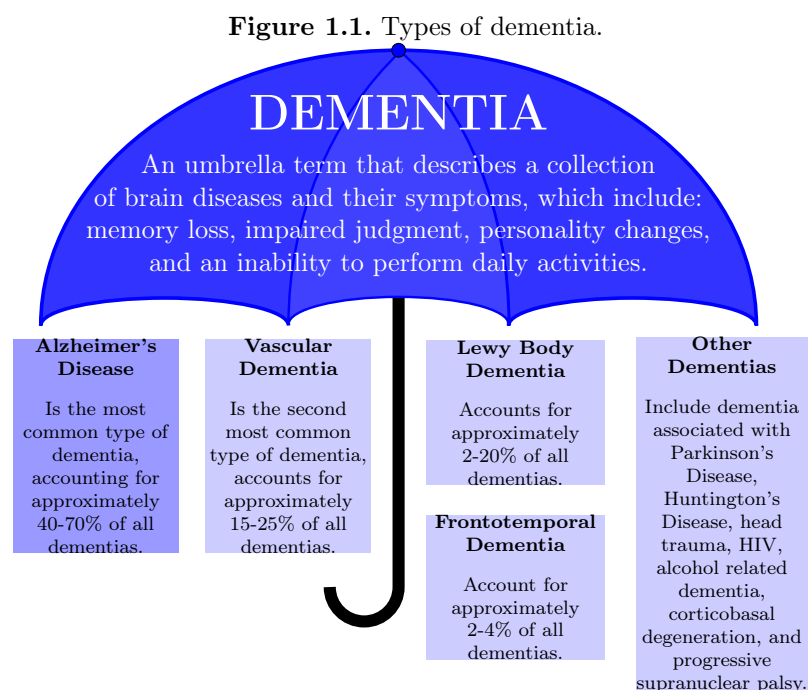
LSTM	Long short-term memory
MCI	Mild cognitive impairment
ML	Machine learning
MRI	Magnetic resonance imaging
MINC	Montreal Neurological Institute
MTA	Medial temporal lobe atrophy
NIFTI	Neuroimaging Informatics Technology Initiative
NRRD	Nearly Raw Raster Data
PET	Positron Emission Tomography
PAR/REC	Philips MRI scanner formats
RNN	Recurrent neural networks
RepL	Representation Learning
SVM	Support vector machine

This chapter presents a historical review, contextualization, and justification for our study, and defines the objectives, scientific contribution and organization of the thesis.

1.1 ALZHEIMER'S DISEASE

Alzheimer's disease (AD) is a progressive cognitive dementia that progresses from mild cognitive impairment to a final acute phase that robs patients of their independence and functional capacity (WELLER J., 2018). The word dementia comes from Latin and means lack of reason.

AD is the most frequent type of dementia among all known types, affecting 70% of all diagnosed patients (SWAFFE, 2016). Figure 1.1 shows the most common types and their predominance across the spectrum of dementia.



Source: Adapted from (SWAFFE, 2016).

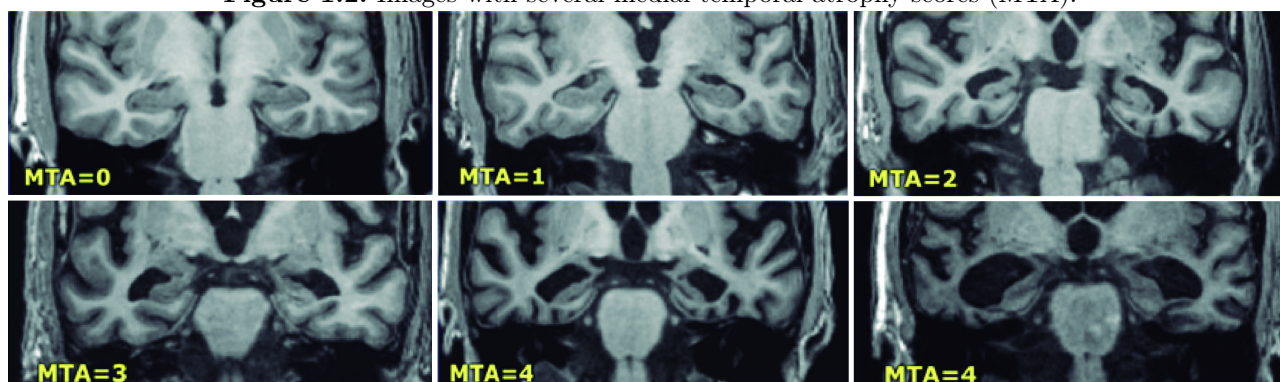
There are more than 100 different types of known dementia that, due to the prevalence of AD, are popularly and collectively labeled as Alzheimer's. The disease has become shorthand for the disorientation, memory loss and slow responses that can arise in the elderly. Other types of dementia include:

1. Vascular dementia: this can appear after a stroke that compromises the function of an important area of the brain. It can also appear because of long-term, untreated high blood pressure and diabetes that can cause small ischemias that progressively affect the brain.
2. Lewy body dementia: characterized by symptoms similar to those of Parkinson's disease such as tremors, muscle rigidity and slow movement. Symptom severity tends to fluctuate daily and may include hallucinations and memory loss.
3. Frontotemporal dementia: affects the frontal and temporal lobes of the brain that are responsible for regulating mood and behavior. Patients readily fluctuate between states of apathy and excitement.
4. Parkinson's dementia: toxic proteins are deposited in areas responsible for movement. Lewy body and Parkinson's dementia can be difficult to distinguish.

AD causes irrecoverable and real loss of nerve cells (neurons) and brain mass. The process appears earlier and progresses faster than the milder loss that is considered a normal part of natural aging. In AD, the brain begins to attack itself, the reasons for which are still unknown. The hippocampus, which is responsible for recording new information, is the most affected part of the brain. For this reason, at early stages of the disease, AD patients retain memories from years ago but may forget what happened yesterday.

Studies on changes in hippocampal structures show that changes in neuronal plasticity, myelination and interneuronal connectivity are important (BALESTRIERI J., 2020). The medial temporal atrophy (MTA) score is used to assess these structures in neuroimaging. Figure 1.2 shows brain slices with MTA ranging from 0 to 4. An MTA score above 3 is considered abnormal in patients aged 75 years or older.

Figure 1.2. Images with several medial temporal atrophy scores (MTA).



Source: Adapted from (BALESTRIERI J., 2020).

MTA SCORE

The score is calculated from a visual analysis of the thickness of the choroidal fissure, thickness of the temporal horn and height of the hippocampal formation in the coronal section of the T1 sequence of cranial magnetic resonance images.

- MTA 0: no atrophy.
- MTA 1: widening of the choroidal fissure.
- MTA 2: widening of the choroidal fissure and temporal horn of the lateral ventricle.
- MTA 3: moderate loss of hippocampal volume (reduced hippocampal height).
- MTA 4: severe loss of hippocampal volume.

1.1.1 Stages

Dementia and AD are frequently ascribed stages that describe disease progression. Defining stages help clinicians determine appropriate treatments and facilitates communication between physicians and caregivers.

Several scales are used to classify dementia and better understand the different stages of AD. These scales are based on how well a person thinks (cognitive decline) and functions (physical abilities).

Examples of scales include:

1. Global Deterioration Scale / Reisberg Scale (GDS) - GDS uses seven stages that are based on levels of cognitive decline. The test is most relevant for people with AD since not all types of dementia include memory loss.
2. Functional Assessment Staging Test (FAST) - Like the GDS scale, FAST is a seven-stage system based more on the level of functional ability and the ability to perform daily living activities than on cognitive decline.
3. Clinical Dementia Rating (CDR) - CDR uses a five-point system based on cognitive (thinking) skills and functional capacity. Stages are assigned by interviewing the subject and others, including family members, to assess six areas: memory, orientation, judgment and problem solving, community affairs, home and hobbies, and personal care (HUGHES C. P., 1982).

Table 1.1 shows the characteristics and expected average duration of each stage of the CDR scale.

Some studies have used other stage classifications for AD. For example, the Australian Imaging, Biomarker & Lifestyle Flagship Study of Aging (AIBL) (ELLIS *et al.*, 2009) is the result of an Australian study with hundreds of patients that diagnoses disease progression in three stages: cognitively normal (CN), mild cognitive impairment (MCI) and probable Alzheimer’s disease (AD). Another system is the Alzheimer’s Disease Neuroimaging Initiative (ADNI) (MARINESCU *et al.*, 2018), a North American study, which also classifies disease progression into three stages: CN, MCI and AD. Data from both AIBL and ADNI were used in the present study. Another study, the Open Access Series of Imaging Studies (OASIS) (MARCUS *et al.*, 2007) uses the CDR scale.

Table 1.1. Clinical Dementia Rating (CDR) Scale

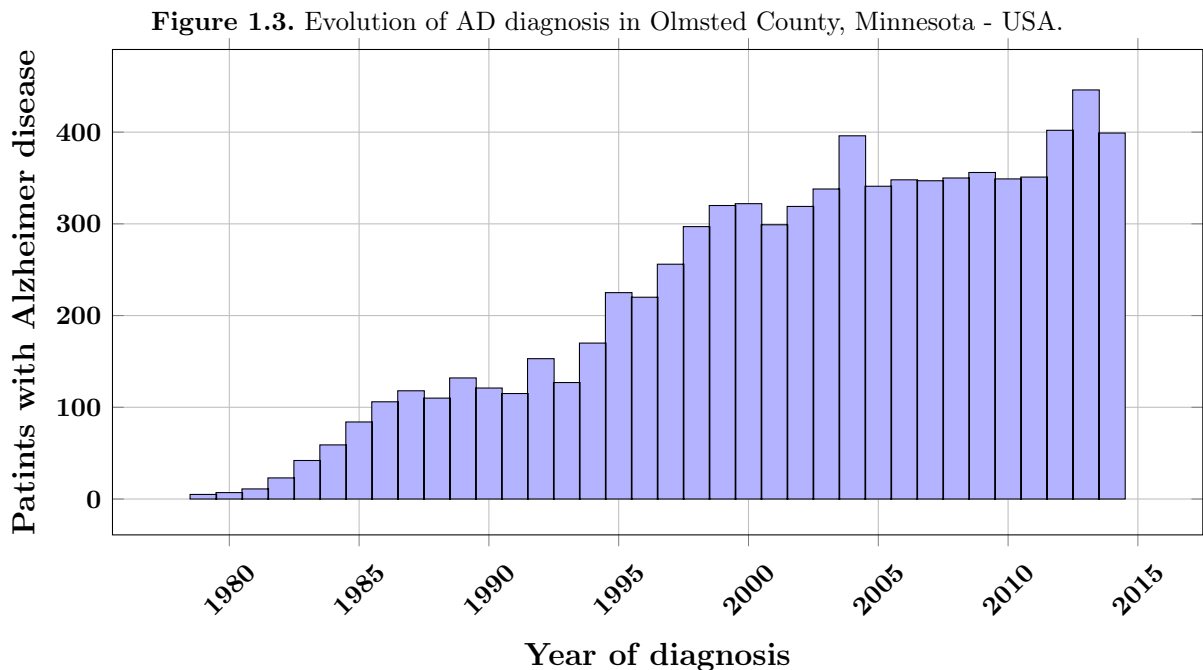
Stage	Description	Expected duration of stage
CDR-0	No Dementia	N/A
CDR-0.5	Very Mild Dementia – Memory problems are slight, but consistent. - Some difficulty with time and problem-solving - Daily life is slightly impaired - Individuals can perform personal care activities	Average duration is a few years up to 7 years.
CDR-1	Mild Dementia – Memory loss is moderate, especially for recent events, and interferes with daily activities. - Moderate difficulty with solving problems - Cannot function independently in community affairs - Difficulty with daily activities and hobbies, especially complex ones	Average duration is 2 years.
CDR-2	Moderate Dementia – More profound memory loss, only retaining highly learned material. - Disorientation with respect to time and place - Impaired judgment, with difficulty handling problems - Little to no independent function at home - Can only do simple chores - Fewer interests	Average duration is just under 2 years to 4 years.
CDR-3	Severe Dementia – Severe memory loss. - Disorientation with respect to time or place - No judgment or problem-solving abilities - Cannot participate in community affairs outside home - Requires help with all tasks of daily living - Requires help with most personal care - Frequent incontinence	Average duration is 1 year to 2.5 years.

Source: Adapted from (CENTRAL, 2022).

1.1.2 A short history

AD was first recognized in 1906 when the German physician Alois Alzheimer diagnosed dementia in a 55-year-old patient after performing a brain biopsy and relating the brain sequelae to the patient’s medical records and symptoms. Alois Alzheimer announced these results at the 37th Conference of Psychiatry in Tübingen, Germany (YANG H. D., 2016). The term AD was later named by Emil Kraepelin in his 1910 book on psychiatry. Despite the disease description comes from the beginning of the 20th century, it became a social stigma, marked by the prejudice and lack of understanding of the distant past (YANG H. D., 2016).

It was only in the second half of the 1970s that AD was officially recognized as a disease after receiving an international disease identification code (ICD - International Classification of Disease). Data from the Rochester Epidemiology Project was used to identify increasing numbers of AD diagnoses in patients over 65 in Olmsted County, Minnesota, United States, from 1980 to 2014 (KNOPMAN D. S., 2019). This data is shown in Figure 1.3. Here, the horizontal axis represents the year and the vertical axis the number of unique people diagnosed that year with AD. The data reveal a clear increase in the number of cases, probably due to a better understanding of the disease by medical professionals, or an increase in cases resulting from an aging population, improved diagnostic tools, and other factors.



Source: Adapted from (KNOPMAN D. S., 2019).

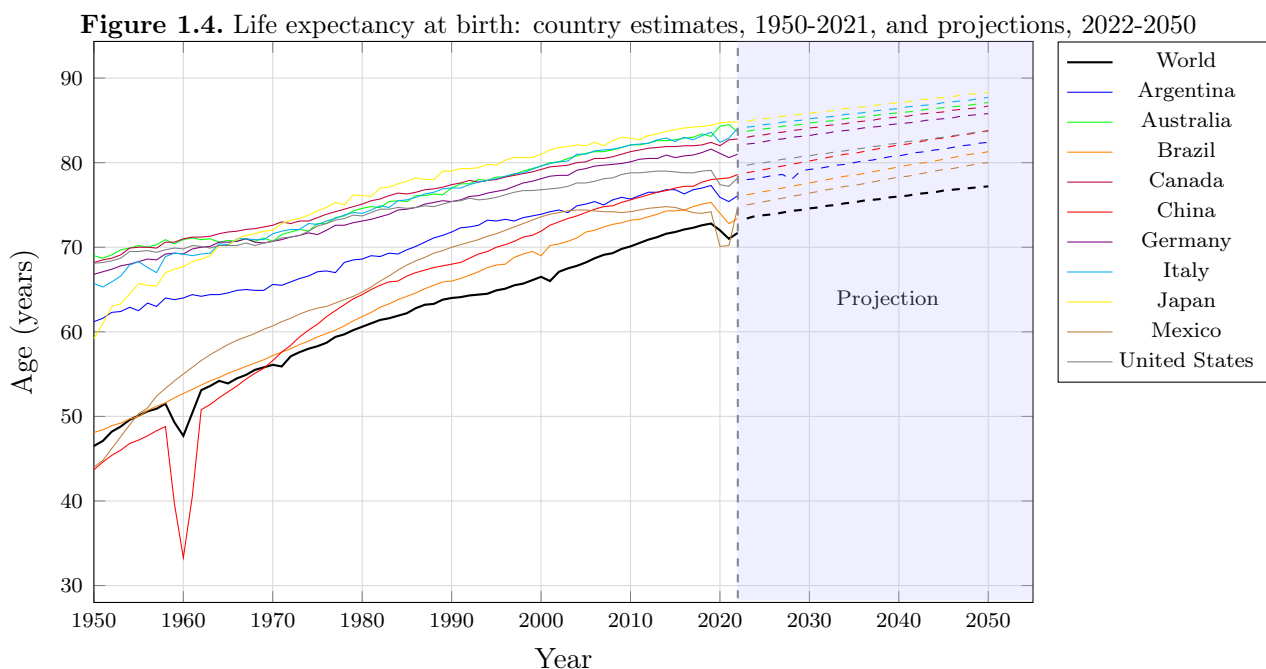
Since AD is mainly associated with cognitive aging, the disease should become even more prevalent with the progressive aging of the world population. Significant increases in life expectancy across the planet have made AD an important public health problem with rising medical care costs. In 2020, the annual cost of treatment in the United States was estimated at U\$ 50 billion (SCHACHTER MD; KENNETH L. DAVIS, 2000).

The 2022 World Population Prospects report from the United Nations (UN) estimated that life expectancy at birth will increase on all continents over the next 30 years. By 2050, average worldwide life expectancy is projected to reach 77.2 years, with a difference of 31.8

years between countries with the lowest and highest values.

As countries with lower rates complete the demographic transition and reach historically low levels of mortality, further reductions in mortality rates will become more difficult to achieve and increases in life expectancy will stabilize, leading to long-term convergence among countries and regions (ONU, 2022).

Figure 1.4 is based on UN data (ONU, 2022) and shows Australia and Japan with the highest life expectancies. At the other extreme, the lowest rates (not shown in Figure 1.4) were found in sub-Saharan Africa.



Source: Adapted from (ONU, 2022).

1.1.3 Diagnosis

Currently, AD diagnosis depends on a variety of tools, of professionals such as neurologists, neuropsychologists, geriatricians and geriatric psychiatrists (REPORT, 2021), and different tests, including medical history, behavioral change assessment, cognitive testing, physical and neurological examinations, blood tests, imaging tests, and determination of beta-amyloid levels.

Clinical anamnesis (interview with the patient by the health professional), neurological examinations and complementary tests are extremely important in diagnosing cognitive

impairment and specific types of dementia. Advances in biomarkers and new diagnostic techniques have improved the accuracy of assessments; however, their use is still restricted to specific cases and to within the scope of scientific research (PARMERA; NITRINI, 2015).

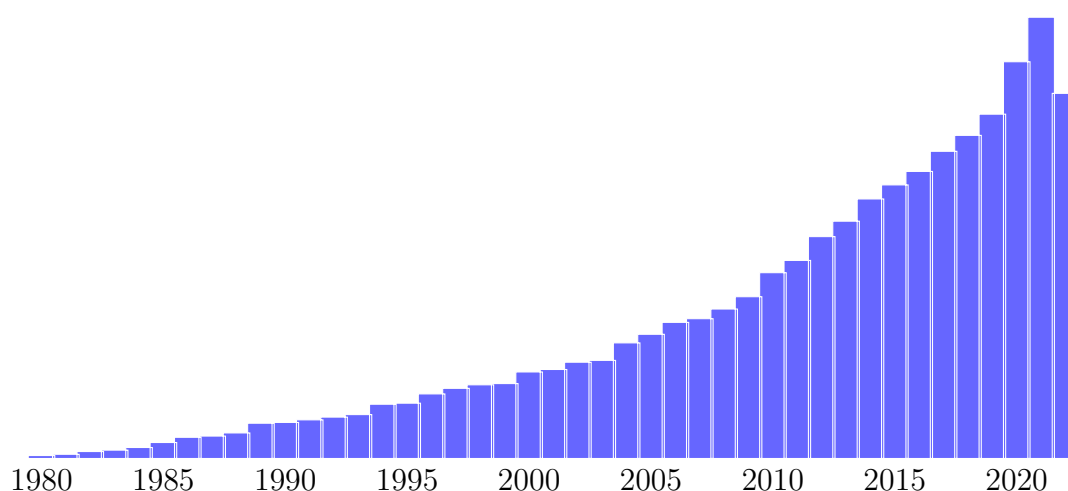
The next step after satisfactory anamnesis is an in-depth general and cognitive neurological examination. The American Academy of Neurology recommends testing vitamin B12, thyroid hormones, creatinine, and blood counts. Some types of structural imaging such as cranial tomography or cranial resonance are also recommended (PARMERA; NITRINI, 2015).

Nevertheless, diagnosing AD is a challenging task that could benefit from other diagnostic strategies. Since AD is multifactorial disease with causes that are difficult to identify, and one that can be confused with other types of dementia, precise diagnosis is still an open field for science.

Diagnosing AD at an early stage, when cognitive impairment is mild, is also important so that treatment can delay disease progression or reduce its financial impacts.

AD related research grows every year. A search on PubMed¹ with the word “Alzheimer’s” showed 200 thousand research articles registered from 1980 to September 2022. Figure 1.5 shows the distribution of scientific articles over the years and demonstrates an evident increase in the scientific community’s interest in AD.

Figure 1.5. PubMed portal search results with the word "Alzheimer's" from 1980 to September 2022.



Source: Own authorship.

¹<https://pubmed.ncbi.nlm.nih.gov/>

1.1.4 Alzheimer's in Brazil

According to the website² of the Brazilian Institute of Geography and Statistics (IBGE), the population of Brazil will reach nearly 233 million inhabitants by 2050. Roughly 35% of these will be 65 years or older, which means approximately 81 million people.

The website³ of the Brazilian Society of Geriatrics and Gerontology (SBGG) estimates that 6.8% of the population over 60 years of age has a chance of developing some type of dementia, 40% to 60% of which would be Alzheimer's.

Thus, approximately 2.2 to 3.3 million Brazilians may develop AD in 2050.

1.1.5 Alzheimer's and Covid-19

A recent study (TAHIRA *et al.*, 2021) at the Butantã Institute concluded that AD is a risk factor for those who contract Covid-19, regardless of age. According to the survey, AD did not increase the risk of hospitalization relative to other risk factors. However, when a patient was hospitalized and already had the disease, the risk of developing a more serious condition due to SARS-CoV-2 was three times greater compared to those who did not have dementia, and six times greater for patients over 80 years old.

1.2 RATIONALE FOR THE PRESENT STUDY

A 2019 report by Alzheimer's Disease International estimated that the global number of people with dementia will reach 78 million by 2030 and 139 million by 2050.

According to a 2021 report by Alzheimer's Disease International (Gauthier S. *et al.*, 2021), 75% of dementia cases go undiagnosed and 46% of these and their caregivers identified fear and stigma as barriers to diagnosis.

The 2019 Alzheimer's Disease International report also reported that the total costs of dementia in 2015 were US\$ 818 billion dollars, US\$ 1 trillion dollars in 2018 and estimated at US\$ 2 trillion dollars in 2030.

²<https://www.ibge.gov.br/apps/populacao/projecao/>

³<https://sbgg.org.br/>

Thus, there is ample justification for improving AD diagnosis by building tools that can assist and complement the efforts of healthcare professionals.

1.3 OBJECTIVES

The general objective of this thesis is to investigate machine learning strategies that use clinical data and magnetic resonance images to contribute to AD diagnosis. The primary data used in the study were obtained from scientific studies carried out in the United States and Australia.

The following objectives were proposed to achieve the general goal of this study:

1. Acquire licenses for reliable clinical and magnetic resonance imaging databases.
2. Develop a framework for reading and processing databases and implementing machine learning techniques.
3. Perform exploratory experiments on clinical and magnetic resonance imaging databases.

SCENARIO 1: Clinical data

4. Investigate and propose a feature extraction technique that fits the data based on clinical diagnoses.
5. Define and carry out experiments on clinical data using the feature selection method.

SCENARIO 2: Medical images

6. Evaluate images in the sagittal, coronal and axial planes.
7. Propose an automatic image (slice) selection algorithm for MRI.
8. Develop strategies for building image datasets that best represent the challenges of diagnosing AD.

9. Set up and carry out classification experiments using convolutional networks.
10. Analyze and generate results using the proposed methods.
11. Discuss results.

1.4 THESIS ORGANIZATION

This thesis has been organized into six chapters and one appendix:

- Chapter [1](#): INTRODUCTION. Presents an introduction, a review of Alzheimer’s disease, and contextualization, justification, and objectives for the study.
- Chapter [2](#): FOUNDATIONS. Reviews concepts of artificial intelligence, machine learning and deep learning. Presents the challenges of diagnosing Alzheimer’s disease, and a review of literature and state of the art practices.
- Chapter [3](#): CLINICAL DATA. Presents some of the technological basis for the treatment of clinical data used in Alzheimer’s diagnosis. Describes the databases and the pre-processing used for building the final datasets. Provides context for the challenges of feature selection and presents the proposed algorithm. Presents performance evaluation metrics for the trained models, the cross-validation methodology, and the machine learning models used.
- Chapter [4](#): IMAGES. Presents some of the technological basis for the treatment of magnetic resonance images used in Alzheimer’s diagnosis. Provides an evaluation of the axial plane and proposes an automated image selection algorithm. Provides an evaluation of the coronal plane and proposes a method for augmenting input data by incorporating neighboring slice images. Shows how the image datasets are generated and prepared for processing by machine learning algorithms.
- Chapter [5](#): RESULTS. Presents the reduced clinical datasets generated by the new feature selection algorithm. Describes the image datasets generated using a strategy to augment input data and enable evaluation via convolutional networks. Presents an evaluation of the models, an explanation of results and a comparison with other state-of-the-art results.

-
- Chapter [6](#): CONCLUSION. Presents an overall evaluation of the contributions of this study and provides suggestions for future studies.
 - Appendix [A](#): Portuguese thesis.

CHAPTER 2

THEORETICAL FOUNDATION

This chapter presents theoretical foundations, a review of research and discussions by other authors on improving AD diagnosis via computational techniques, with emphasis on artificial intelligence and state-of-the-art practices.

2.1 ARTIFICIAL INTELLIGENCE IN DIAGNOSING NEURODEGENERATIVE DISEASES

Artificial intelligence (AI) is a thriving field with many practical applications and avenues for research. There is a relentless search for intelligent software systems that can automate routine processes, interpret speech or images, provide medical diagnoses and support basic scientific research. Early on, the AI field quickly tackled and solved problems that were intellectually difficult for humans but relatively simple for computers—problems that can be described by a list of formal, mathematical rules. The real challenge for AI is solving tasks that are easy to perform but difficult to formally describe - problems that humans solve intuitively, almost automatically, such as recognizing spoken words or faces in images (GOODFELLOW *et al.*, 2016).

AI systems need the capacity to acquire their own knowledge by extracting patterns from raw data. This ability is known as machine learning (ML). The advent of ML allowed computers to tackle problems that involve understanding the real world and making decisions that appear subjective.

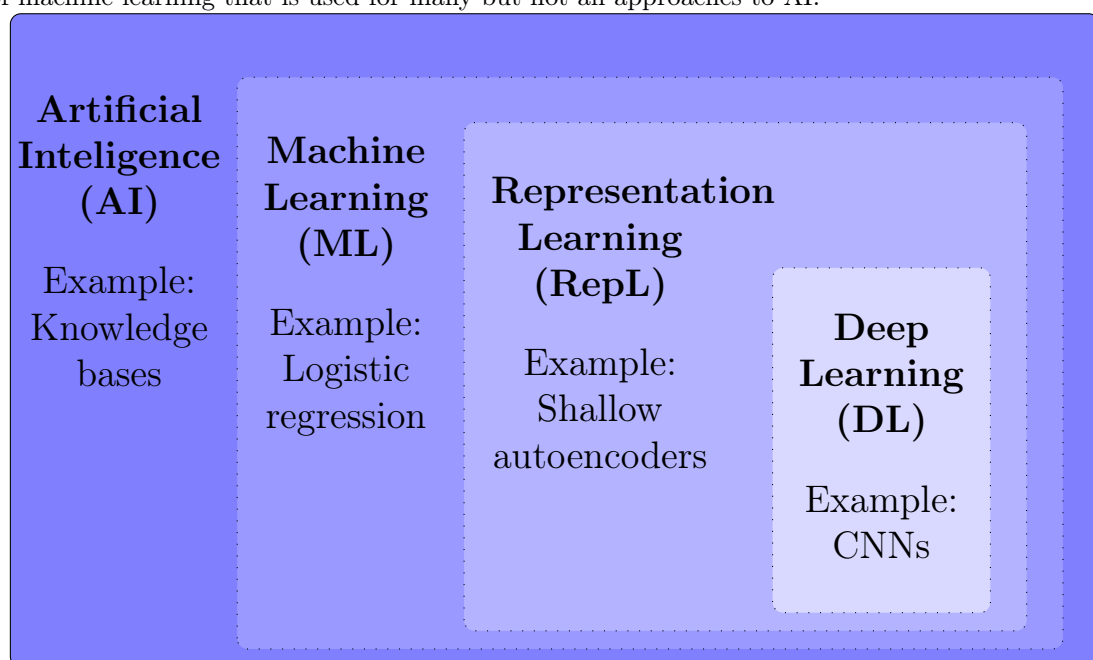
The challenge of more complex problems is deciding which features to extract. One solution to this problem is to use ML to discover not only the mapping from representation to output, but also the representation itself. This approach is known as Representation Learning (RepL). RepL frequently performs better than manually extracted representations. RepL also allows AI

systems to quickly adapt to new tasks with minimal human intervention. A RepL algorithm can discover a set of features for a simple task in minutes, or a complex task in hours. Conversely, manual representation of features for a complex task requires significant human time and effort, which can take decades for an entire community of researchers (GOODFELLOW *et al.*, 2016).

A major challenge for many real-world AI applications is that too many factors influence each piece of data. Many of these variance factors can only be identified using a sophisticated, quasi-human understanding of the data. In these cases, RepL is rather ineffective. Deep learning (DL) solves this problem by introducing representations that express themselves in terms of other, simpler representations.

Figure 2.1 shows how these AI techniques are related and presents an example algorithm for each case.

Figure 2.1. A Venn diagram showing how deep learning is a kind of representation learning, which is in turn a kind of machine learning that is used for many but not all approaches to AI.



Source: Own authorship.

For some time, attempts have been made to develop computerized systems that assist clinical diagnosis. More than 50 years ago, one such system was developed to evaluate patient data and suggest steps to reestablish hydroelectrolyte balance (BLEICH, 1969).

AI can assist the investigation of neurodegenerative disorders more deeply, providing a comprehensive overview of the disease and creating avenues for the application of precision

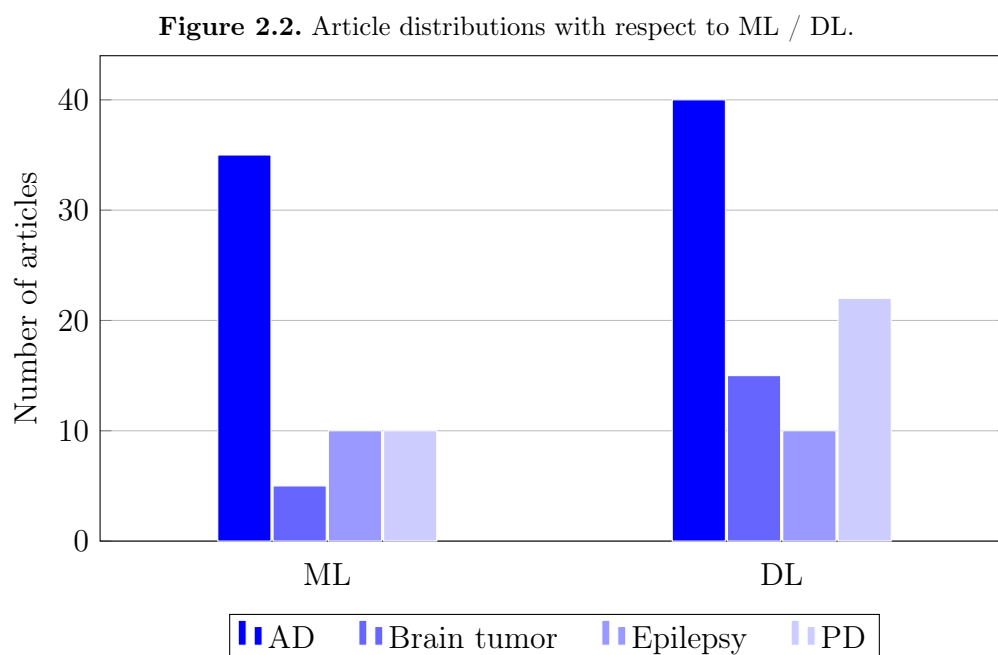
medicine. Thus, AI is a powerful resource that can help diagnose neurodegenerative diseases by revolutionizing the medical environment and providing greater certainty and practicality in the management of these pathologies (BRITO *et al.*, 2021).

2.2 COMPUTATIONAL SUPPORT IN DIAGNOSING ALZHEIMER’S DISEASE

ML and DL techniques have been important allies in aiding AD diagnosis (NIYAS; P, 2021a; NIYAS; P, 2021b; NGUYEN M., 2020; ALBRIGHT, 2019; IDDI *et al.*, 2019; MOORE *et al.*, 2019; GHAZI *et al.*, 2019; NGUYEN *et al.*, 2018; ZHANG *et al.*, 2011).

In general, the use of artificial intelligence techniques to improve the accuracy of diagnosing brain diseases has grown significantly in recent years (KHAN *et al.*, 2021). Several recent studies have used ML or DL to diagnose AD and other brain diseases, such as brain tumors, epilepsy and Parkinson’s disease. (KHAN *et al.*, 2021) searched publications on IEEE Xplore, ScienceDirect and Google Scholar between 2018 and 2020, and found 75 articles on AD, 46 using DL and the remaining 35 using other ML techniques.

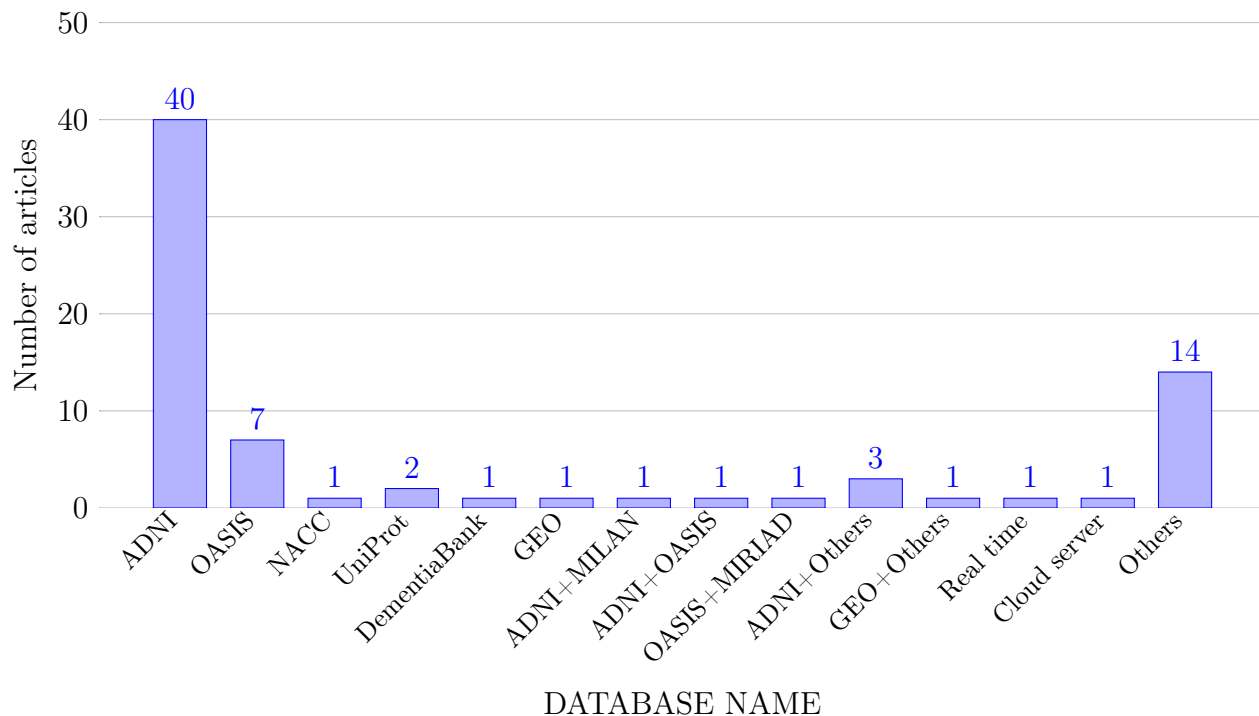
Figure 2.2 shows the number of articles using ML or DL to diagnose AD, brain tumors, epilepsy and Parkinson’s disease.



Source: Adapted from (KHAN *et al.*, 2021).

(KHAN *et al.*, 2021) also determined which databases were used in each of these studies. Figure 2.3 shows that 60% of the studies employed data from the ADNI database, which is also used in the current study, in addition to AIBL and OASIS.

Figure 2.3. Databases used in studies on AD diagnosis.



Source: Adapted from (KHAN *et al.*, 2021).

Several studies have used the AIBL and/or ADNI databases with ML techniques, (NIYAS; P, 2021a), (NIYAS; P, 2021b), (IDDI *et al.*, 2019), (MOORE *et al.*, 2019), (ZHANG *et al.*, 2011) or DL techniques (NGUYEN M., 2020), (ALBRIGHT, 2019), (GHAZI *et al.*, 2019), (NGUYEN *et al.*, 2018), (LIU J LI, 2020), (HAN *et al.*, 2020), (LI *et al.*, 2022), (RAZZAK *et al.*, 2022), (HAZARIKA *et al.*, 2022), (ZHANG *et al.*, 2023), (WU *et al.*, 2022), (GAO *et al.*, 2023a), (GAO *et al.*, 2023b), (SHARMA *et al.*, 2022). The results of these recent state-of-the-art studies will be compared to the results of the present study in Chapter 5 (Results and discussion).

2.3 LITERATURE REVIEW

As shown in Figure 1.5, the number of scientific articles on AD has increased significantly from 1980 to 2022, demonstrating the scientific community's growing interest in this topic.

(KHAN *et al.*, 2021) found 75 articles published from 2018 to 2020 on AD diagnosis that used computational intelligence techniques. Of these, 40 employed DL while the remaining 35 studies used ML techniques. Some of these articles, and others published before 2022, are reviewed below.

The authors of (NIYAS; P, 2021a) compared the performance of state-of-the-art classification algorithms using Dynamic Ensemble Selection (DES) for diagnoses in three categories: CN, MCI, and AD. In this study, a set of the most common ML classifiers was used as DES input. The classification results were compared with and without the use of DES. The authors concluded that DES improved the classification evaluation metrics. The study used the ADNI database.

In (NIYAS; P, 2021b), the authors used the Fischer score to select features in the ADNI and AIBL databases and analyzed the classification evaluation metrics for the CN, MCI and AD classes using the support vector machine (SVM) and k-nearest neighbors (kNN) classifiers. The authors concluded that this feature selection method was efficient and improved the classification results. The classification metrics for the CN v/s AD, MCI v/s AD and MCI v/s CN classes were also evaluated. This was done using balanced classification accuracy (BCA) given the imbalance of the classes in the datasets.

(IDDI *et al.*, 2019) proposed a two-phase approach for modeling and predicting measures of cognition, function, brain imaging, fluid biomarkers and diagnosis of individuals using multiple domains simultaneously. In the first phase, multivariate mixed effects models were used to simultaneously model multiple markers over time. In the second phase, random forests (RF) were used to predict CN, MCI and AD diagnoses from continuous markers based on the first-phase model. The ADNI database was used in this study.

AD prediction based on time series was proposed in (MOORE *et al.*, 2019). Here, RF was used to discover the relationship between pairs of data points within different time intervals. The input vector was composed of a summary of historical time series and included both demographic and non-demographic variables, such as genetic data. The database used was ADNI. The results showed that the method was effective and comparable to other methods.

A technique for predicting the progression of Alzheimer’s disease using recurrent neural networks (RNN) was proposed in (NGUYEN M., 2020). In this study, multimodal AD markers

and a patient's clinical diagnosis from one or more points in time were used to predict clinical diagnosis, cognition, and ventricular volume diagnosis for each month in the future. The Alzheimer's Disease Prediction of Longitudinal Evolution (TADPOLE) challenge database from ADNI was used. The authors stated that their approach ranked second in the TADPOLE challenge as of June 3, 2020.

(ALBRIGHT, 2019) proposed a prediction method for Alzheimer's disease progression using neural networks and a new pre-processing algorithm called "All-Pairs", which involved comparing all possible pairs of temporal data points for each patient. According to the authors, the trained neural network could be used to identify patients in early stages of AD, which are good candidates for clinical trials of AD therapies. This study also used the ADNI database.

(GHAZI *et al.*, 2019) proposed a long-short-term memory (LSTM) algorithm that was used to model the progression of Alzheimer's disease (AD) using six imaging biomarkers (ventricular volumes, hippocampus, whole brain, fusiform, middle temporal gyrus, and entorhinal cortex) from the MRI database. The ADNI database was used. The study concluded that the integrated treatment of missing values in the formation of an LSTM network improved the application of RNNs in neurodegenerative modeling of disease progression in longitudinal sections.

(NGUYEN *et al.*, 2018) also modeled Alzheimer's disease progression using RNN. Here, the ADNI database was used to propose strategies to train RNN via data that used RNN to fill in incomplete data or missing points on a timeline.

(LIU J LI, 2020) also used the ADNI database and examined the problem of predicting disease progression by measuring cognitive scores and selecting predictive biomarkers of progression. Most previous work on predicting AD progression ignored the issue of missing data. Missing data represents a major challenge to modeling longitudinal data, since most statistical models assume complete data sets.

New diagnostic models that use magnetic resonance imaging to identify the various stages of AD were presented in (HAN *et al.*, 2020). Here, experiments on magnetic resonance images collected from the ADNI website were used to evaluate the proposed models.

(LI *et al.*, 2022) used a 3D convolutional neural network to process entire 3D magnetic resonance images, using data from ADNI and magnetic resonance imaging.

(RAZZAK *et al.*, 2022) proposed an integrative framework based on deep learning to improve the predictive performance of Alzheimer’s diagnosis. The study used ADNI data and magnetic resonance imaging.

(HAZARIKA *et al.*, 2022) presented a new model based on convolutional networks to distinguish patients with AD and MCI from cognitively normal individuals. The new model was based on VGG-19 and DenseNet and used data from ADNI.

An accuracy of 93.4% for AD v/s CN diagnosis using AIBL data was achieved by (ZHANG *et al.*, 2023), who presented a three-dimensional multilayer perceptron mixer DL model.

A new attention-based 3D multi-scale CNN model was proposed in (WU *et al.*, 2022), achieving 91.3% accuracy on ADNI data.

(GAO *et al.*, 2023a) presented a hybrid multi-scale attention convolution network and an aging transform network for AD diagnosis, with sMRI data. The study reports 90.5% accuracy in diagnosing AD v/s CN on ADNI data.

In this article (GAO *et al.*, 2023b), the authors proposed a generative brain state transfer adversarial network (BrainStatTrans-GAN) for AD diagnosis. The paper reports 94.9% accuracy for AD v/s CN diagnosis in ADNI data.

A new approach to structural and metabolic image fusion based on wavelet packet transform using MRI and PET scans is proposed in (SHARMA *et al.*, 2022). The paper reports accuracy of 97.33% for AD v/s CN diagnosis for ADNI data.

Chapter 5 will compare some of these studies to the results of the present study.

This chapter presented theoretical foundations and reviewed research by other authors who also used computational techniques to aid in AD diagnosis. The following chapter will present technological proposals for developing computational techniques that use clinical data to aid in the diagnosis of AD.

ANALYSIS BASED ON CLINICAL DATA

This chapter details the clinical data used and the technological proposals developed.

A patient clinical data includes data from anamnesis, laboratory tests, weight, age, images, heart rate, blood pressure, medications, clinical history, diagnoses and others. The relative importance of these data sources depends on the diagnosis type.

In the present study, the AD clinical data were extracted from the studies detailed below.

3.1 DATABASE DESCRIPTIONS

We selected the ADNI, OASIS, and AIBL databases from the set of databases related to AD (Figure 2.3). ADNI and OASIS were among the most used databases in scientific articles from 2018 to 2020 (KHAN *et al.*, 2021). AIBL has also been recently referenced in (NIYAS; P, 2021b).

3.1.1 OASIS

The Open Access Series of Imaging Studies (OASIS) is a project that aims to make brain neuroimaging datasets freely available to the scientific community. The clinical database can be found on the OASIS¹ website (MARCUS *et al.*, 2007). This database consists of a cross-sectional collection of 416 patients aged from 18 to 96 years. For each patient, three or four individual T1-weighted MRIs obtained in single scan sessions are included. The subjects are all right-handed and include both men and women. Patients are diagnosed according to the CDR assessment scale (HUGHES C. P., 1982). The database is available in the comma-separated values (CSV) format.

¹<https://www.oasis-brains.org>

Table 3.1 lists the features in the OASIS1 database and their descriptions.

Table 3.1. OASIS: Features descriptions

Feature	Description
Age	Age at time of image acquisition (years)
Sex	Sex (male or female)
Education	Years of education
Socioeconomic status	Assessed by the Hollingshead Index of Social Position and classified into categories from 1 (highest status) to 5 (lowest status)
MMSE score	Ranges from 0 (worst) to 30 (best)
CDR scale	0 = no dementia, 0.5 = very mild AD, 1 = mild AD, 2 = moderate AD
Atlas scaling factor	Computed scaling factor (unitless) that transforms native-space brain and skull to the atlas target (i.e., the determinant of the transform matrix)
eTIV	Estimated total intracranial volume (cm^3)
nWBV	Expressed as the percent of all voxels in the atlas-masked image that are labeled as gray or white matter by the automated tissue segmentation process

Source: Own authorship.

3.1.2 AIBL

The Australian Imaging, Biomarker & Lifestyle Flagship Study of Ageing (AIBL) (ELLIS *et al.*, 2009) is a study to discover which biomarkers, cognitive characteristics, and health and lifestyle factors determine the subsequent development of symptomatic AD. This Australian study started in 2006 with more than 1000 patients diagnosed in three stages of disease progression: cognitively normal (CN), mild cognitive impairment (MCI) and established Alzheimer’s disease (AD). Further information can be found on the study’s website².

Table 3.2 lists the features in the AIBL database and their descriptions.

²<https://aibl.csiro.au/>

Table 3.2. AIBL: Features descriptions

Feature	Description
HMT3	Red Blood Cell Count
HMT7	White Blood Cell Count
HMT13	Platelets
HMT40	Hemoglobin
HMT100	Mean Cell Hemoglobin
HMT102	Mean Cell Hemoglobin Concentrate
AXT117	Thyroid Stimulate Hormone
BAT126	Vitamin 12
RCT6	Urea Nitrogen
RCT11	Serum Glucose
RCT20	Cholestrol
RCT392	Creatinine
LIMMTOTAL	Total number of story units recalled-Logical Memory Immediate Recall
LDELTOTAL	Total number of story units recalled-Partial Score of Logical Memory test
MMSCORE	Mini Mental State Examination Score
CDGLOBAL	Clinical Dementia Rating Global
APGEN1	Apoe gene 1
APGEN2	Apoe gene 2
DXCURREN	Diagnosis status

Source: Own authorship.

3.1.3 ADNI

The Alzheimer’s Disease Neuroimaging Initiative (ADNI) (MARINESCU *et al.*, 2018) is a multicenter, longitudinal North American study designed to develop clinical, imaging, genetic, and biochemical markers for the early detection and screening of AD. Since its launch over a decade ago, this historic public-private partnership has made major contributions to AD research and enabled data sharing among researchers around the world. Additional information can be found on the study’s website³. This study also classifies the stage of disease progression into three categories: CN, MCI, and AD.

Table 3.3 lists the features in the ADNI database and their descriptions.

Data from the ADNI and AIBL studies can be downloaded from the Image and Data Archive (IDA)⁴ platform, which is managed by the Laboratory of Neuro Imaging (LONI) of the Mark and Mary Stevens Neuroimaging and Informatics Institute, associated with the University of

³<https://adni.loni.usc.edu/>

⁴<https://ida.loni.usc.edu/login.jsp>

Southern California.

Table 3.3. ADNI: Features descriptions

Feature	Description
Ventricles	Ventricles Volume
Hippocampus	Hippocampus Volume
CDRSB	Clinical Dementia Rating Scale Box
MMSE	Mini-Mental State Examination
WholeBrain	WholeBrain Volume
Entorhinal	Entorhinal Volume
Fusiform	Fusiform Volume
MidTemp	Med Temp Volume
ICV	Intracranial Volume
ADAS11	Alzheimer’s Disease Assessment Cognition Scale 11
ADAS13	Alzheimer’s Disease Assessment Cognition Scale 13
Age	Age at baseline
PTEDUCAT	Education
RAVLT_learning	Rey Auditory Verbal Learning Test
RAVLT_immediate	Rey Auditory Verbal Learning Test (5 sum)
RAVLT_Forgetting	Rey Auditory Verbal Learning Test Forgetting
RAVLT_Perc_Forgetting_bl	Rey Auditory Verbal Learning Test Percentile Forgetting
AV45	Average AV45 SUVR of frontal, AC, precuneus, and parietal cortex relative to the cerebellum
APOE4	Apolipoprotein E4
FAQ	Functional Activities Questionnaire
FDG	Average FDG-PET of angular, temporal, and posterior cingulate
DX	Diagnosis Status

Source: Own authorship.

Further details on the features in each of these databases can be seen in the reference articles ([ELLIS *et al.*, 2009](#)), ([MARINESCU *et al.*, 2018](#)) and ([MARCUS *et al.*, 2007](#)).

3.1.4 Preprocessing

Preprocessing is an essential step in preparing ML data and involves transforming raw data into a more suitable and effective format for building ML models. Preprocessing improves the quality and efficiency of the learning process and can have a significant impact on the performance and accuracy of the resulting model.

Eighteen features were chosen from AIBL, sixteen features used in ([NIYAS; P, 2021b](#)) and two genetic features, and assigned to three classes (AD, MCI and CN). Patient data contained

neuropsychology evaluations, blood tests and genetic data from each patient evaluation. The resulting dataset is referred to as AIBL-18 (Table 3.4).

Table 3.4. AIBL-18: dataset composition

Quantity	Features
18	CDGGLOBAL AXT117 BAT126 HMT3 HMT7 HMT13 HMT40 HMT100 HMT102 RCT6 RCT11 RCT20 RCT392 MMSCORE LIMMTOTAL LDELTOTAL APGEN1 APGEN2

Source: Own authorship.

Twenty-one features were chosen from the ADNI-TADPOLE study, features used in (NIYAS; P, 2021b), and the data was assigned to three classes (AD, MCI, and CN) for each patient evaluation. The resulting dataset is referred to as ADNI-21 (Table 3.5).

Table 3.5. ADNI-21: dataset composition

Quantity	Features
21	Ventricles Hippocampus WholeBrain Entorhinal Fusiform MidTemp ICV CDRSB MMSE RAVLT_learning RAVLT_immediate RAVLT_forgetting RAVLT_perc_forgetting_bl ADAS11 ADAS13 FAQ FDG AV45 APOE4 AGE PTEDUCAT

Source: Own authorship.

Eight features were chosen from the OASIS study, features used in (PAIVA; ESCOVEDO, 2021), and grouped into four classes: CDR0, CDR0.5, CDR1 and CDR2, for each patient assessment. This dataset is referred to as OASIS-8 (Table 3.6).

Table 3.6. OASIS-8: dataset composition

Quantity	Features
8	Age Sex Education Socioeconomic_status MMSE_score Atlas_scaling_factor eTIV nWBV

Source: Own authorship.

The AIBL-18 dataset initially contained 1639 instances. However, preprocessing was needed given that:

- Data was lacking in some instances. This may have occurred because the recording of an exam result was forgotten, or the exam was not carried out.
- Occasionally, there were up to six instances per patient that resulted from assessments made for the same patient on six different dates. Naturally, in these cases the diagnosis

may have changed from CN to MCI, MCI to AD or even AD to MCI. Patients with many evaluations tend to produce instances with greater statistical dependence.

These cases were treated using the following strategies:

1. Choose to not impute the data, considering that the number of instances is sufficient for classification, while completely excluding all instances that have at least one missing piece of data.
2. Select only one instance per patient, choosing the most recent record from those available. This strategy helps minimize dependence in the dataset.

The overall process yielded a dataset with 785 instances, 18 features and three classes for AIBL-18, representing 95.03% of the total number of patients (i.e., a significant use of the original data), without having to impute data or use data from more than one consultation of any patient.

The same preprocessing strategy was applied to ADNI-21, resulting in a dataset with 21 features, three classes and 1033 instances, or 59.47% of the total number of patients in the study.

The OASIS study originally had four classes (CDR 0, CDR 0.5, CDR 1 and CDR 2). However, CDR 1 and CDR 2 were grouped together into a new class called CDR 1-2 since the number of instances in the CDR 2 class was very small. This process yielded a dataset with 216 instances, three classes and seven features.

3.2 DATA ANALYSIS

The post-processed AIBL-18 dataset resulted in 785 instances, 18 features and three classes (Table 3.7).

Table 3.7. AIBL-18: instance distribution by class

Class	Number of instances
CN	547
MCI	125
AD	113

Source: Own authorship.

Similarly, the post-processed ADNI-21 dataset resulted in 1033 instances, 21 features and three classes (Table 3.8).

Table 3.8. ADNI-21: instance distribution by class

Class	Number of instances
CN	505
MCI	367
AD	161

Source: Own authorship.

The post-processed OASIS-7 dataset resulted in 217 instances, seven features and three classes (Table 3.9).

Table 3.9. OASIS-7: instance distribution by class

Class	Number of instances
CDR 0	133
CDR 0.5	57
CDR 1-2	26

Source: Own authorship.

It is expected that all datasets would show an unequal distribution of instances among classes given that the presence of the disease is less frequent than the diagnosis in the general population.

Nevertheless, classification models are affected when trained with datasets that have imbalanced classes. In this case, the trained classifier can succeed if it simply chooses the largest class for any given instance.

To prevent this, the classification error function can be weighted in order to balance the errors that occur in the minority classes (CUI *et al.*, 2019).

Another option is to oversample the minority classes, synthesizing new instances from existing instances. This technique is known as SMOTE (the Synthetic Minority Oversampling Technique) (CHAWLA *et al.*, 2002). One variation of SMOTE is ADASYN (the Adaptive Synthetic Sampling Method), which generates synthetic instances in the region of the feature space where the density of minority examples is low (HE *et al.*, 2008). These techniques were evaluated in the present study but did not produce results that would justify their incorporation.

It is essential that the classification model assumes that datasets are unbalanced, and that

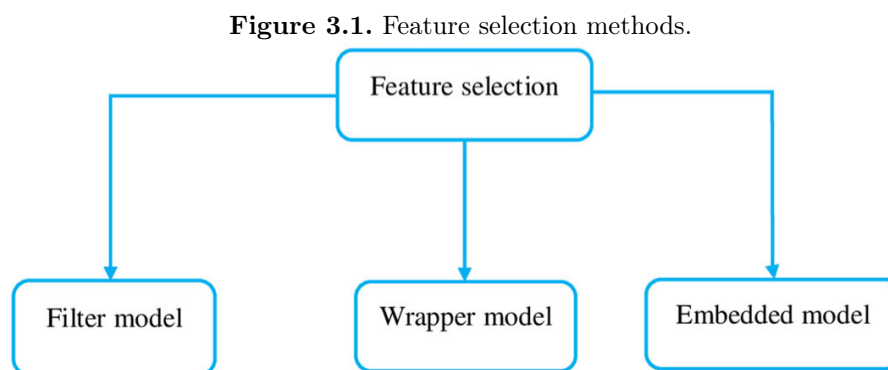
the evaluation metrics account for this.

3.3 FEATURE SELECTION

Feature selection is used to choose a subset of specific features for designing models. This process has many advantages, the most notable of which is improving a machine learning algorithm.

A second advantage is decreased computational complexity. If the number of features in a model is reduced, the calculation of model parameters becomes simpler. Another advantage of fewer features is a clearer understanding of how features relate to each other and contribute to results.

There is no straightforward and simple solution for choosing which features to include in an ML model. However, there are several techniques that can be used to effectively select features (DHAL; AZAD, 2022), which can be grouped as filter, wrapper and embedded methods, as shown in Figure 3.1.



Source: Adapted from (DHAL; AZAD, 2022).

3.3.1 Filter methods

Filter methods are feature selection methods that are executed as a preprocessing stage before running a model. They work by analyzing aspects of relationships between variables. Various indicators are used to decide which elements will be omitted and which will remain, depending on the model. Filter methods usually return feature rankings that show how features

are arranged relative to each other and exclude variables that are deemed obsolete.

A typical filter approach in linear regression is to remove features that are strongly correlated with each other. A variance threshold is another filtering approach that places a specific limit on the variation needed between features in order to be included in a model. The logic is that these variables will not provide significant fit if they do not have a high variance and therefore will not have much effect on the dependent variable.

Some examples of filter methods, used in the results of this work (Chapter 5), are Chi-square (JIN *et al.*, 2006), Mutual Information (HOQUE *et al.*, 2014), Anova F-value (DING *et al.*, 2014), Fisher Score (HART *et al.*, 2000) and MultiSURF (URBANOWICZ *et al.*, 2018).

3.3.2 Wrapper methods

Wrapper methods use various combinations of features to train models and then measure efficiency. These methods find an optimal subset of features by training models with a given subset, which is then evaluated on a test set. Wrapper methods are highly effective, but time consuming. Using them for large sets of features can be difficult and computationally expensive.

One type of wrapper method in linear regression is called recursive feature elimination. This method starts with all the features in a model and eliminates them one by one. Direct selection is the inverse of this process, starting with a single feature and adding others with the goal of improving model efficiency.

Some examples of wrapper methods, used in the results of this work (Chapter 5), are Recursive Feature Elimination (SEIJO-PARDO *et al.*, 2019), Permutation Importance (ALTMANN *et al.*, 2010), Shapley additive explanations (LUNDBERG; LEE, 2017) and Boruta (KURSA; RUDNICKI, 2010).

3.3.3 Embedded methods

Embedded methods are feature selection methods used in machine learning algorithm design. Regularization, particularly Lasso regularization, is the most popular embedded method as it can automatically reduce a feature set. Lasso regression is often referred to as L1 norm

regularization.

Some examples of embedded methods, used in the results of this work (Chapter 5), are Embedded Random Forest and Embedded LightGBM (SAEYS *et al.*, 2008).

3.3.4 The proposed algorithm

(NIYAS; P, 2021b) achieved promising results by using the Fischer score (HART *et al.*, 2000) to select features from the ADNI and AIBL databases and then evaluating the classification evaluation metrics for the CN, MCI, and AD classes.

Given a data set (\mathbf{x}, y) with n instances, d features and c classes, represented by $\{(\mathbf{x}_i, y_i)_{i=1}^n\}$, where $\mathbf{x}_i \in \mathbb{R}^d$ corresponds to the values of the features of instance i and $y_i \in \{1, 2, \dots, c\}$ corresponds to the label received by instance i , the feature selection problem is finding a subset of size $m < d$ that contains the maximum amount of representative information from the data set.

One possible solution to this problem is to use the Fisher score to select features from the data set, such that, within the data space delimited by the selected features, the distances between data points in different classes are as large as possible, while the distances between data points of the same class are as small as possible. Thus, this feature selection mechanism functions as a filter.

Considering μ_k^j and σ_k^j as the mean and standard deviation of the k^{th} class of the j^{th} feature, the Fisher score can be calculated as:

$$F(\mathbf{x}^j) = \frac{\sum_{k=1}^c n_k (\mu_k^j - \mu^j)^2}{(\sigma^j)^2}, \quad (3.1)$$

where $(\sigma^j)^2 = \sum_{k=1}^c n_k (\sigma_k^j)^2$, and μ^j and σ^j represent the mean and standard deviation of the entire dataset regarding the j^{th} feature, and n_k represents the size of the k^{th} class in the reduced data space.

After calculating the Fisher score for each feature, the m features with the highest Fisher score are selected. Given that the score for each feature is calculated independently, the features selected by the Fisher score criterion are suboptimal.

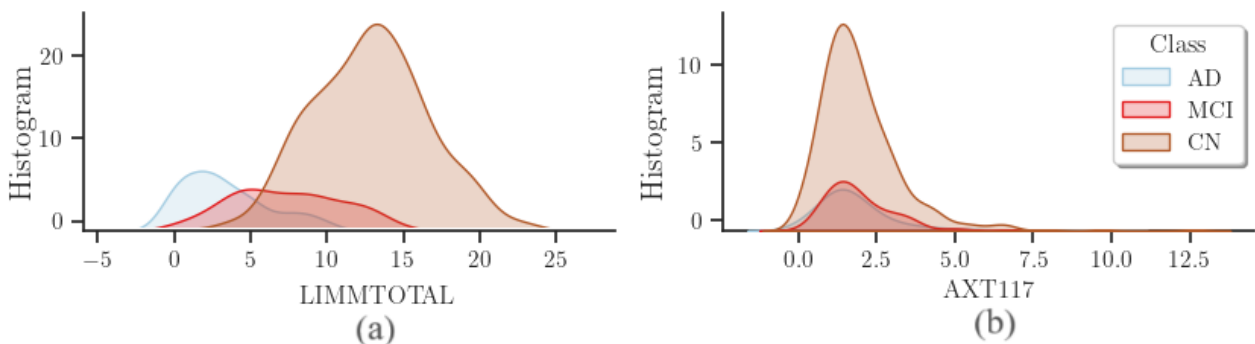
Using a feature selection algorithm based on the Fischer score was reviewed here as a reminder that the results of this process are based on two statistics (mean and variance, or standard deviation).

3.3.4.1 The FMFS algorithm

We proposed a new feature selection method named FMFS (first moment feature selection). This method is based on the first statistical moment about the origin, for the feature values relative to their classes.

The idea for FMFS arose from observing a histogram of the classes for each feature. Figure 3.2 shows that the LIMMTOTAL feature classes show greater separation than the AIBL feature AXT117. This suggests that the AXT117 feature can be excluded from the database since it contributes less to class distinction than does the LIMMTOTAL feature.

Figure 3.2. Class histograms for the LIMMTOTAL(a) and AXT117(b) features in the AIBL dataset.



Source: Own authorship.

The FMFS algorithm consists of the following steps, which will be detailed in the following paragraphs:

1. Scaling all features according to the global average of the database.
2. For each feature, calculate class averages and accumulating the Euclidean distance between averages.
3. Performing a logarithmic sweep of 100 thresholds between maximum and minimum distances.

4. Selecting features that exceed the threshold, scanned in descending order.
5. Evaluating the model at each threshold for a given database.
6. Selecting the dataset that produced the most accurate result.

The FMFS algorithm can be built from the following equations, given the nomenclature defined in Figure 3.3, where \mathbf{x} represents the entire database with \mathbf{M} instances and \mathbf{N} features, $x_{i,j}$ represents the sample of the i^{th} instance and j^{th} feature. The y_{i,c_k} value represents the c_k class, associated with the i^{th} instance. Possible classes range from c_1 to c_p .

Figure 3.3. Nomenclature definition.

$$\mathbf{x} = \mathbf{M} \text{ instances} \left\{ \begin{array}{l} \overbrace{\left[\begin{array}{cccc} \cdot & & & \\ \cdot & & & \\ \cdot & \cdot & x_{i,j} & \cdot \\ \cdot & & & \\ \cdot & & & \end{array} \right]}^{\mathbf{N} \text{ features}} \\ i \in [0 \dots \mathbf{M}-1] \\ j \in [0 \dots \mathbf{N}-1] \\ \mathbf{x} \in R^{\mathbf{M} \times \mathbf{N}} \end{array} \right. \quad \mathbf{y} = \begin{bmatrix} \cdot \\ \cdot \\ y_{i,c_k} \\ \cdot \\ \cdot \end{bmatrix} \begin{array}{l} c_k \in [c_1 \dots c_p] \\ i \in [0 \dots \mathbf{M}-1] \\ \mathbf{y} \in R^{\mathbf{N} \times 1} \end{array}$$

Source: Own authorship.

Therefore, the overall mean of the data set - μ_g , can be calculated using Equation 3.2.

$$\mu_g = \frac{1}{\mathbf{N}} \frac{1}{\mathbf{M}} \sum_{i=0}^{\mathbf{M}-1} \sum_{j=0}^{\mathbf{N}-1} x_{i,j} \quad (3.2)$$

The mean of each feature - μ_j can be calculated using Equation 3.3.

$$\mu_j = \frac{1}{\mathbf{M}} \sum_{i=0}^{\mathbf{M}-1} x_{i,j} \quad (3.3)$$

With μ_g and μ_j , all samples can be scaled, resulting in $z_{ij} = x_{ij}\mu_g/\mu_j$, such that all features have the same mean μ_g .

The average of each c_k class is calculated for each of the features j - μ_{j,c_k} , using Equation 3.4.

$$\mu_{j,c_k} = \frac{1}{\mathbf{M}_{j,c_k}} \sum_{i=0}^{\mathbf{M}_{j,c_k}-1} z_{i,j} \text{ where } (y_i = c_k) \quad (3.4)$$

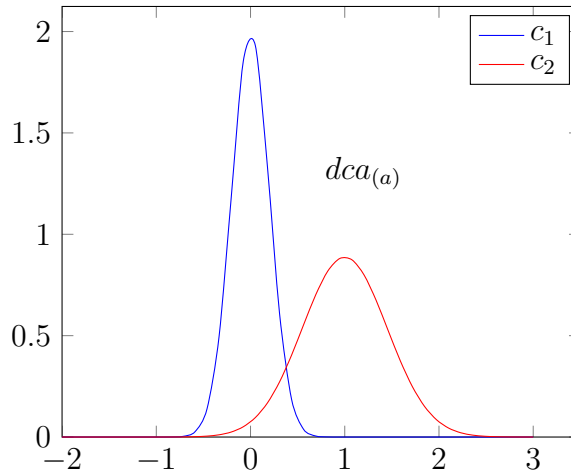
where \mathbf{M}_{j,c_k} represents the number of samples that feature j has of class c_k .

Finally, the distance between all class means of each feature j - dca_j can be calculated using Equation 3.5.

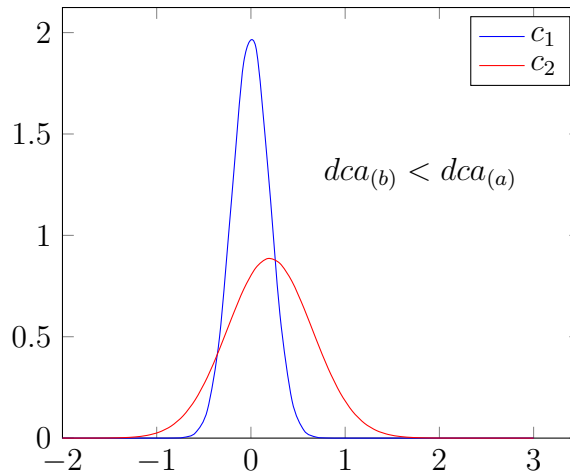
$$dca_j = \sqrt{\sum_{\substack{0 \leq k \leq p \\ 0 \leq q \leq p \\ k \neq q}} (\mu_{j,c_k} - \mu_{j,c_q})^2} \quad (3.5)$$

Figure 3.4 shows a hypothetical example of the histogram of classes c_1 and c_2 in two features 1 and 2 (a and b , respectively). In feature 1, the means of classes c_1 and c_2 are farther apart than the means of the same classes in feature 2. In this case, dca_a will certainly be greater than dca_b , which is the expected result of the FMFS method.

Figure 3.4. Features with different class means.



(a) Histogram of feature 1's classes c_1 and c_2 .



(b) Histogram of feature 2's classes c_1 and c_2 .

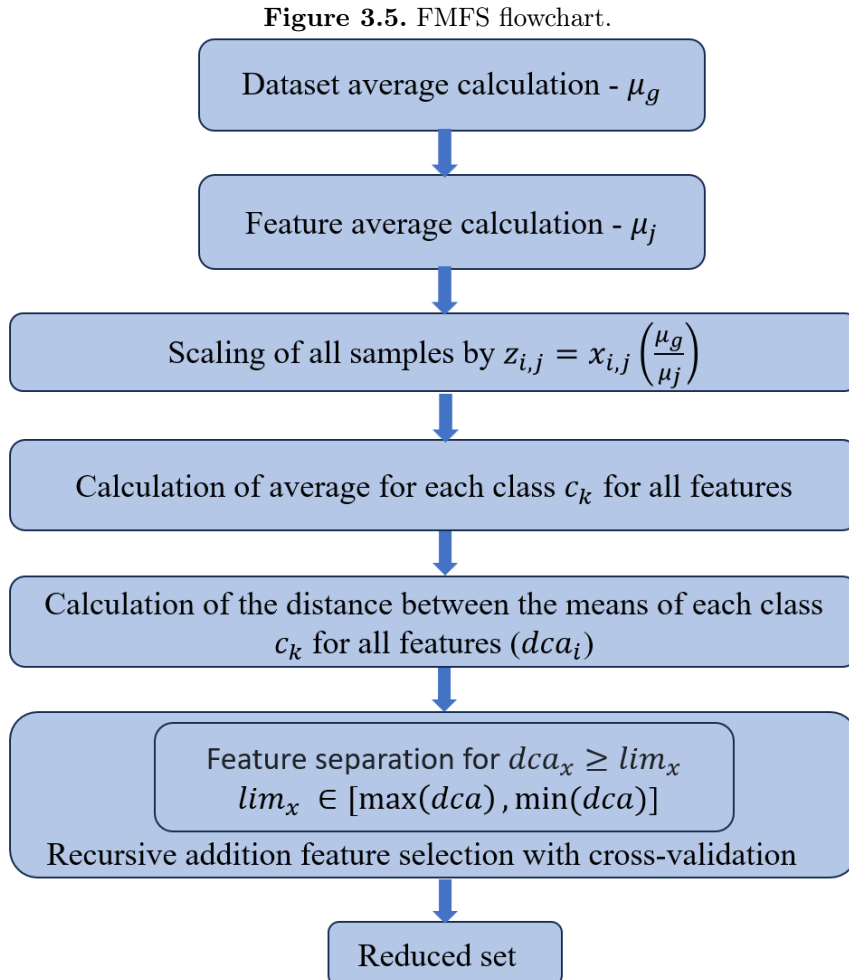
Source: Own authorship.

The features that make up a selected dataset should be those with the highest dca values, given that dca reflects the degree of separation among the classes in each feature. Thus, a top-down sweep was proposed, starting from the highest value $\max(dca)$ to the lowest value $\min(dca)$. The number of thresholds to evaluate and the type of scale (linear or logarithmic) can produce reasonably adequate results. In general, given w thresholds, the values can be calculated for a linear sweep (Equation 3.6) or a logarithmic sweep (Equation 3.7).

$$lim_x = \max(dca) - x (\max(dca) - \min(dca)) / w \quad (3.6)$$

$$lim_x = \max(dca) - x \frac{(\max(dca) - \min(dca))}{w} \log \left(\frac{(x+1)}{w} - 1 \right) \quad (3.7)$$

Figure 3.5 shows a flowchart of the FMFS algorithm, highlighting the main blocks.



Source: Own authorship.

For each lim_x , features are selected when $dca \geq lim_x$. The performance of each dataset is then evaluated using classifiers. The dataset with the best performance and smallest possible feature set is chosen.

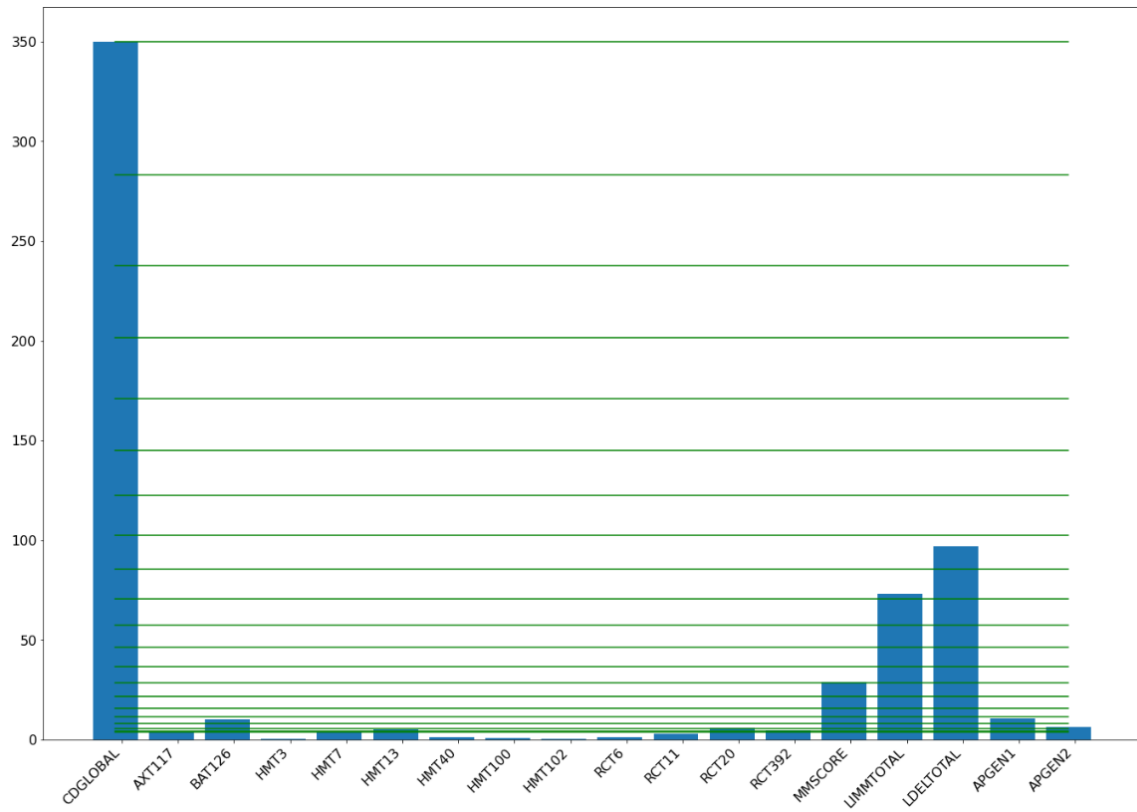
3.3.4.2 Applying FMFS to the databases

The FMFS algorithm was applied to the ADNI, AIBL and OASIS databases using a logarithmic sweep with $w = 100$, experimentally determined. For each of the databases, the ranking of feature importance and the accuracy measure of the kNN and SVM classifiers were plotted for each set of features selected by the logarithmic sweep.

Figure 3.6 shows the ranking of each feature based on the proposed algorithm for the AIBL database. Note the clear significance of the CDGLOBAL, LDELTOTAL, LIMMTOTAL and MMSCORE feature metrics compared to the other features. The features with the smallest metrics tended to have very close values, which meant that a logarithmic sweep from the highest to the lowest values was used to build the dataset.

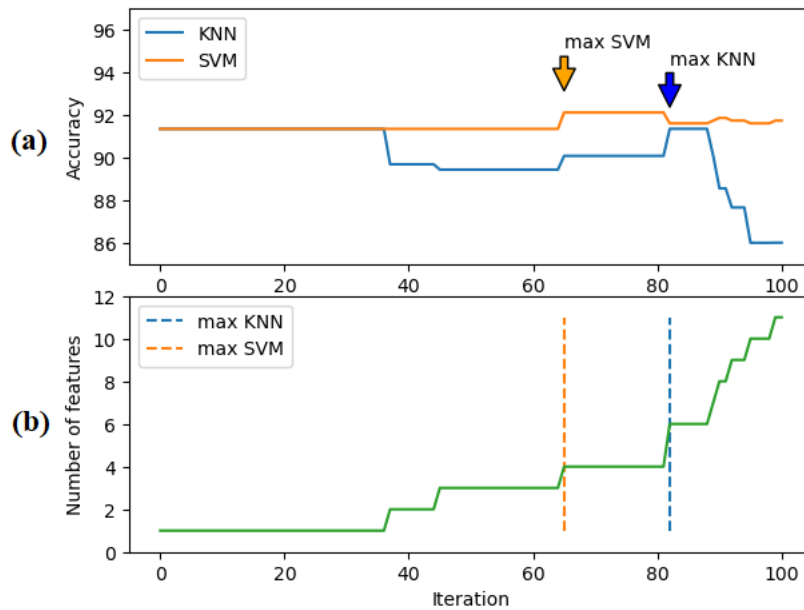
Figure 3.7 shows the evolution of the performance of each data set selected by the sweep and evaluated by the kNN and SVM algorithms. The vertical lines identify the optimal performance thresholds for each of the algorithms.

Figure 3.6. AIBL feature ranking - *dca* values.



Source: Own authorship.

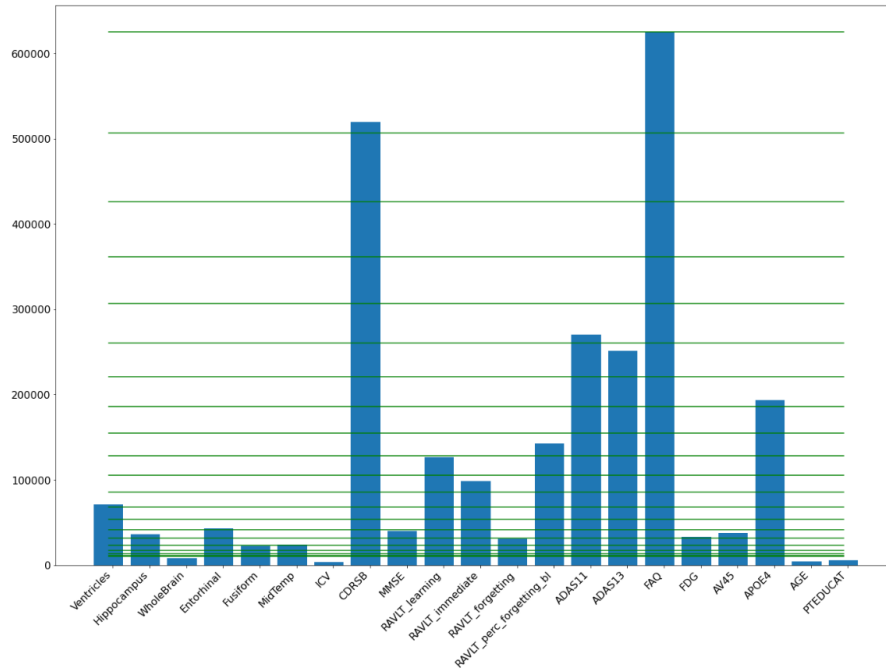
Figure 3.7. AIBL performance measurement. a) The orange and blue curves represent the evolution of SVM and kNN accuracy, respectively, with indicated maximums. b) The green curve represents the evolution of the number of selected features. The orange and blue vertical lines represent the point of maximum accuracy for SVM and kNN respectively.



Source: Own authorship.

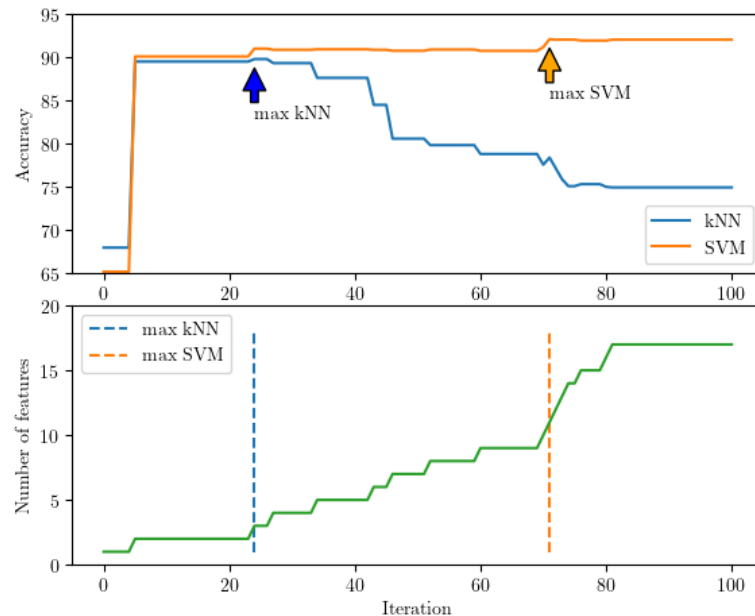
Figure 3.8 shows the same evaluation using the proposed algorithm for each feature in ADNI while Figure 3.9 shows the evolution of the performance of each selected data set and the optimal performance selection points for the kNN and SVM algorithms

Figure 3.8. ADNI feature ranking - *dca* values.



Source: Own authorship.

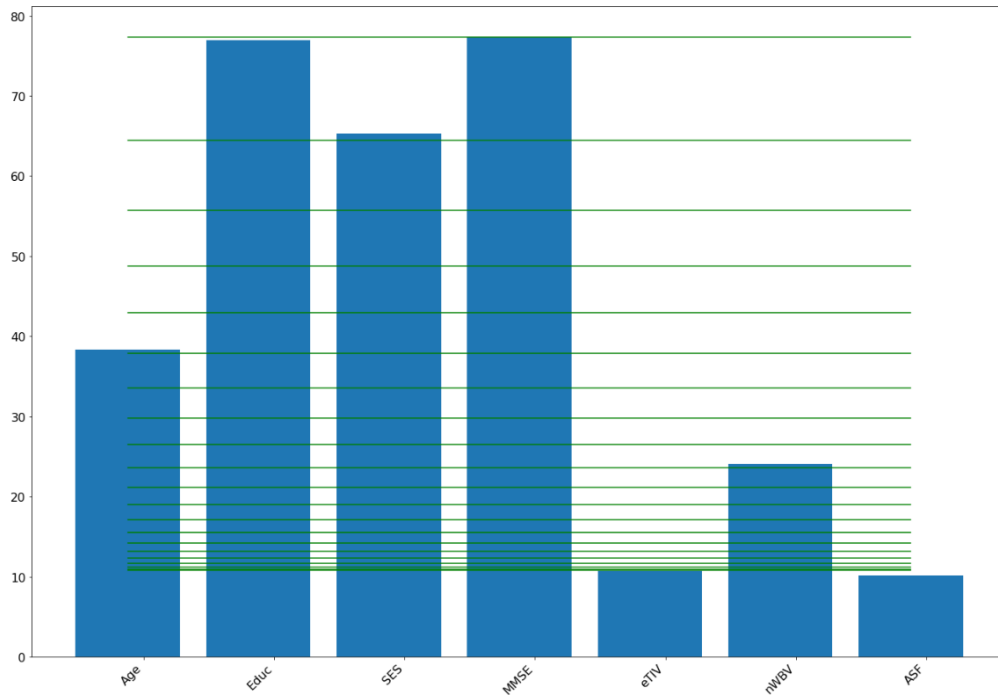
Figure 3.9. ADNI performance measurement. a) The orange and blue curves represent the evolution of SVM and kNN accuracy, respectively, with indicated maximums. b) The green curve represents the evolution of the number of selected features. The orange and blue vertical lines represent the point of maximum accuracy for SVM and kNN respectively.



Source: Own authorship.

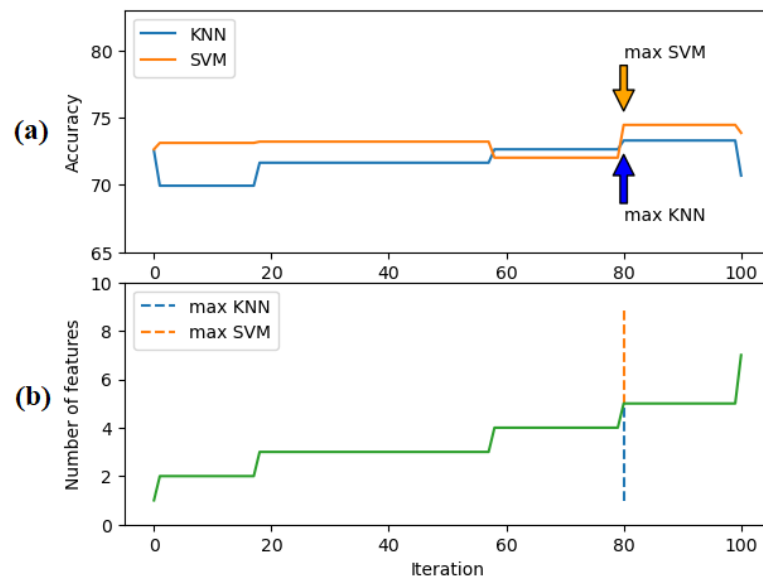
Finally, Figure 3.10 shows the ranking of each feature according to the proposed algorithm for the OASIS database while Figure 3.11 shows the performance evolution of each selected data set and the optimal performance selection points for each kNN and SVM algorithm.

Figure 3.10. OASIS feature ranking - *dca* values.



Source: Own authorship.

Figure 3.11. OASIS performance measurement. a) The orange and blue curves represent the evolution of SVM and kNN accuracy, respectively, with indicated maximums. b) The green curve represents the evolution of the number of selected features. The orange and blue vertical lines represent the point of maximum accuracy for SVM and kNN respectively.



Source: Own authorship.

3.4 PERFORMANCE METRICS

A confusion matrix (Figure 3.12) can be used to provide numerical results for the errors and successes that an evaluated model makes with a test set, resulting in the following possible outcomes: TP - true positives, TN - true negatives, FP - false positives, and FN- false negatives.

Figure 3.12. Confusion matrix.

		Predicted class	
		<i>P</i>	<i>N</i>
Actual Class	<i>P</i>	True Positives (TP)	False Negatives (FN)
	<i>N</i>	False Positives (FP)	True Negatives (TN)

Source: Own authorship.

With this approach, it was possible to define the metrics in the current study by calculating metrics for each class and arriving at a final value from the macro average of these values.

3.4.1 Accuracy - A_{cy}

The base evaluation metric for any classification system is accuracy. Accuracy is calculated as the ratio of the number of correct results to the sum of all results (Equation 3.8).

$$A_{cy} = \frac{TP+TN}{TP+FP+FP+TN} \quad (3.8)$$

3.4.2 Precision - P_{rn}

Precision is a measure of the accuracy of a diagnosis, i.e., the percentage of patients diagnosed by a system who are affected by a disease. This can be calculated as the ratio of the number of correct positive results to the total number of positive results (Equation 3.9).

$$P_{rn} = \frac{TP}{TP+FP} \quad (3.9)$$

3.4.3 Sensitivity - S_{ny}

While A_{cy} deals with both positive and negative results, the performance of a specific model in detecting positive results is evaluated using S_{ny} (or recall). S_{ny} is also called the true positive rate and is calculated as ratio of positive results that are correctly considered positive to all possible positive results (Equation (3.10)).

$$S_{ny} = \frac{TP}{TP+FN} \quad (3.10)$$

3.4.4 Specificity - S_{py}

S_{py} is also called the true negative rate and is calculated as the proportion of negative results that are correctly considered negative relative to all possible negative results. Equation (3.11) shows this calculation, which measures the performance of a specific model by detecting negative results.

$$S_{py} = \frac{TN}{TN+FP} \quad (3.11)$$

3.4.5 Balanced accuracy - B_{ca}

B_{ca} is used to measure the accuracy of a dataset that has an unbalanced number of examples for each class. In unbalanced datasets, A_{cy} provides overly optimistic model evaluations. B_{ca} can be calculated as the arithmetic mean between the true positive rate and the true negative rate (Equation (3.12)).

$$B_{ca} = \frac{S_{ny} + S_{py}}{2} \quad (3.12)$$

3.4.6 F1 score - F1

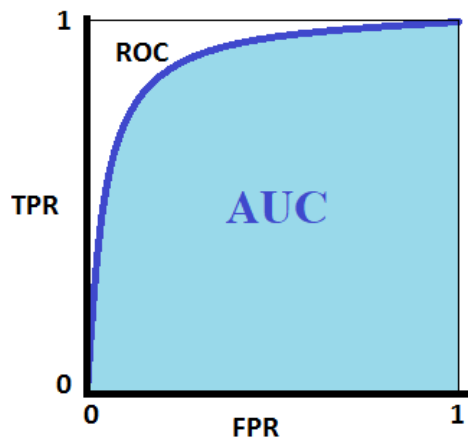
F1 was also designed to work properly on unbalanced datasets. It combines P_{rn} and S_{ny} into a single metric by taking the harmonic mean of both (Equation (3.13)).

$$\mathbf{F1} = \frac{2}{\frac{1}{P_{rn}} + \frac{1}{S_{ny}}} \quad (3.13)$$

3.4.7 Area under the ROC curve - AUC

The Receiver Operating Characteristic (ROC) is a probability curve that is evaluated at threshold settings between the true positive rate (TPR) and the false positive rate (FPR). TPR is S_{ny} , while FPR is $1 - S_{py}$, forming the TPR x FPR plane (Figure 3.13 - a hypothetical example). The area under the ROC curve (AUC) represents the degree or measure of separability between classes. It indicates how much a model can distinguish between classes. The greater the AUC, the better the model is at predicting class 0 as 0 and class 1 as 1. By analogy, the greater the AUC, the better the model distinguishes between patients with and without a disease.

Figure 3.13. Receiver Operating Characteristics (ROC) curve.



Source: Own authorship.

In a multiclass model, an N number of ROC curves can be plotted for each of N classes using the One vs All methodology. Consequently, the metric is called MAUC (multi-class AUC) and was used in the present study given that the datasets have three classes of AD stages.

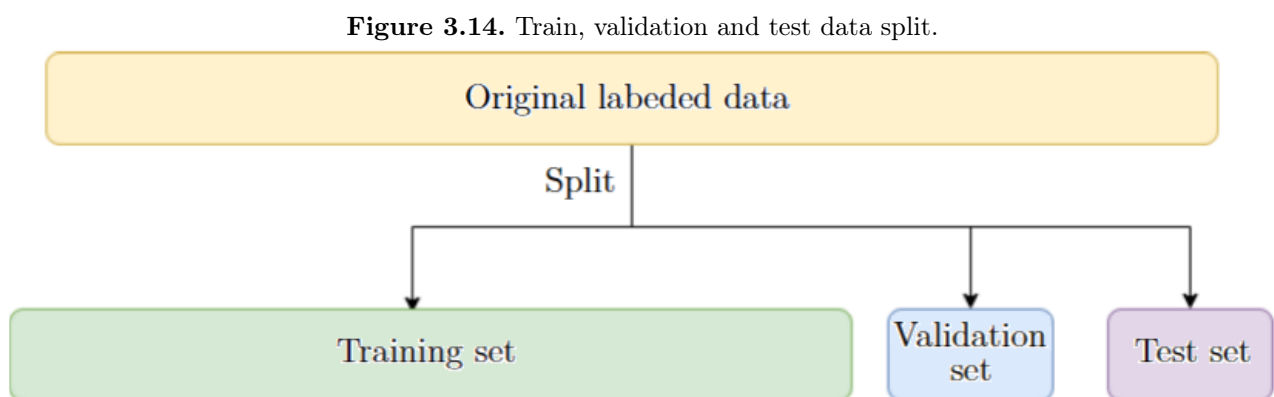
Different performance metrics imply different conclusions for a classification model. While a model can give excellent results in terms of \mathbf{P}_{rm} , it can give very poor results in terms of \mathbf{S}_{py} . Therefore, the results, including the recommended performance metrics, have been summarized in tabular form.

3.5 NESTED CROSS-VALIDATION

It is always necessary to check the consistency of an ML model. One cannot simply fit the model to training data and expect it to work accurately with real-world data. Some assurance is needed that the model has learned most of the data patterns correctly and demonstrates low bias and variance. To achieve this, the original data can be separated into three parts: training, validation and testing (Figure 3.14).

Splitting the data into three parts naturally reduces the amount of data the model can be trained on and increases the risk of missing important patterns or trends in the dataset, which in turn increases bias error.

This problem is solved by the k-Fold technique, where k represents the number of times the process is repeated on the training and validation data. This significantly reduces bias, as most of the data is being used for fit, and significantly reduces variance, as most of the data will also be used in the test set. Swapping the training and testing sets also increases the effectiveness of this method. This technique is called cross-validation and seeks to reduce model underfitting or overfitting.



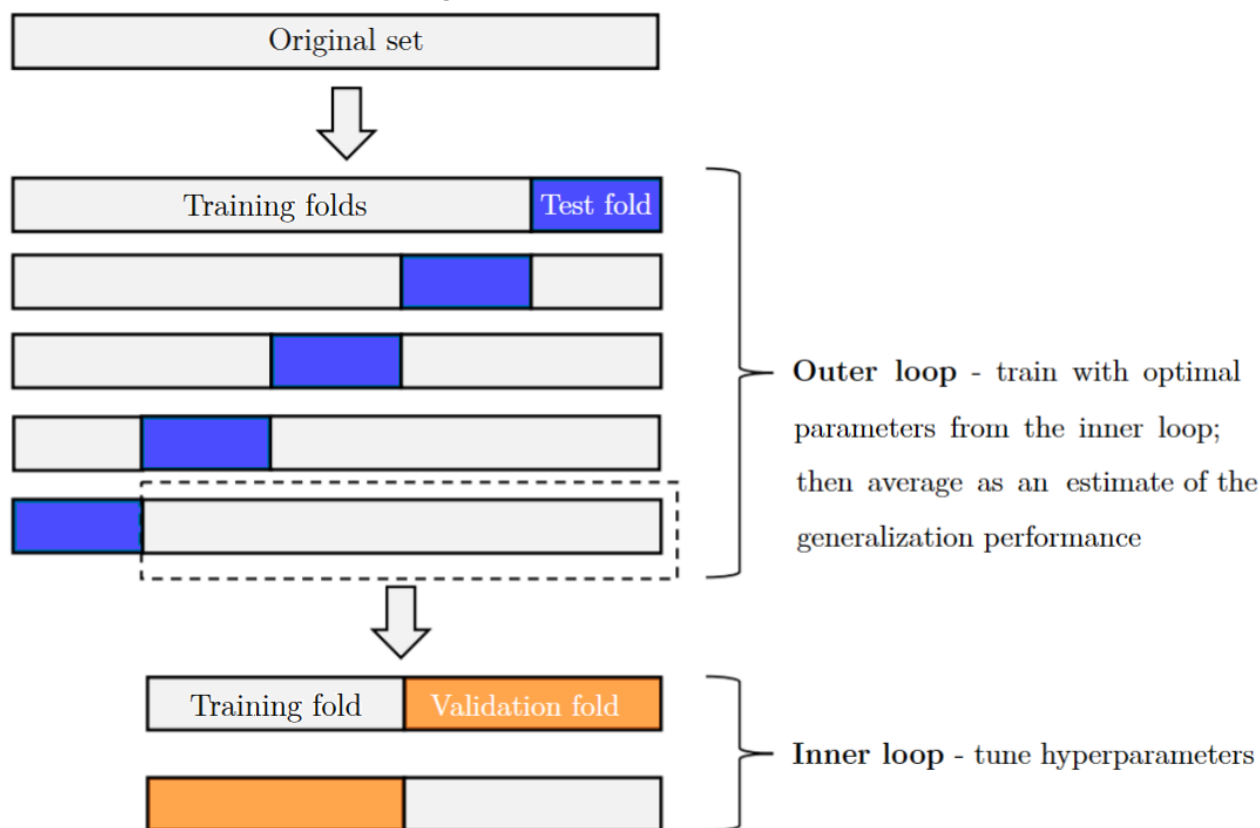
Source: Own authorship.

A standard model selection process usually includes a hyper-parameter optimization phase in which a validation technique, such as k-Fold cross-validation, is used to select an optimal model. However, this process is vulnerable to a form of selection bias that makes it unreliable in many applications (CAWLEY; TALBOT, 2010).

Therefore, we used a method based on k-Fold called nested cross-validation, which performs two cycles of k-Fold in a nested manner (PARVANDEH *et al.*, 2020). Figure 3.15 shows how nested cross-validation is applied, in six steps, as described below:

- Step 1: Split data into training and testing in the outer loop (five outer loops in the illustration). For each outer loop, perform step 2, starting with the first set of Training folds.
- Step 2: split data into training and validation sets in the inner loop (two inner loops in the illustration), which are used to tune hyperparameters using the grid search technique.
- Step 3: evaluate the best performing model on the outer loop using test data (Test fold).
- Step 4: save the parameters from the best model in this outer loop. Repeat steps 2 to 4 in the remaining outer loops.
- Step 5: choose the best model from the outer loop and then train on the complete data set to create a final model.
- Step 6: validate the final model on independent data.

Figure 3.15. Nested cross validation.



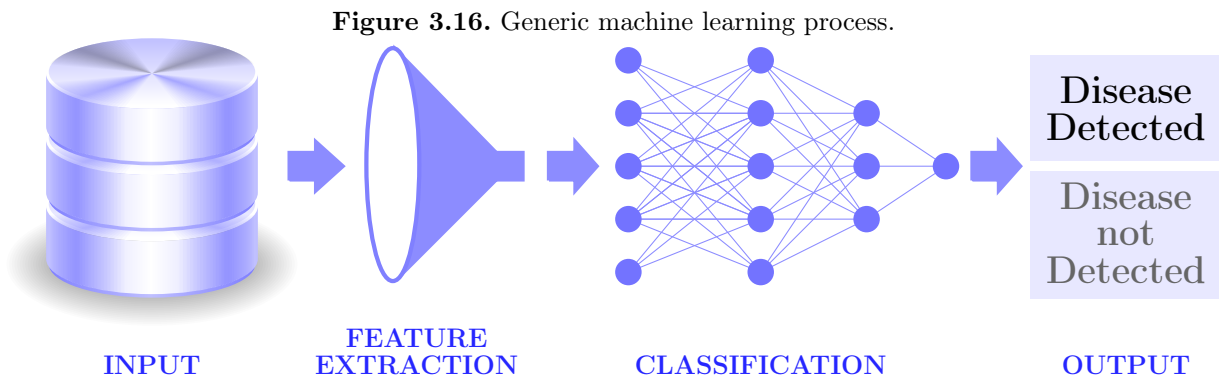
Source: Adapted from (PARVANDEH *et al.*, 2020).

Proper use of cross-validation to estimate the real error of a classifier requires that all steps of the algorithm, including the fit of the classifier parameters, be repeated in each cross-validation loop (VARMA; SIMON, 2006). Therefore, nested cross-validation provides a nearly unbiased estimate of the real classification error.

3.6 ML MODELS USED

Figure 3.16 shows the general pattern of the ML classification task. In the first stage, pre-processing and data analysis are performed to build a dataset that represents the desired objective. The next step is to select features that can facilitate the classification process by removing features of lesser importance, reducing dimensionality, and simplifying the dataset to improve performance. One of our objectives was to develop a feature selection algorithm - FMFS. Finally, the classification stage is performed to improve the performance of disease

diagnosis. While the classifier scheme shown in Figure 3.16 looks more like a neural network, it simply represents all other classification models available in ML.



The SVM and kNN classifiers were chosen for initial tests to facilitate comparisons with other studies (NIYAS; P, 2021b) (NGUYEN *et al.*, 2018), which will be briefly reviewed in the topics below.

3.6.1 SVM: support vector machine

The SVM algorithm (CORTES; VAPNIK, 1995) is a very popular algorithm in data Science due to its power and wide applicability.

Figure 3.17 shows the support vectors (gray squares) that define the margin of greatest separation, or optimal margin, between, in this case, binary classes. The optimal hyperplane separates the two classes and divides the optimal margin in half. The hyperplane is described by Equation (3.14), where $\mathbf{x} = [x_1 \ x_2]^T$ and $\mathbf{w} = [a \ 1]^T$, for the binary case.

$$\mathbf{w}^T \mathbf{x} + b = 0 \quad (3.14)$$

Thus, given an \mathbf{x}_i , the SVM classifier will decide between $h(\mathbf{x}_i) = 1$ or $h(\mathbf{x}_i) = -1$ (Equation (3.15)).

$$h(\mathbf{x}_i) = \begin{cases} 1 & \mathbf{w}^T \mathbf{x} + b \geq 0 \\ -1 & \mathbf{w}^T \mathbf{x} + b < 0 \end{cases} \quad (3.15)$$

The problem of finding the values of \mathbf{w} and b is called a constrained optimization problem,

which corresponds to finding the optimal hyperplane and is performed during the training phase of SVM.

Given a data set $D = (\mathbf{x}_i, y_i) | \mathbf{x}_i \in \mathbb{R}^n, y_i \in \{-1, 1\}$, where \mathbf{x}_i represents instances with n features and y_i refers to m labels, \mathbf{w} and b can be calculated to find the optimal hyperplane using Equations (3.16) and (3.17).

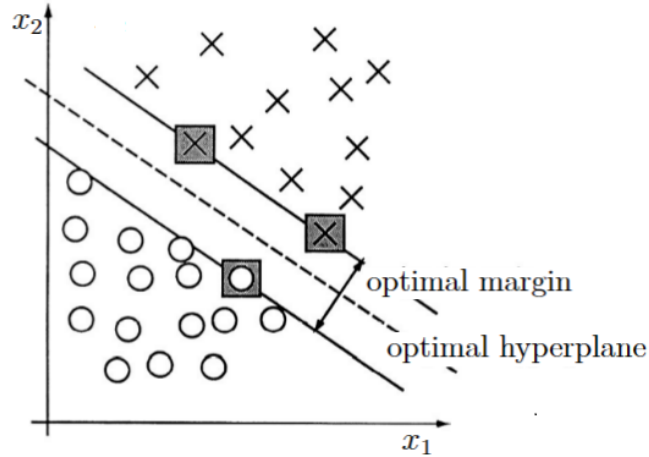
$$\mathbf{w} = \sum_{i=1}^m \alpha_i y_i \mathbf{x}_i, \quad (3.16)$$

where α_i represents the vector containing the Lagrange multipliers for each instance.

$$b = \frac{1}{S} \sum_{i=1}^S (y_i - \mathbf{w}^T \mathbf{x}_i), \quad (3.17)$$

where S is the number of support vectors.

Figure 3.17. SVM algorithm - the support vectors, marked with grey squares, define the margin of largest separation between the two classes.



Source: Adapted from (CORTESE; VAPNIK, 1995).

The strategy for calculating optimal SVM parameters is to use slack variables for the constraints of the optimization problem. A regularization parameter C is also added to determine the importance of slack, or the degree to which each training sample can be incorrectly classified. Thus, the regularized optimization problem becomes:

$$\max_{\alpha_i} \sum_{i=1}^m \alpha_i - \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m \alpha_i \alpha_j y_i y_j \mathbf{x}_i^T \mathbf{x}_j, \quad (3.18)$$

where

$$0 \leq \alpha_i \leq C, i = 1 \dots m, e \sum_{i=1}^m \alpha_i y_i = 0.$$

C is a regularization parameter included in the SVM optimization process that provides the flexibility needed to comply with all constraints for the calculation of the optimal hyperplane. Without it, the presence of outliers or non-linearly separable data can make the calculation of the optimal hyperplane impossible. Thus, with values of C ranging from zero to infinity, the strictness in meeting the constraints is controlled. Smaller C values will result in wider margins at the cost of some misclassifications, while greater C values force the classifier to tolerate fewer constraint violations. C values need to minimize noisy data and their consequent impact on solutions.

The kernel function helps transform data that is not linearly separable into another dimension that eventually facilitates the discovery of a hyperplane that can separate the data. Examples of kernel functions include the linear kernel:

$$K(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{x}_i^T \mathbf{x}_j \tag{3.19}$$

and the polynomial kernel:

$$K(\mathbf{x}_i, \mathbf{x}_j) = (\mathbf{x}_i^T \mathbf{x}_j + c)^d, \tag{3.20}$$

where c is a constant and d represents the degrees of freedom of the function.

Also, the Radial Basis Function (RBF) or Gaussian kernel, which is given as:

$$K(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\gamma \|\mathbf{x}_i - \mathbf{x}_j\|^2), \tag{3.21}$$

where low γ values cause the model to behave like a linear SVM and high values cause the model to be strongly impacted by the examples of support vectors.

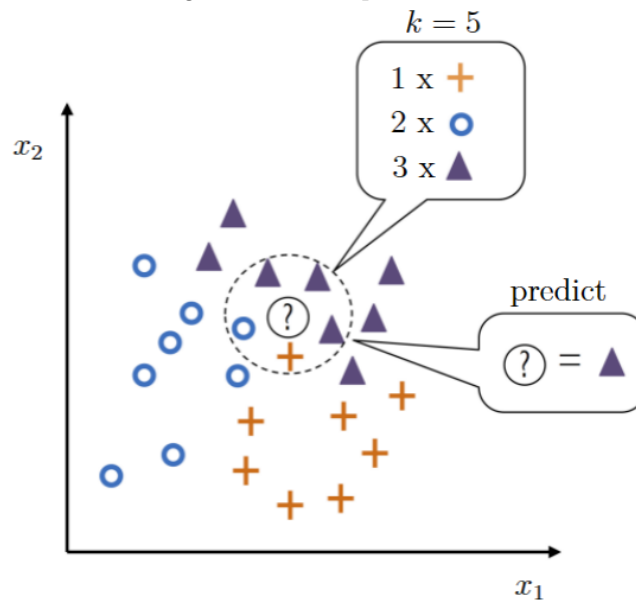
The SVM algorithm belongs to a category called kernel methods, which includes the radial base functions and linear discriminant analysis algorithms.

3.6.2 kNN: k-nearest neighbors

The kNN algorithm (COVER; HART, 1967), which stands out for its simplicity, is a non-parametric supervised learning classifier that uses proximity to make classifications or predictions on the clustering of an individual data point. kNN works on the assumption that points of the same class will be found close to each other.

Figure 3.18 shows an example with three classes (triangle, circle, cross) and the point to be classified, represented by a question mark (?). The algorithm evaluates the k nearest neighbors ($k = 5$, in this example), sums the number of points belonging to each class and selects the class with the most points (triangle, in this example).

Figure 3.18. kNN algorithm - example with three classes and $k = 5$.



Source: Adapted from (RASCHKA, 2018).

To determine which data points are closest to a given query point, the distance between the query point and other data points is calculated. Several distance measurements can be used, including the Euclidean, Manhattan, Hamming and Minkowski type, defined as:

$$dist(\mathbf{x}, \mathbf{z}) = \left(\sum_{i=1}^d |x_i - z_i|^p \right)^{1/p}. \quad (3.22)$$

Given a test point x and a set S_x of the k nearest neighbors of x , selected by calculating the distances between x and the data set, the kNN classifier will decide using:

$$h(x) = \text{mode}(y : (x,y) \in S_x), \quad (3.23)$$

where the $\text{mode}(\cdot)$ function selects the label with the greatest frequency in S_x .

The kNN algorithm is part of a category called instance-based methods, which includes learning vector quantization and self-organizing map algorithms.

3.6.3 Other ML models

Additional classifiers were also used in order to explore other options and expand the results of the study. These included Random Forest (BREIMAN, 2001), Logistic Regression (LAVALLEY, 2008), Naive Bayes (RISH, 2001), Decision Tree (FREUND; MASON, 1999), Ada Boosting (SCHAPIRE, 2013), Linear Discriminant Analysis (BALAKRISHNAMA; GANAPATHIRAJU, 1998), Gradient Boosting (FRIEDMAN, 2001), Histogram Gradient Boosting (SHAFER *et al.*, 1996) and Extreme Gradient Boosting (CHEN; GUESTRIN, 2016).

This chapter presented the primary clinical databases used, the treatments used to prepare the databases for AD diagnosis assessment, and the technological proposals developed, with emphasis on the proposed feature selection algorithm. In the next chapter, technological proposals are also made for using magnetic resonance images to aid in AD diagnosis.

IMAGE BASED ANALYSIS

This chapter details the image data used and the technological proposals developed.

Cross-sectional imaging techniques such as computed tomography and magnetic resonance imaging provide a valuable non-invasive method for detecting the cortical atrophy that is commonly seen in AD. Spectroscopy of magnetic resonance imaging (MRI) and metabolic imaging techniques such as positron emission tomography (PET) are even more specific and provide evidence of unique metabolic alterations in memory pathways, thus providing a wealth of new information for researchers (NORFRAY; PROVENZALE, 2004). An important advantage of magnetic resonance imaging is that it does not emit ionizing radiation.

4.1 AD DIAGNOSIS BASED ON MRI

The main role of magnetic resonance imaging (and computed tomography) in diagnosing Alzheimer's disease is assessing characteristic changes in cerebral volume that can yield diagnostic accuracy of up to 87% (NORFRAY; PROVENZALE, 2004).

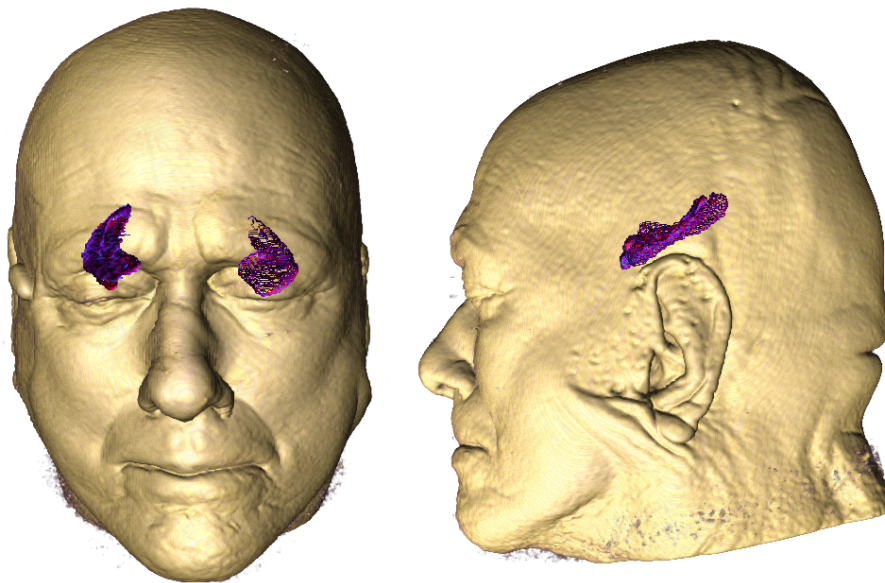
Atrophy of the mesial temporal lobe, particularly of the hippocampus, entorhinal cortex and perirhinal cortex, can be used to diagnose AD (PATEL *et al.*, 2020) in a non-invasive way.

Brain volume measurements, assessed with segmentation, demonstrate that patients with Alzheimer's disease have accelerated rates of brain volume loss, roughly twice normal rates (1% vs. 0.5% per year). This is even more pronounced in the hippocampus where yearly volume losses can be three times greater (4.5% vs 1.5% per year) (MILLER-THOMAS *et al.*, 2016).

Figure 4.1 illustrates the position and structure of the hippocampus in the brain cavity (CC license BY-NC-ND) (ESTUDIOS, 2017). The hippocampus is a complex structure in the mesial temporal lobe and participates in the limbic system and memory circuit. It measures about 4.5 cm long and is composed of three parts: the head, also called the foot, body and tail

(ROGACHESKI *et al.*, 1998).

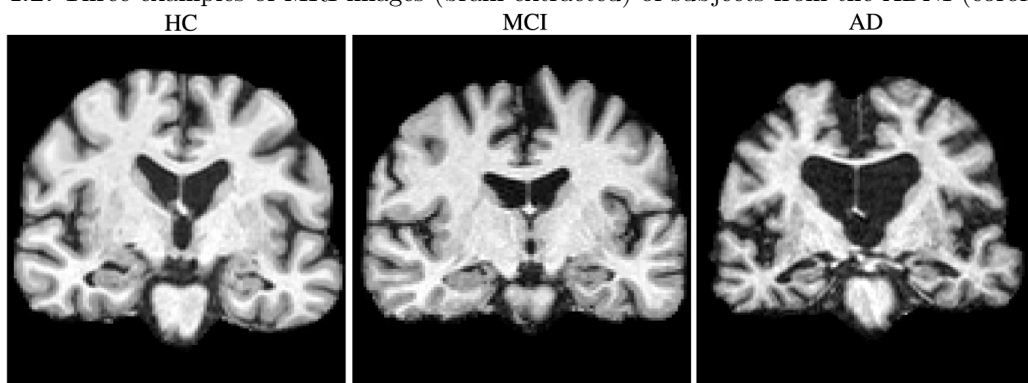
Figure 4.1. Illustration of the hippocampus in the brain.



Source: Own authorship.

(LEDIG *et al.*, 2018) studied brain structure volumes in a dataset of 5,074 MRI images. Three cases, at different stages of AD stood out: a healthy 84.8-year-old male control (HC - healthy control), a 71.8-year-old female with MCI, and a 77.5-year-old male with AD. Figure 4.2 shows a coronal view of each of these patients, each at different stages of AD.

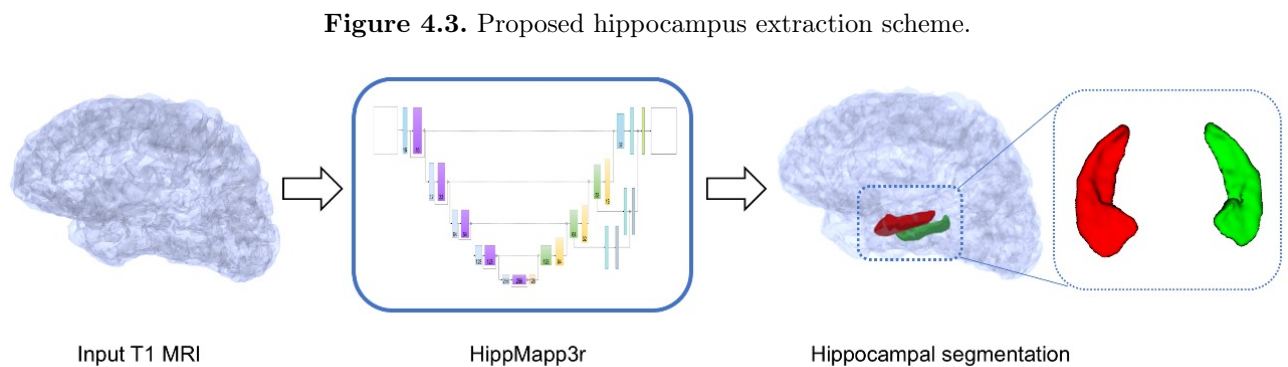
Figure 4.2. Three examples of MRI images (brain-extracted) of subjects from the ADNI (coronal slices).



Source: Adapted from (LEDIG *et al.*, 2018).

The present study will explore perceived changes and losses in brain structures that are evidenced by MRI images.

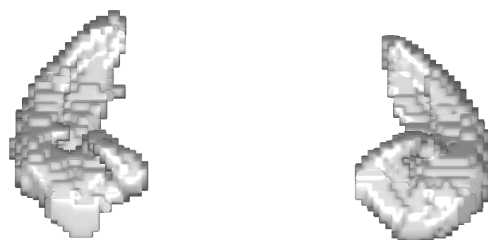
In (GOUBRAN *et al.*, 2020), total hippocampal segmentation from MRI was proposed. The proposed algorithm, based on a convolutional neural network, was called HippMapp3r. Figure 4.3 shows the general scheme of the processing carried out.



Source: (GOUBRAN *et al.*, 2020).

An example of the result generated by HippMapp3r is shown in Figure 4.4.

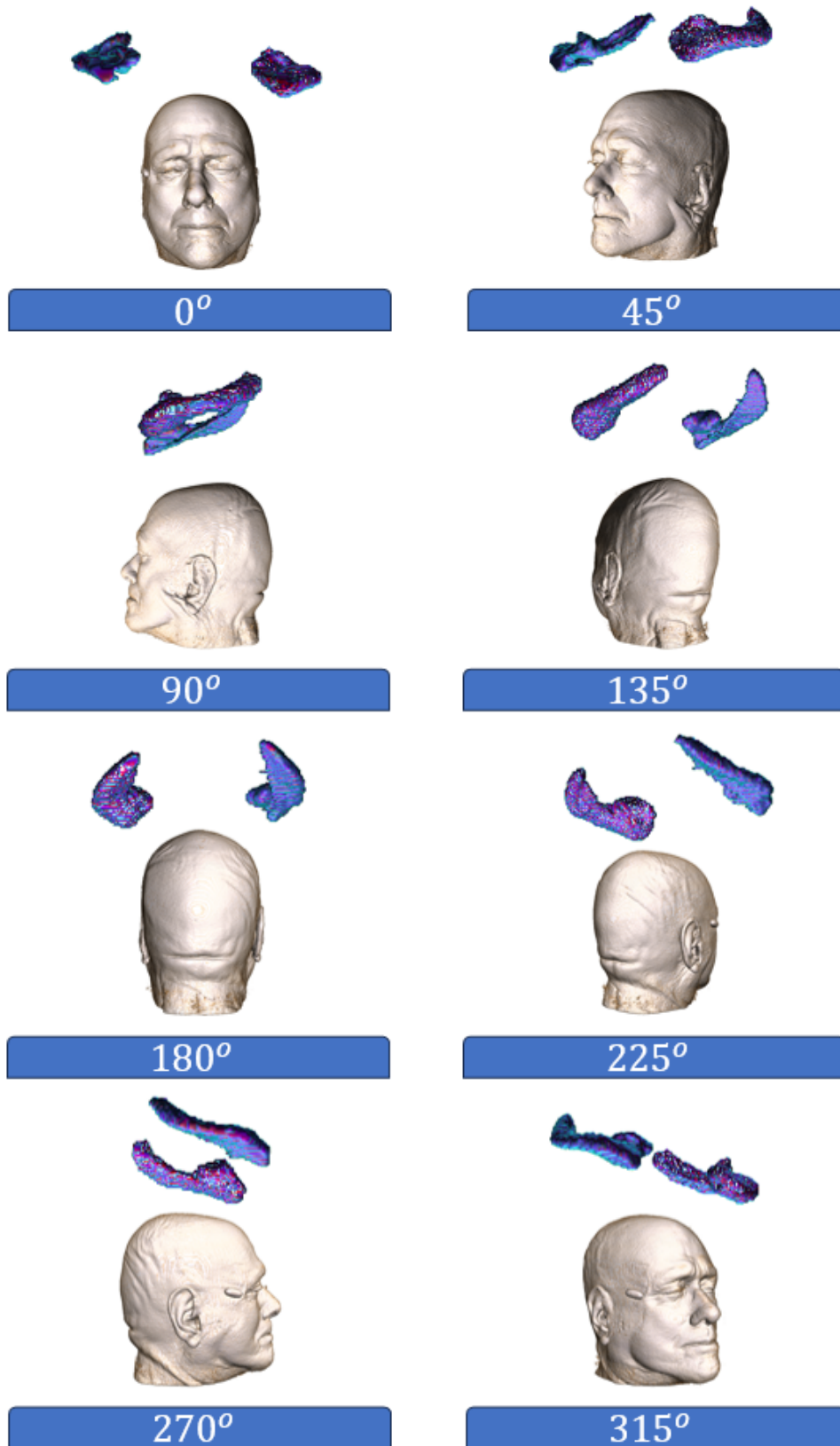
Figure 4.4. Extracted hippocampus.



Source: Own authorship.

Figure 4.5 shows the hippocampus highlighted above the subject's head at various viewing angles. The subject's head and the hippocampus are at different scales, but at the same angle.

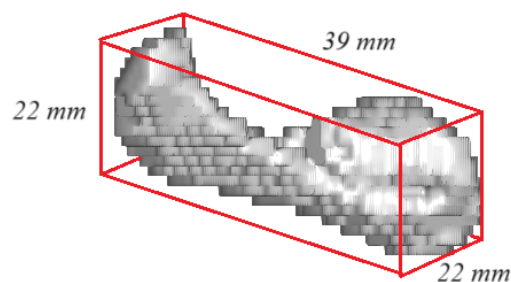
Figure 4.5. Visualization of the hippocampus from various angles..



Source: Own authorship.

Using this same example, it was possible to measure the maximum dimensions of the hippocampus in the 3 planes (Figure 4.6), resulting in 22 mm when measured from left to right, 22 mm when measured from top to bottom and 39 mm when measured from the front to the back of the subject's head.

Figure 4.6. Maximum dimensions of the hippocampus in the 3 planes. Right hippocampus rotated 300 degrees.



Source: Own authorship.

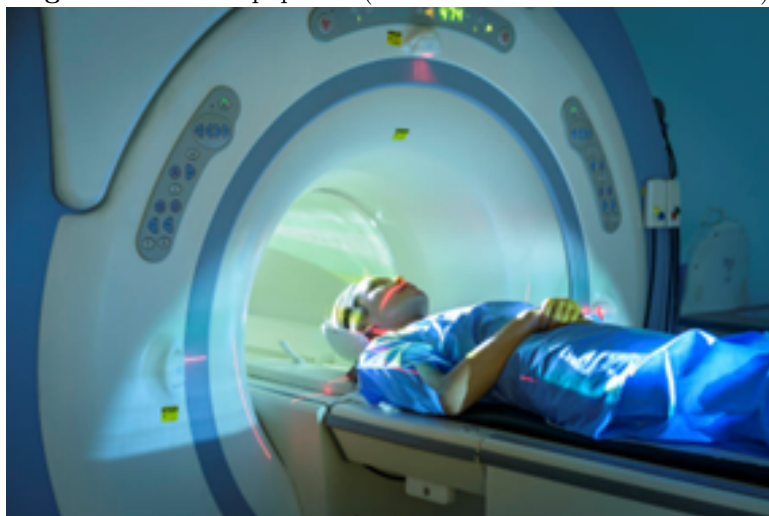
These measurements reveal that the coronal plane has the largest number of MRI *slices* that can cross some region of the hippocampus.

4.2 MAGNETIC RESONANCE IMAGING

MRI is a non-invasive imaging technology that produces detailed three-dimensional anatomical images. It is often used for disease detection, diagnosis, and treatment monitoring. The technology is based on sophisticated mechanisms that excite and detect changes in the axis of rotation of protons found in the water that makes up living tissue.

An MRI image is obtained by placing a patient inside a large magnetic field where they must remain still during the imaging process. Figure 4.7 shows the typical MRI equipment and environment (CC license BY-ND)(MELBOURNE, 2018). MRI is preferred over x-rays when frequent imaging is required for diagnosis or therapy, especially of the brain, since it does not emit ionizing radiation.

Figure 4.7. MRI equipment (creative commons license BY-ND).



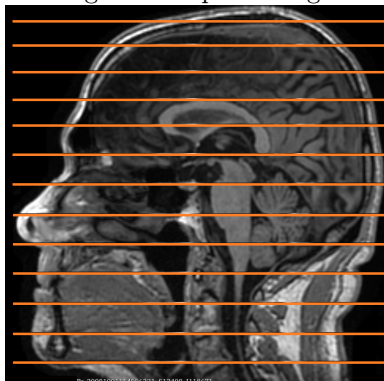
Source: (MELBOURNE, 2018).

4.2.1 Digital format of magnetic resonance images

A digital image file of the internal structures under evaluation is produced at the end of an MRI examination. The file consists of a header and slice data in three dimensions (X, Y and Z).

Figure 4.8 shows a representation of the MRI slices in one plane of a patient's head.

Figure 4.8. Orange lines representing axial MRI slices.



Source: Own authorship.

The image file format is often a confusing aspect of medical image processing. The predominant formats in radiology are: Digital Imaging and Communications in Medicine (DICOM), Neuroimaging Informatics Technology Initiative (NIFTI), Philips MRI scanner formats (PAR/REC), Analyze 7.5 (ANALYZE), Nearly Raw Raster Data (NRRD) and

Montreal Neurological Institute (MINC). According to (LAROBIANA; MURINO, 2013), the main formats used for medical images are: ANALYZE, NIFTI, MINC and DICOM.

4.2.1.1 NIFTI format

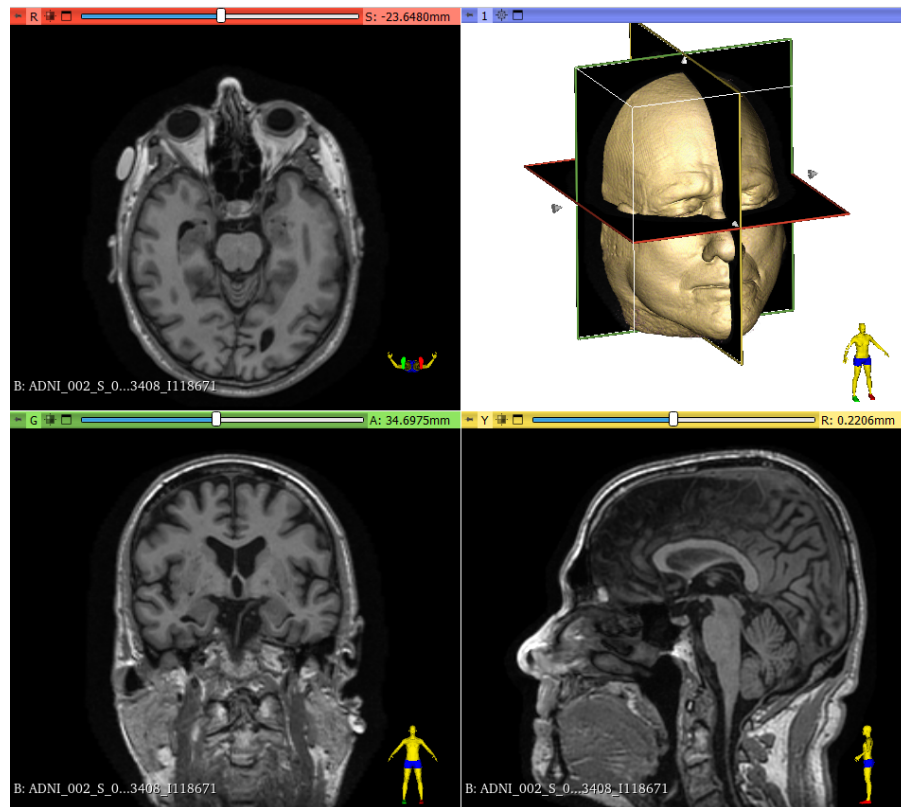
The NIFTI format is widely used in image informatics for neuroscience and neuroradiology research and was also used in the present study. DICOM is the standard format used in clinical care and can be converted to NIFTI with some imaging tools (LI *et al.*, 2016).

NIFTI stands for the Neuroimaging Informatics Technology Initiative, which was a working group at the NIH (National Institutes of Health in the USA) in the early 2000s. The Neuroimaging Informatics Technology Initiative strove to create a file type that was like the ANALYZE file format but enhanced. At that time, MINC and ANALYZE were the most common file formats used in neuroimaging research.

A NIFTI file can be used to visualize and process MRI images for specific objectives. Figure 4.9 shows the axial (red bar), coronal (green bar), and sagittal (yellow bar) views of an MRI in NIFTI format from an 85-year-old male patient with a normal neurological diagnosis. A 3D model of the patient's skull (blue bar) was also built using the images from the three planes. The image was captured using the 3D Slicer image computing platform¹, which provides a free and open-source software solution for visualization, processing, segmentation, registration, and analysis of medical, biomedical, and other 3D images and meshes. The platform can also be used for planning and navigating image-guided procedures. The 3D Slicer software is referenced in (FEDOROV *et al.*, 2012).

¹<https://www.slicer.org/>

Figure 4.9. Axial (red bar), coronal (green bar) and sagittal (yellow bar) slice views of an MRI in NIFTI format.



Source: Own authorship.

4.3 DATABASE DESCRIPTION

The Kaggle² website provides some MRI databases for AD³ diagnosis. Nevertheless, we chose to extract data directly from referenced studies on AD, such as the Alzheimer’s Disease Neuroimaging Initiative (ADNI).

4.3.1 ADNI

The Alzheimers Disease Neuroimaging Initiative (ADNI) (MARINESCU *et al.*, 2018) is a multicenter, longitudinal North American study designed to develop imaging, genetic, and biochemical clinical biomarkers for the early detection and screening of AD. Since its launch over a decade ago, this historic public-private partnership has made major contributions to AD

²<https://www.kaggle.com/>

³<https://www.kaggle.com/datasets/tourist55/alzheimers-dataset-4-class-of-images>

research, enabling data sharing among researchers around the world. Additional information can be found at the study’s website⁴. This study also classifies disease progression into three categories: CN, MCI, and AD.

ADNI study data can be downloaded from the Image and Data Archive (IDA)⁵ platform, which is managed by the Laboratory of Neuro Imaging (LONI) of the Mark and Mary Stevens Neuroimaging and Informatics Institute associated with the University of Southern California.

4.3.1.1 MRI in ADNI

Image collection is an essential part of ADNI’s goal of developing biomarkers for tracking AD progression. Image acquisition is carried out in several stages, ADNI 1, ADNI GO, ADNI 2 and ADNI 3, with several other variations in technological and acquisition conditions.

We selected MRI data from ADNI phase 1, particularly at the time of patient screening. This dataset is standardized by ADNI, denoted as ADNI1: Screening 1.5T, and has the following general characteristics:

- contains data from 1075 patients, aged 55 to 91 years old.
- collected during their first visit to the study.
- 619 male and 456 female patients.

Data in ADNI1: Screening 1.5T were collected from January 3rd to September 29th, 2006 and the 1075 participants were diagnosed at the CN, MCI and AD stages (Table 4.1).

Table 4.1. ADNI1:Screening 1.5T - Number of patients per class

Class	Quantity
CN	307 patients
MCI	525 patients
AD	243 patients

Source: Own authorship.

According to ADNI, the ADNI1: Screening 1.5T magnetic resonance protocol aimed to acquire consistent structural images on 1.5 Tesla scanners using T1-weighted and T2-weighted

⁴<https://adni.loni.usc.edu/>

⁵<https://ida.loni.usc.edu/login.jsp>

sequences. In most clinical environments, a 1.5T MRI scanner is more than adequate for generating clear images and providing evidence for diagnosis and prognosis. A 3T scanner was also used by ADNI in data acquisition phases, such as ADNI 3, to achieve other objectives.

Pulse radio frequencies and gradient waves can be used to control highlighting in the scanned images. T1-weighted images show the anatomy of soft and fatty tissue while T2-weighted images highlight fluids and pathologies.

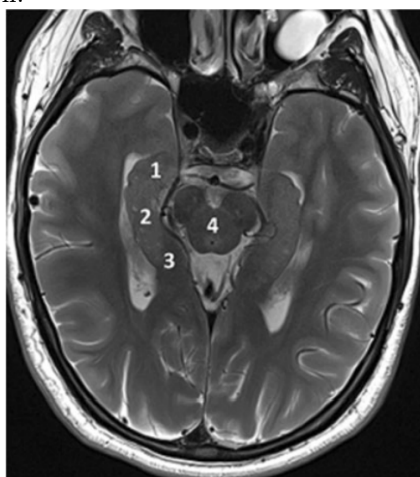
The images in ADNI1: Screening 1.5T are available in the NIFTI format and can be downloaded from the IDA platform. Figure 4.9 shows an MRI from anonymous patient 002_S_0295 in the ADNI1: Screening 1.5T dataset, highlighting specific slice views in the axial, coronal and sagittal planes.

4.3.1.2 Slice selection

After defining access and primary data, our next challenge was to make appropriate choices for image analysis. Note that not all MRI slices are relevant for the study of AD. The slices of interest are those that show the structures altered by AD evolution, such as the hippocampus (DEKEYZER; BOSSCHE, 2017), which shows yearly volume losses that are up to three times greater than those of normal individuals (MILLER-THOMAS *et al.*, 2016).

For example, Figure 4.10 identifies various hippocampal regions in an axial slice of a 3T T2-weighted MRI (DEKEYZER; BOSSCHE, 2017).

Figure 4.10. Anatomy of the hippocampal formation: 1 = hippocampal head, 2 = hippocampal body, 3 = hippocampal tail, 4 = mesencephalon.

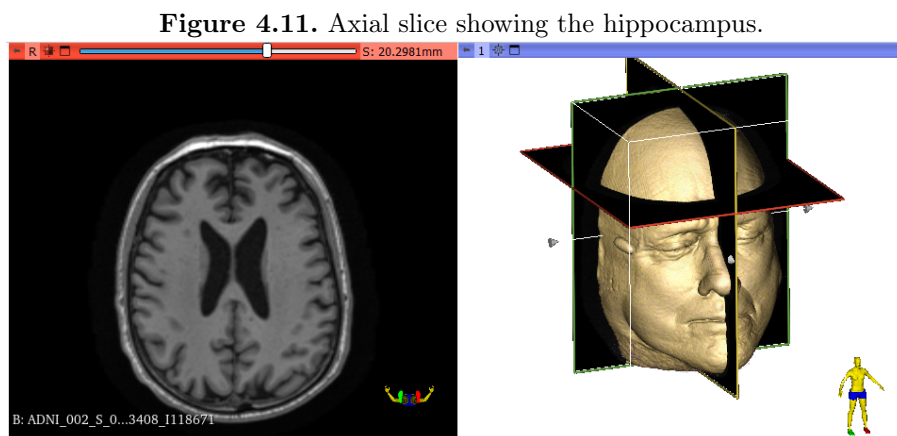


Source: Adapted from (DEKEYZER; BOSSCHE, 2017).

Slices in any of the three axial, coronal or sagittal planes will certainly reveal aspects of the hippocampus, assuming that the data was acquired at a very fine spatial resolution and over the entire region of the patient’s head. However, we chose axial and coronal slices to generate the images that composed the analysis dataset. Sagittal slices were not evaluated because they do not show the right and left parts of the hippocampus simultaneously.

4.4 AXIAL PLANE EVALUATION

Figure 4.11 shows an axial slice selected using 3D Slicer software.



Source: Own authorship.

In the current study, a neuroradiologist could have inspected each of the MRIs in the ADNI1: Screening 1.5T database (1075 patients) and chosen slices that best represent the hippocampal region.

However, since our premise was the autonomous development of computational intelligence tools to aid in AD diagnosis, we decided to develop a slice selection algorithm that scanned all MRIs and selected the most promising slices.

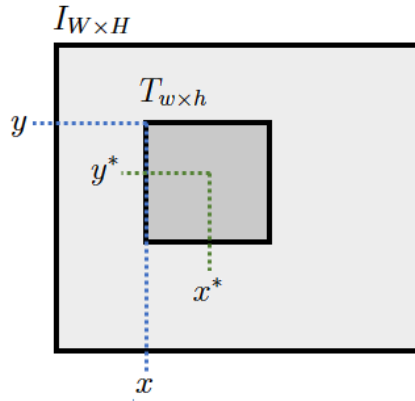
4.4.1 Proposed slice selection algorithm

Given an axial slice image I (dimensions $W \times H$), determine whether this image shares similarities with a template image T that represents the anatomy of the hippocampus (dimensions $w \times h$, where $w \leq W$ e $h \leq H$), and then select the slice with greatest similarity.

4.4.1.1 SMR algorithm

We proposed an algorithm for selecting slices based on similarity, called SMR (similarity matching rule), that uses a combination of three template matching functions. The three functions slide the template T over image I (Figure 4.12), for all $x \leq W - w + 1$ and $y \leq H - h + 1$. This yields a similar image $R(x,y)$, with the dimensions $[W - w + 1 \times H - h + 1]$.

Figure 4.12. Image matching.



Source: Own authorship.

The first function is $R_{diff}(x,y)$ (Equation 4.1), which performs the normalized sum of squared differences for each point on the image template, starting from point (x,y) . The minimum value of $R_{diff}(x,y)$ determines the point (x,y) of greatest similarity between the image and the template.

$$R_{diff}(x,y) = \frac{\sum_{x^*,y^*} (T(x^*,y^*) - I(x + x^*,y + y^*))^2}{\sqrt{\sum_{x^*,y^*} T(x^*,y^*)^2 \sum_{x^*,y^*} I(x + x^*,y + y^*)^2}} \quad (4.1)$$

The second function, $R_{dot}(x,y)$ (Equation 4.2), calculates the normalized sum of the product for each point on the template with that of the image, starting from point (x,y) . The maximum value of $R_{dot}(x,y)$ determines the point (x,y) of greatest similarity between the image and the template.

$$R_{dot}(x,y) = \frac{\sum_{x^*,y^*} (T(x^*,y^*)I(x + x^*,y + y^*))}{\sum_{x^*,y^*} T(x^*,y^*)^2 \sum_{x^*,y^*} I(x + x^*,y + y^*)^2} \quad (4.2)$$

The third function, $R_{cdot}(x,y)$ (Equation 4.3), calculates the normalized sum of the product for each point on the template, subtracted from its average, with that of the image, subtracted

from its average, starting from point (x,y) . The maximum value of $R_{cdot}(x,y)$ determines the point (x,y) of greatest similarity between the image and the template.

$$R_{cdot}(x,y) = \frac{\sum_{x^*,y^*} (T^*(x^*,y^*) I^*(x+x^*,y+y^*))}{\sum_{x^*,y^*} T(x^*,y^*)^2 \sum_{x^*,y^*} I(x+x^*,y+y^*)^2} \quad (4.3)$$

Specifically, $R_{cdot}(x,y)$ calculates the product of the pixel values subtracted from the mean (Equations [4.4](#) and [4.5](#)).

$$T^*(x^*,y^*) = T(x^*,y^*) - 1/(w h) \sum_{x^{**},y^{**}} T(x^{**},y^{**}) \quad (4.4)$$

$$I^*(x+x^*,y+y^*) = I(x+x^*,y+y^*) - 1/(w h) \sum_{x^{**},y^{**}} I(x+x^{**},y+y^{**}) \quad (4.5)$$

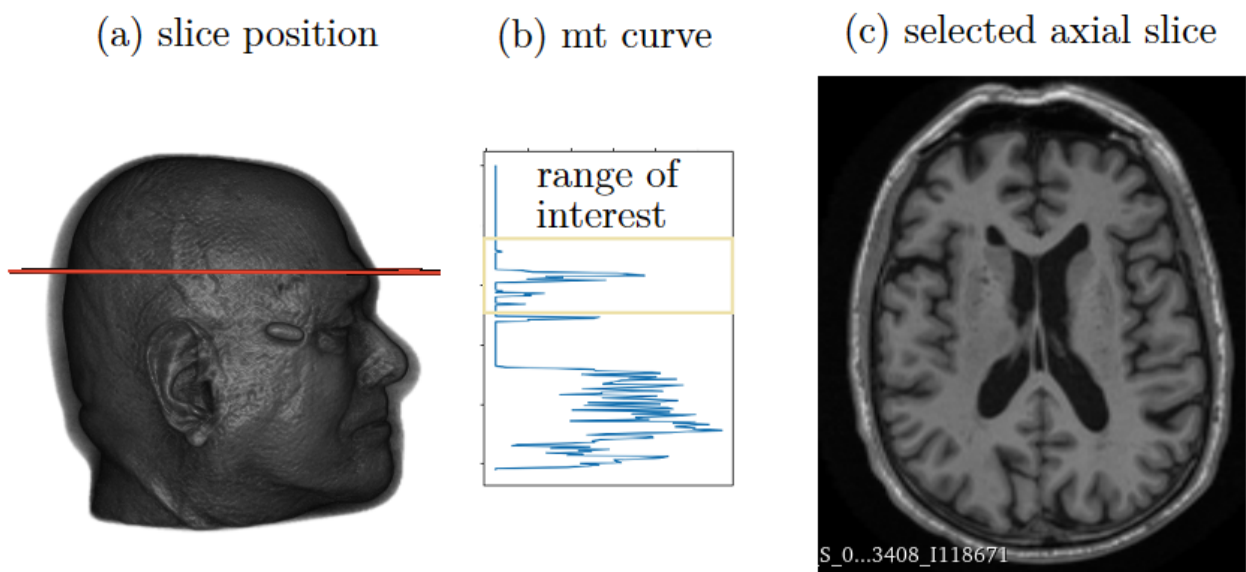
Therefore, $R_{cdot}(x,y)$ calculates the normalized covariance between image I and template T.

Finally, the SMR algorithm combines these three functions for each slice using Equation [4.6](#), where th is a threshold of 0.01, experimentally determined.

$$mt = \max(R_{dot}) \max(R_{cdot}) (1 - \min(R_{diff}) + th) \quad (4.6)$$

Figure [4.13](#) shows an example of the SMR algorithm with an MRI from anonymized patient 002_S_0295, (ADNI1: Screening 1.5T database). The red line in (a) shows the position of the selected slice relative to the patient's head. The search range for the slice is highlighted in yellow on the mt curve in (b). Naturally, the position of the selected slice coincides with the peak of the mt curve, within the range of interest. Finally, the image of the selected slice is shown in (c).

Figure 4.13. Example SMR result. The red line in (a) shows the position of the selected slice with respect to the patient’s head. The range of interest on the mt curve (b) for the slice search is outlined in yellow. The image of the selected slice is shown in (c).



Source: Own authorship.

The SMR algorithm can also be used to augment input data by selecting one or more slices from the MRI, if necessary. Since the ADNI1: Screening 1.5T dataset contains 1075 MRI images and convolutional networks require large volumes of data for training, the slices that are lateral to a slice selected by SMR can also be used to train the convolutional network.

4.4.1.2 Applying SMR to the data base

The ADNI1: Screening 1.5T database contains 1075 MRI images from three classes of patients CN, MCI, and AD. These images are in the NIFTI format, with axial, coronal and sagittal slices.

The SMR algorithm was used to scan all axial slices from the 1075 MRI images. The slice with the largest mt value was then selected from within the range of interest.

The T template measured 70 by 103 pixels (Figure 4.14). Image I has dimensions 166×256 pixels, corresponding to a specific slice in the axial plane.

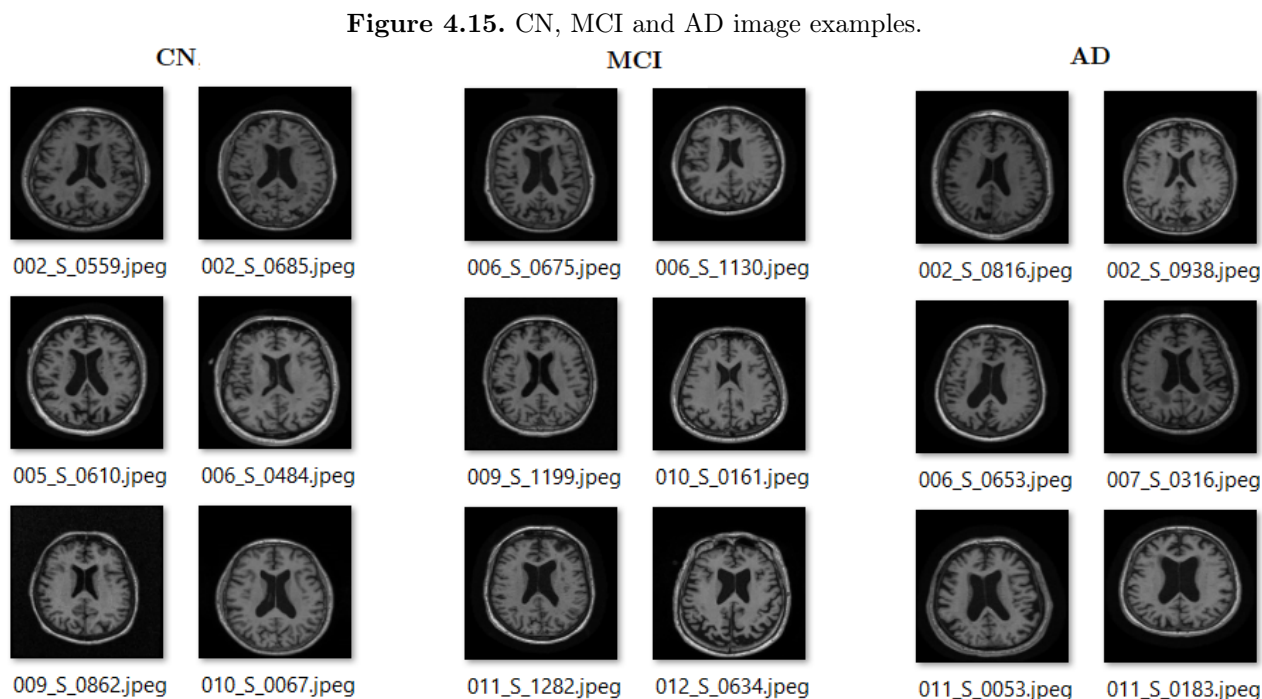
Figure 4.14. Template.



Source: Own authorship.

4.4.2 Resulting image dataset

After processing the ADNI1: Screening 1.5T database, CN, MCI and AD directories were created to organize the images by class. Figure 4.15 shows examples of the resulting images before subsequent processing.



Source: Own authorship.

All images were saved in the JPEG format at 176 pixels by 176 pixels.

4.5 EVALUATION IN THE CORONAL PLANE

The coronal plane simultaneously shows both the left and right structures of the hippocampus (Figure 4.16)

Figure 4.16. Red circles identify the left and right structures of the hippocampus in the coronal plane.

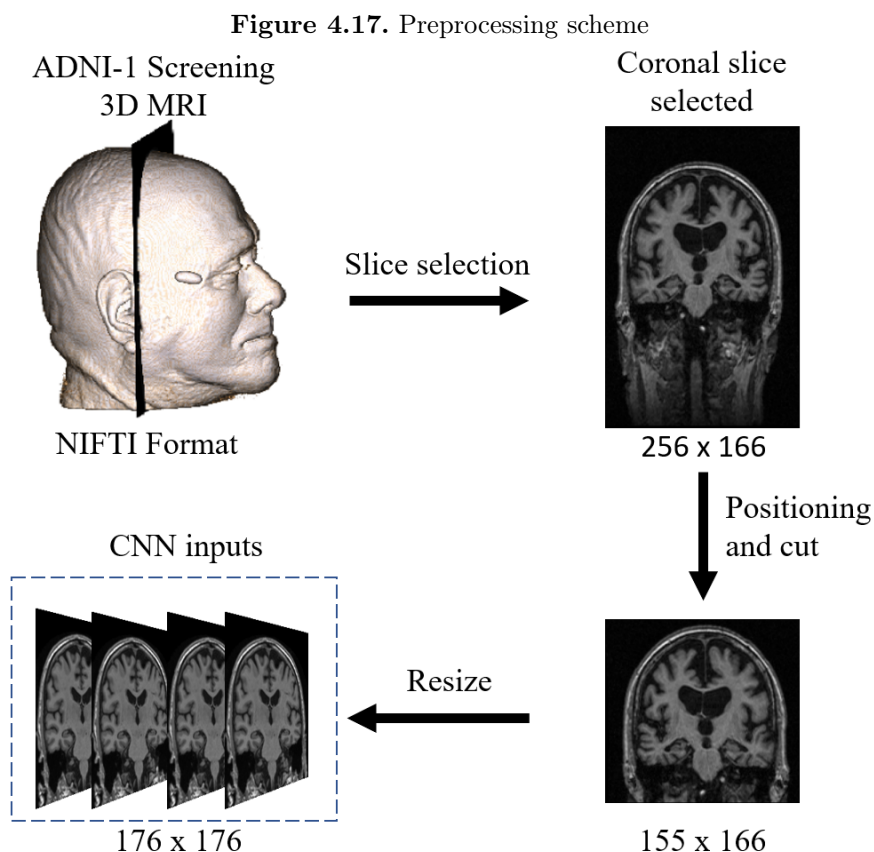


Source: Own authorship.

Note that in Figure 4.16, the right and left sides (red circles) are very similar and mirrored across the axis of the coronal plane. This aspect was used in slice selection.

4.5.1 Strategy for building the image dataset

A different strategy was used in the coronal plane than in the axial. Here, all CN and AD MRI images in the ADNI1: Screening 1.5T database were scanned manually. Figure 4.17 illustrates the process used to identify images for inclusion in the convolutional networks.



Source: Own authorship.

The first processing step consisted of reading the 3D MRI images in the ADNI1: Screening 1.5T database and then scanning coronal slices to identify hippocampal structures. Then, the images were resized to 256 pixels x 166 pixels, since MRI image dimensions are not entirely standardized.

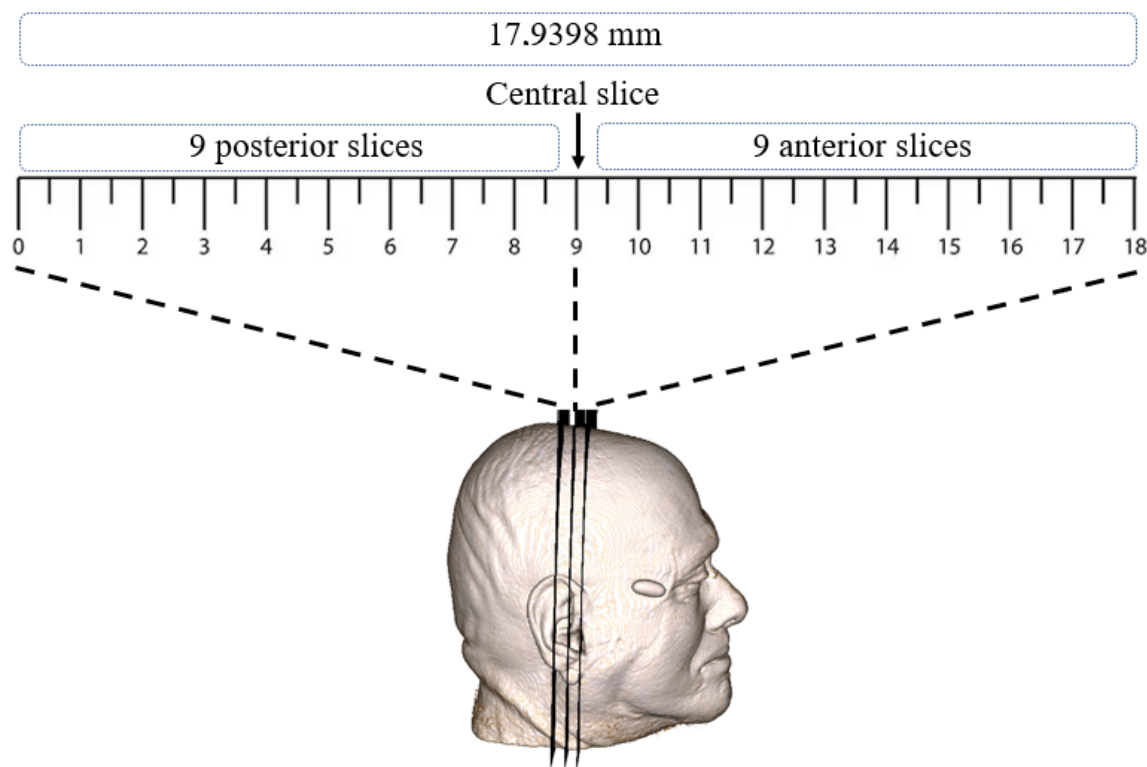
The next step was to frame the target area and eliminate unnecessary regions, especially in the lower part of the skull, favoring instead areas of the brain where hippocampal structures are located. This resulted in images of 155 pixels x 166 pixels.

Finally, the images were converted to 8-bit grayscale, resized to suit processing by convolutional networks at 176 pixels x 176 pixels, and saved to a disk. The conversion to 8 bits was done by rounding, with no substantial loss of quality, given that the original image has a low dynamic range. Therefore, the use of image quality preservation techniques was not evaluated as Tone Mapping (SALIH *et al.*, 2012).

Since convolutional networks need large volumes of data for training, a strategy was used to augment the input data. Figure 4.18 shows this strategy, where as many as nine additional

slices can be chosen in both the anterior and posterior directions from the central target slice, visually determined. This process yields up to 19 slices per patient MRI image, representing a total area of up to 17.9398 mm, since the slices are spaced 0.9442 mm apart.

Figure 4.18. Coronal slices with 0.9442 mm spacing. Central, anterior, and posterior slices are highlighted.



Source: Own authorship.

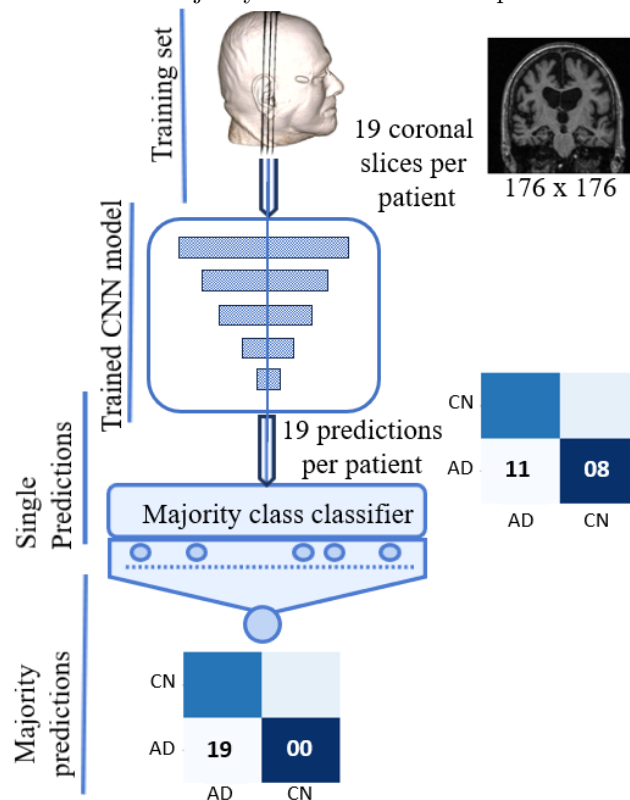
4.5.2 Multiple instance evaluation

The use of multiple instances per patient is justified by the necessary increase in data for the training stage of the network and also to make it possible to decide on the diagnosis by majority. As shown in Figure 4.6, the hippocampus has the longest extension in the direction measured from front to back of the individual's head, reaching up to 39 mm in the example measured. Therefore, for ADNI1: Screening 1.5T MRIs, up to 40 coronal *slices* can be selected to cover the entire length of the hippocampus. In this study, one central *slice* was chosen manually in the central region of the hippocampus, and another nine *slices* in the anterior and posterior directions, totaling 19 *slices* per patient. The strategy used in this study corresponds to using around 50% of the length of the hippocampus. The other 50% was left as a tolerance for the initial estimate of the central *slice* and the anatomical variations inherent in the diversity of

patients.

Figure 4.19 shows an example of a hypothetical patient where 19 instances were used. In this example, 11 instances were individually assessed as AD and 08 instances indicated CN. The majority decision gives the final result of AD for this patient.

Figure 4.19. Multiple instance decision scheme. In this case, a patient was evaluated and 11 instances indicated AD and 08 instances indicated CN. The majority decision is that the patient’s diagnosis is AD.



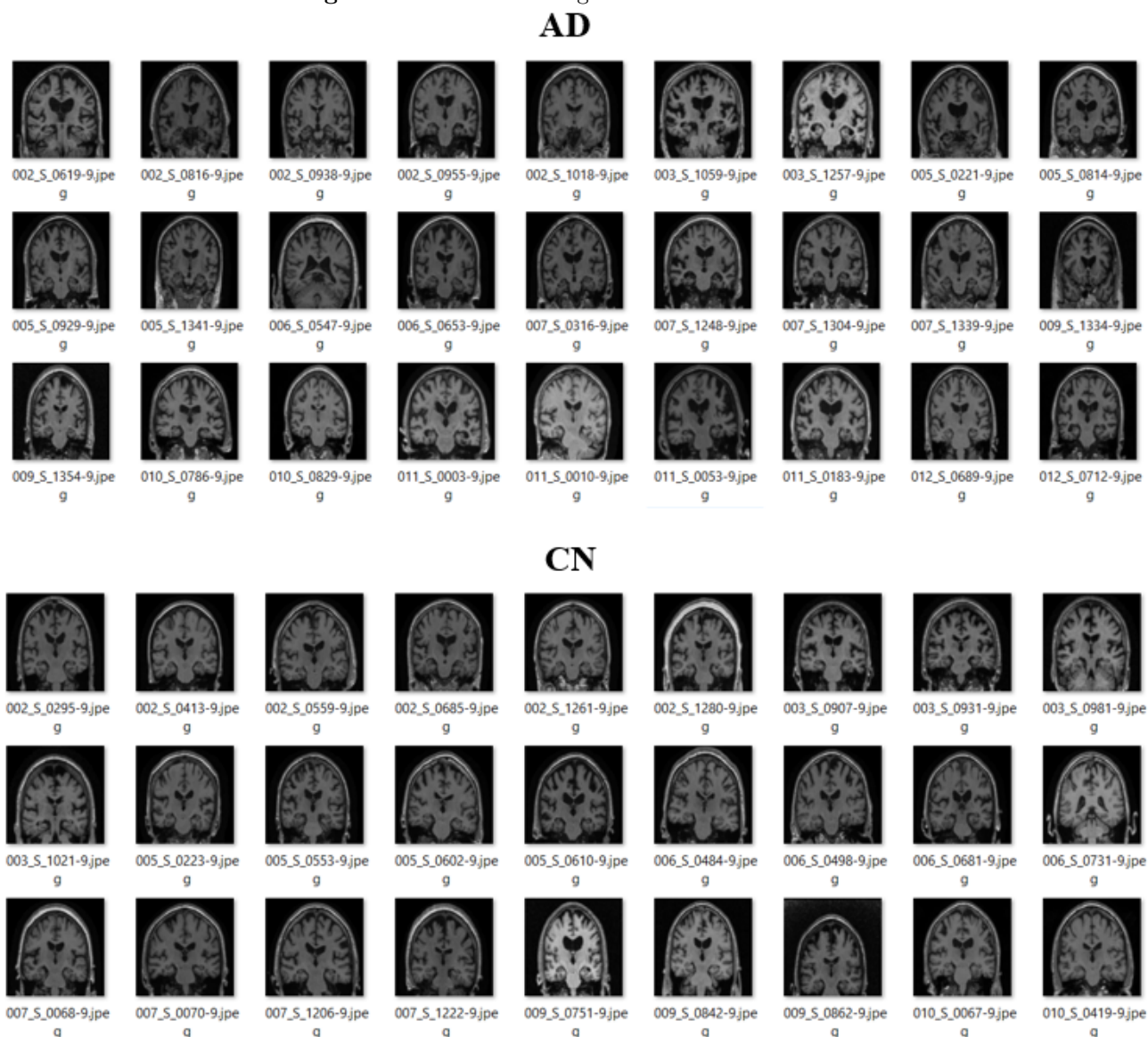
Source: Own authorship.

In addition, it is important to note that the majority decision helps to make the choice of the central *slice* more robust, considering that the preceding and following *slices* expand the spatial region of analysis. Another factor is the greater tolerance to a possible failure in the network’s inference, which can vary the diagnosis in individual *slices* of the same patient.

4.5.3 Resulting image dataset

An image dataset of central slices was generated for patients with CN and AD. A sample is shown in Figure 4.20.

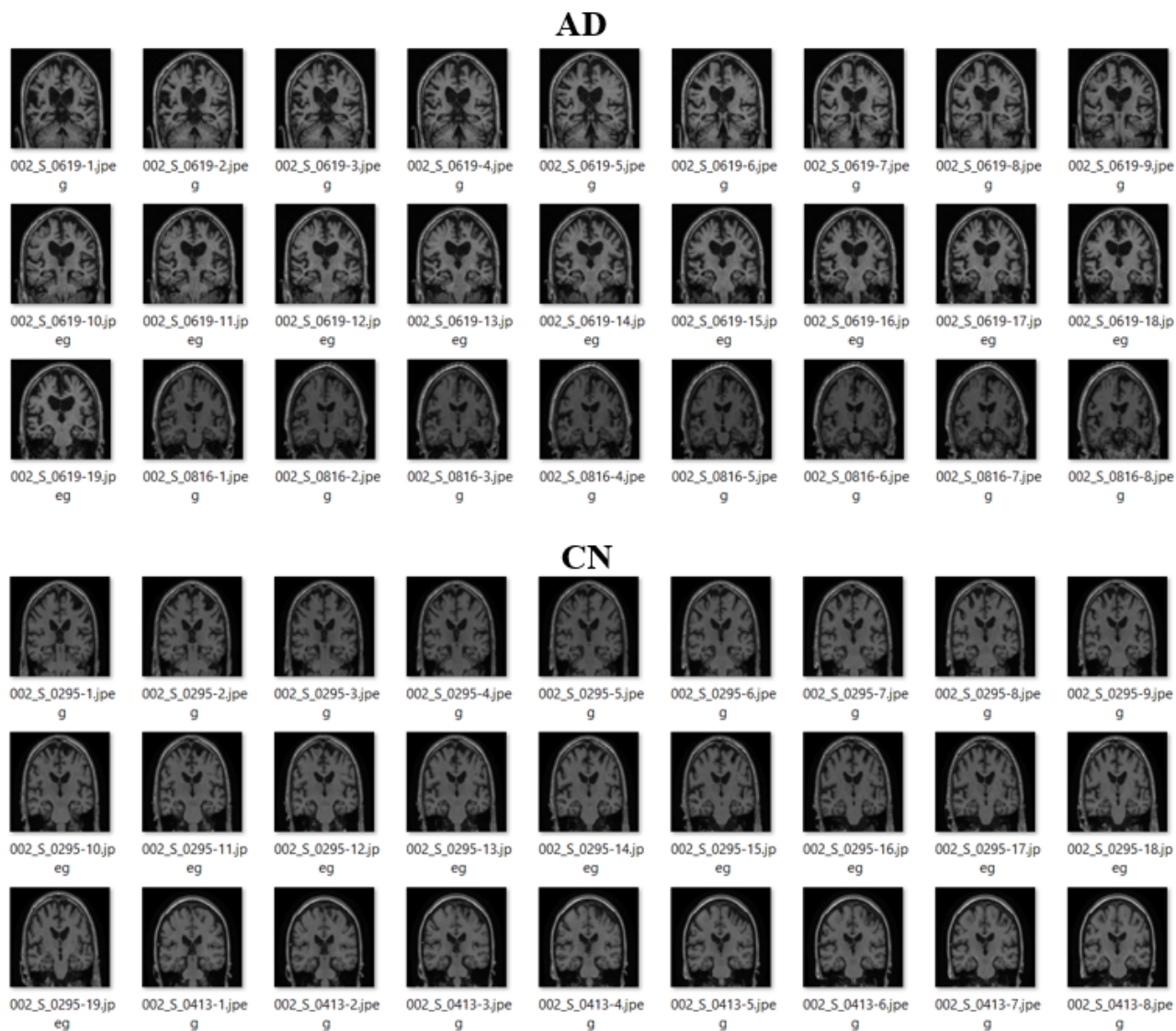
Figure 4.20. Coronal image dataset of central slices.



Source: Own authorship.

Not all the MRI images could be used due to missing data, errors in file reading or data inconsistency. Thus, from the original ADNI1: Screening 1.5T database (containing 307 CN patients, 525 MCI patients and 243 AD patients), images from 226 CN patients and 173 AD patients were generated, which was insufficient for training the convolutional network. Thus, augmenting the data with anterior and posterior slices was considered. Figure 4.21 shows a sample of the image dataset with the central slice, nine anterior slices and nine posterior slices.

Figure 4.21. Coronal dataset with a center slice, 9 anterior slices, and 9 posterior slices. The slice ending in -10 is the central slice chosen.



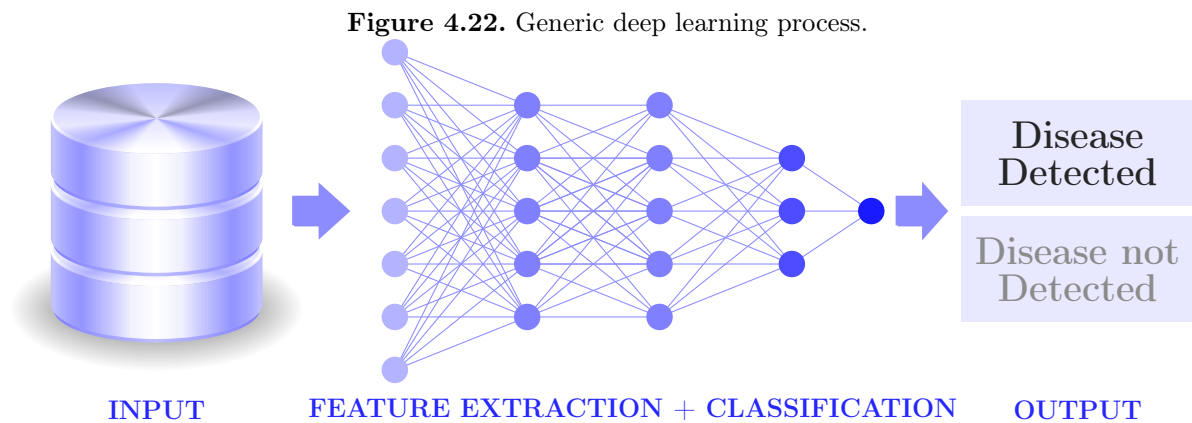
Source: Own authorship.

This strategy for augmenting the image dataset produced up to 4,294 CN images and 3,287 AD images, which was sufficient for convolutional network training.

4.6 DL MODELS

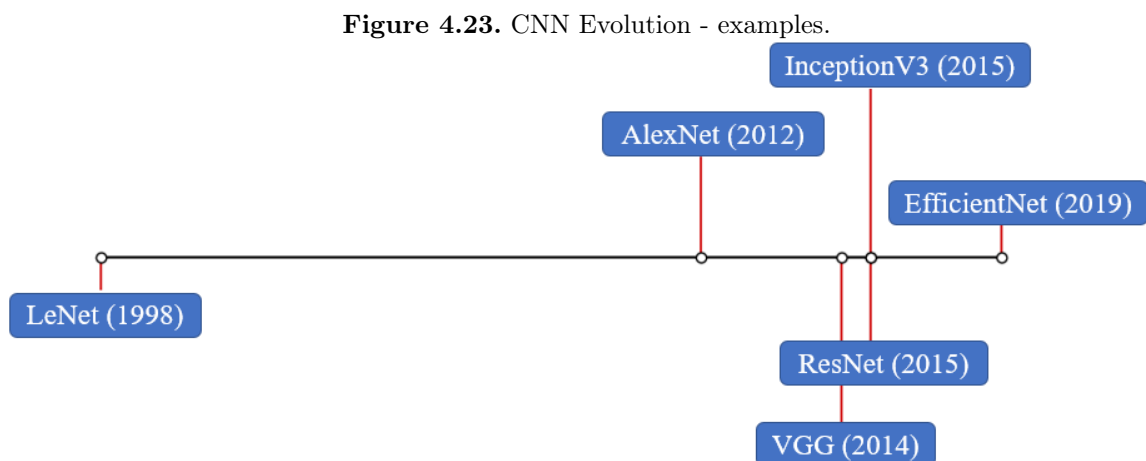
DL models consist of multiple layers of processing to learn data representations with multiple levels of abstraction. DL networks of the convolutional neural network (CNN) type improve the processing of images, video, speech and audio, while recurrent networks are particularly

effective with sequential data such as text and speech (LECUN *et al.*, 2015). Figure 4.22 shows a general scheme for processing DL models. Here, feature extraction is incorporated directly into the model and is performed within the first processing layers.



Source: Own authorship.

Typical CNN architectures stack a few convolutional layers, then a pooling layer, then a few more convolutional layers, then another pooling layer, and so on. Over the years, variants of this fundamental architecture have been developed, resulting in astonishing advances (YAMASHITA *et al.*, 2018). Figure 4.23 shows some examples of CNNs that have marked this evolution.



Source: Own authorship.

4.6.1 Classical CNN models

LeNet (LECUN *et al.*, 1998) was originally developed for handwritten number recognition and consists of two convolutional layers, followed by pooling, and then two fully connected

layers.

AlexNet (KRIZHEVSKY *et al.*, 2012) was the first winner of the “ImageNet” challenge. This process consists of eight layers (five convolutional layers and three other fully connected layers) and introduced the rectified linear unit (ReLU) activation function.

VGG (SIMONYAN; ZISSERMAN, 2014) was proposed by the Visual Geometry Group (VGG) team at the University of Oxford. The VGG-16 variant has 13 convolutional and three fully connected layers. The group also proposed VGG-19 - an even deeper variant, with 16 convolutional layers and three fully connected layers.

ResNet (HE *et al.*, 2015) incorporated the idea of “residual blocks”, which are connected to each other through jump connections. This model manages to go even deeper without compromising the generalization power of the model. ResNet-50 is an example of this strategy.

Instead of stacking convolutional layers, Inception stacks modules or blocks around convolutional layers. Inception V2 and V3 (SZEGEDY *et al.*, 2015) now use batch normalization.

EfficientNet (TAN; LE, 2019) uses a strategy to uniformly scale network width, depth, and resolution.

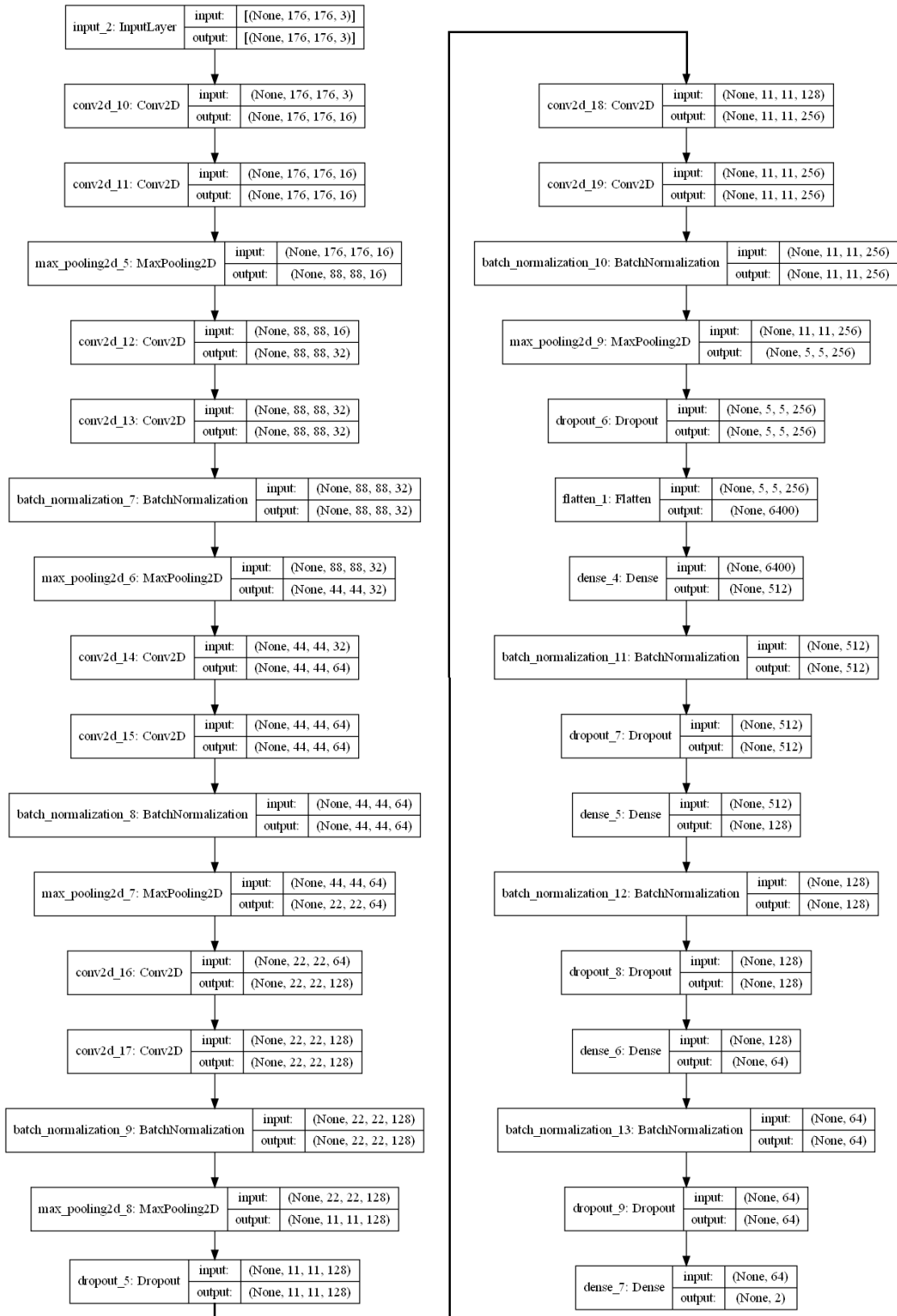
We used the VGG-16, VGG-19, ResNet-50 and InceptionV3 networks to evaluate the results in the image datasets, since these networks are commonly used for image classification.

4.6.2 Proposed CNN models

Five CNN models were also proposed (CNN-1, CNN-2, CNN-3, CNN-4 and CNN-5) and built from the following structural layers: convolutional, batch normalization, max pooling, dropout, flatten and dense.

Figure 4.24 shows the CNN-1 model.

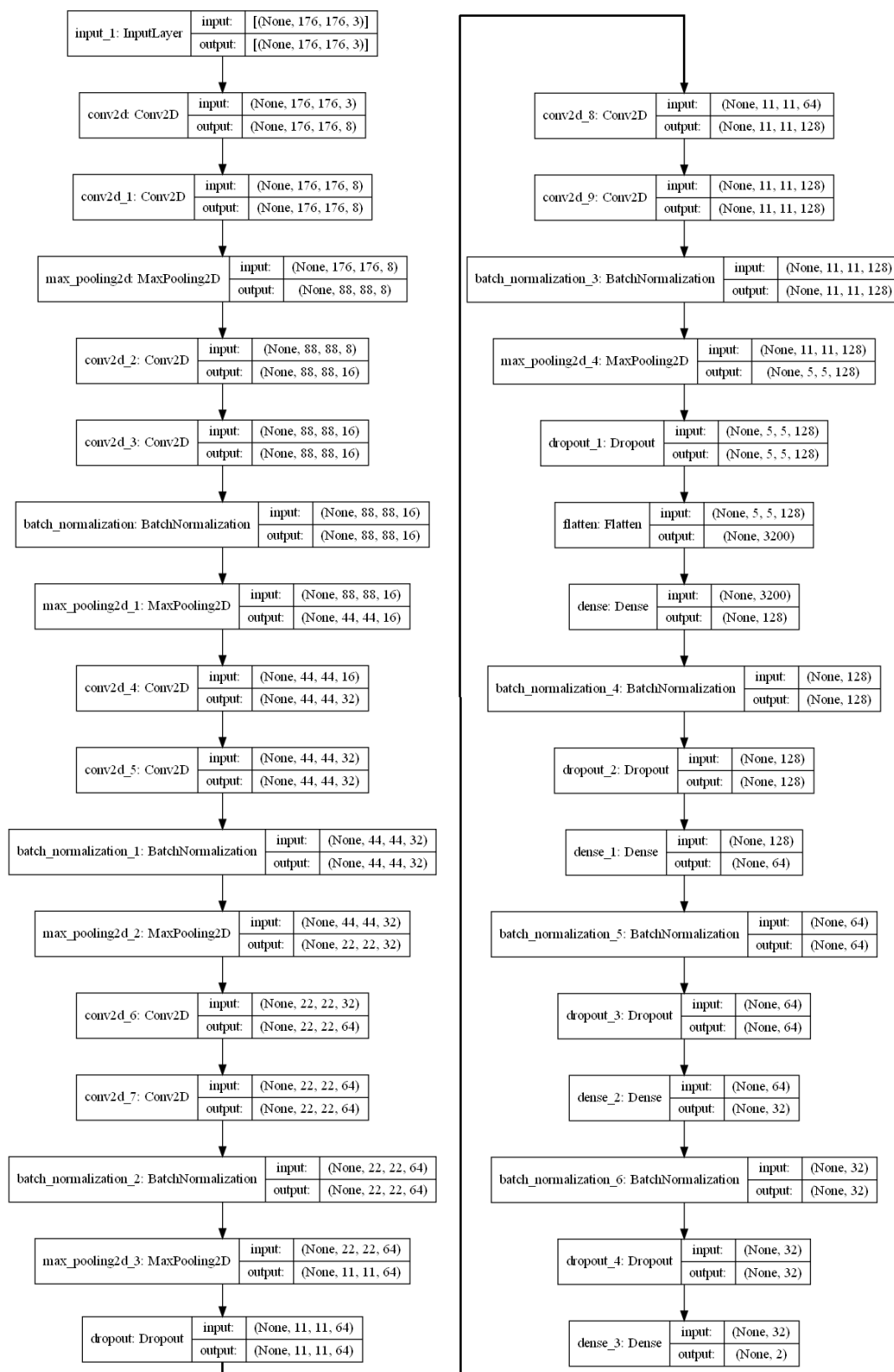
Figure 4.24. CNN 1 model - total parameters: 4,534,866. Initial model.



Source: Own authorship.

Figure 4.25 shows the CNN-2 model.

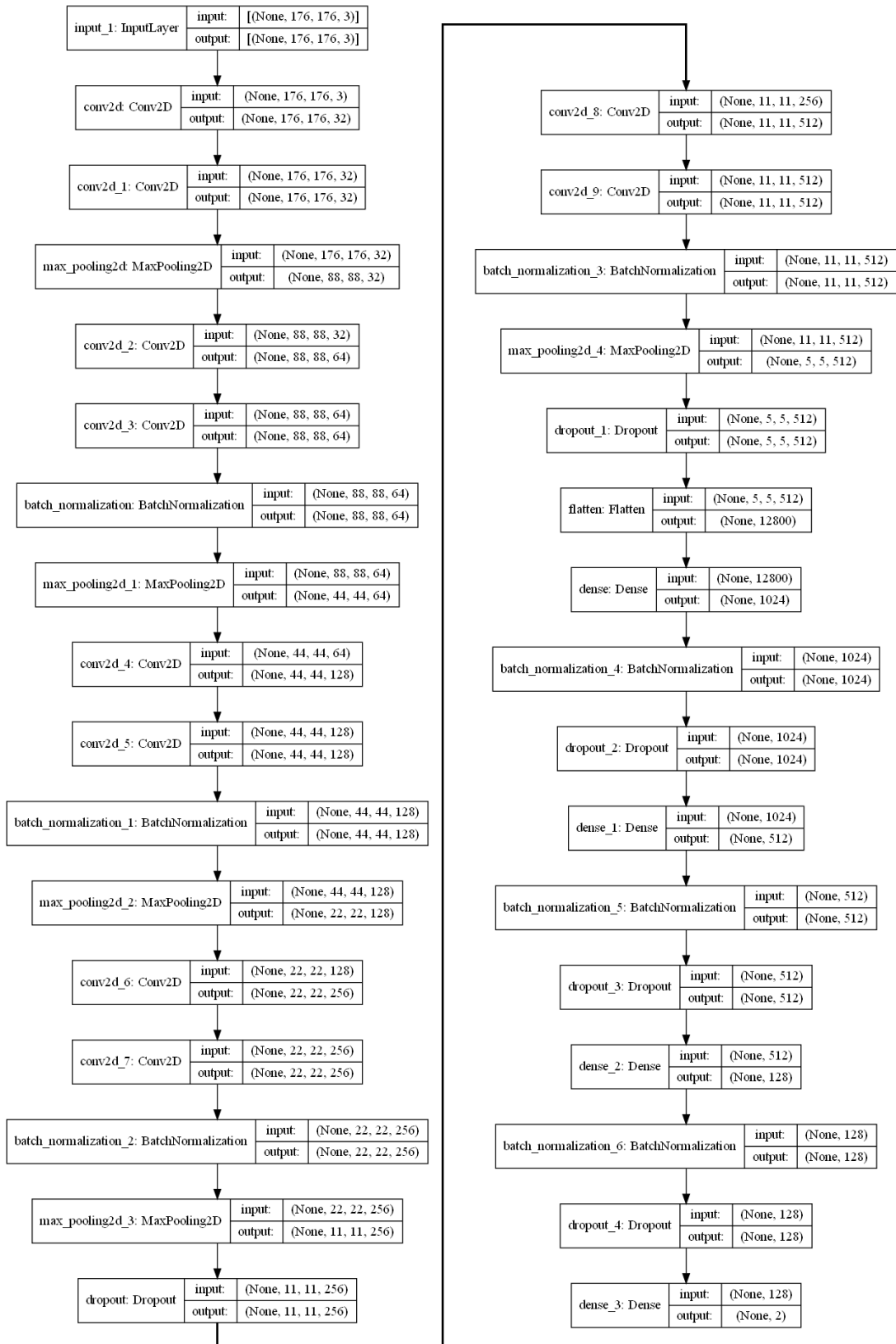
Figure 4.25. CNN 2 model - total parameters: 717,034. Reduced dimensionality with respect to CNN-1.



Source: Own authorship.

Figure 4.26 shows the CNN-3 model.

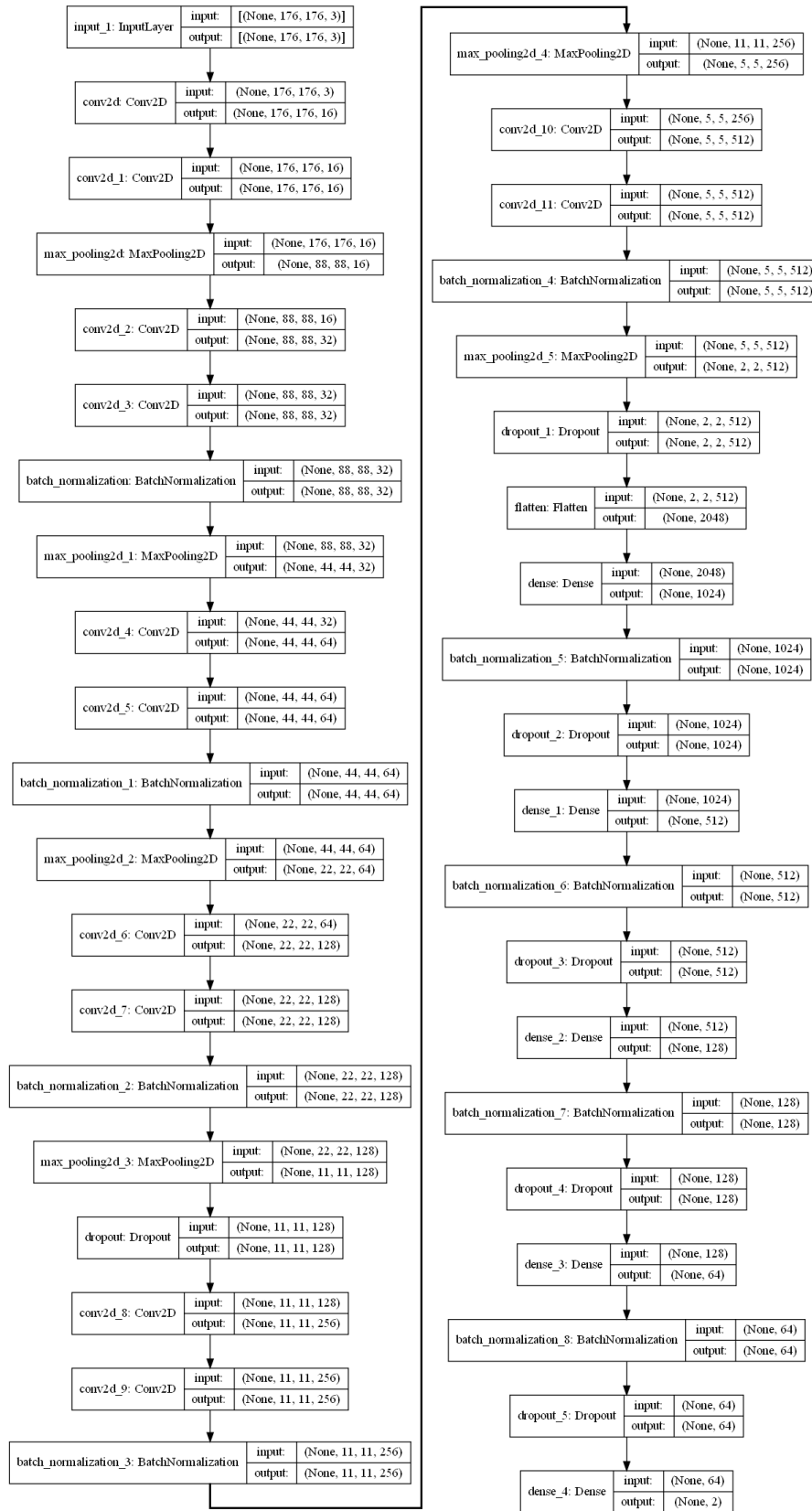
Figure 4.26. CNN 3 model - total parameters: 18,421,666. Increased dimensionality with respect to CNN-1.



Source: Own authorship.

Figure 4.27 shows the CNN-4 model.

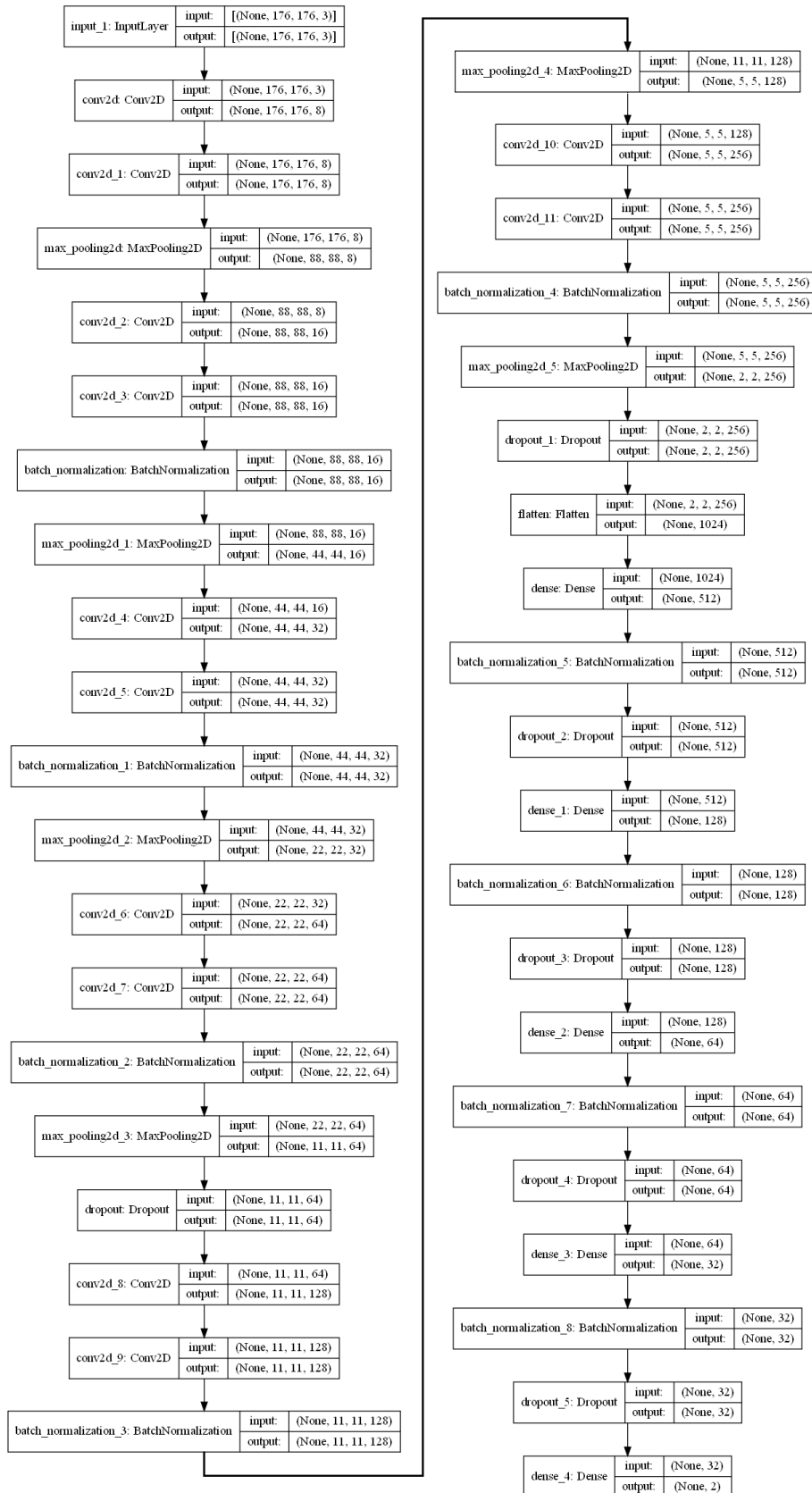
Figure 4.27. CNN 4 model - total parameters: 7,426,642. Increased depth compared to CNN-1.



Source: Own authorship.

Figure 4.28 shows the CNN-5 model.

Figure 4.28. CNN 5 model - total parameters: 1,786,090. Reduced dimensionality with respect to CNN-4.



Source: Own authorship.

CNN-1 consists of 10 convolutional layers, five max pooling layers, seven batch normalization layers, five dropout layers, one flatten layer, and four dense layers (Figure 4.19). This network contains 4,534,866 parameters, of which 4,532,498 are trainable parameters and 2,368 are non-trainable parameters.

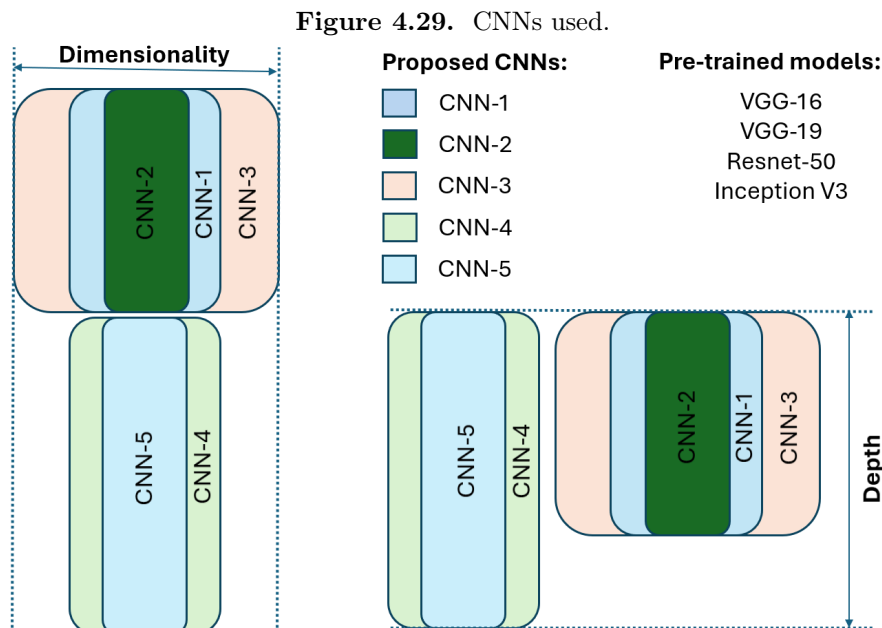
CNN-2 has the same types and quantities of layers as CNN-1 but with reduced dimensionality (Figure 4.20). Consequently, CNN-2 has 717,034 parameters, of which 716,106 are trainable and 928 non-trainable.

While CNN-3 also has the same types and quantities of layers as CNN-1, its layers have greater dimensionality (Figure 4.21). The network resulted in 18,421,666 parameters (18,416,418 trainable and 5,248 are non-trainable).

CNN-4 has more layers than does CNN-1, with two additional convolutional layers, a max pooling layer, two batch normalization layers, two dropout layers and a dense layer (Figure 4.22). The dimensionality of CNN-4 is the same as CNN-1, resulting in a network with 7,426,642 parameters (7,421,202 trainable and 5,440 non-trainable).

Finally, CNN-5 keeps the same number and layers as CNN-4, but with reduced dimensionality (Figure 4.23), resulting in 1,786,090 parameters (1,783,626 trainable and 2,464 non-trainable).

Figure 4.29 shows an overview of the CNNs used in this work.



Source: Own authorship.

This chapter presented the primary magnetic resonance dataset used, the treatment for setting up adequate datasets for assessing AD diagnosis and developed technological proposals, with emphasis on a strategy for augmenting available data. The next chapter will present results, a comparison of these results with those from other studies, and a discussion of the results from both clinical and imaging data.

RESULTS AND DISCUSSION

This chapter presents the results obtained from the clinical analysis and those obtained from the evaluation of magnetic resonance images. It also provides a discussion of the results and comparisons with the results achieved by other researchers.

5.1 CLINICAL DATA ANALYSIS RESULTS

The clinical data results were produced to provide a basis of comparison among the dataset classifiers, with and without feature selection.

The practical procedures were performed in Python version 3.7.3, with numpy libraries version 1.19.5, scikit-learn version 0.24.1, and using a computer with an Intel(R) Core(TM) i7-7500U CPU @ processor 2.70 GHz 2.90 GHz, with 16 GB of RAM.

The proposed feature selection algorithm (FMFS) was applied to each of the databases, resulting in reduced feature sets that were submitted to two classification algorithms: k-nearest neighbors (kNN) and support vector machines (SVM), and evaluated with the following metrics: balanced accuracy, precision, sensitivity, specificity, f1 score and multi-class area under the curve (MAUC). These procedures made it possible to compare the classification performance between the complete and reduced data sets, demonstrate the effect of the feature selection algorithm, and compare it with the results from other studies.

5.1.1 Feature selection in the AIBL database

The AIBL clinical database, described in Chapter 3 and originally consisting of 18 features, 785 instances and three classes (referred to as AIBL-18), was submitted to the proposed feature selection algorithm (FMFS) and optimized for the kNN and SVM classification models. The

results of these procedures are shown in Table 5.1

Table 5.1. Reduced AIBL dataset: selected features

Model	Dataset	Features
kNN	AIBL-6	CDGLOBAL BAT126 MMSCORE LIMMTOTAL LDELTOTAL APGEN1
SVM	AIBL-4	CDGLOBAL MMSCORE LIMMTOTAL LDELTOTAL

Source: Own authorship.

This produced two new datasets, AIBL-6 and AIBL-4, with six and four features, respectively, which will be evaluated in the following sections.

5.1.2 Feature selection in the ADNI database

The ADNI clinical database, described in Chapter 3 and originally consisting of 21 features, 1033 instances and three classes (referred to as ADNI-21), was submitted to the proposed feature selection algorithm (FMFS) and optimized for the kNN and SVM classification models. The results of these procedures are shown in Table 5.2

Table 5.2. Reduced ADNI dataset: selected features

Model	Dataset	Features
kNN	ADNI-3	CDRSB ADAS11 FAQ
SVM	ADNI-11	Ventricles Entorhinal CDRSB MMSE RAVLT_learning RAVLT_immediate RAVLT_perc_forgetting_bl ADAS11 ADAS13 FAQ APOE4

Source: Own authorship.

This resulted in two new datasets, ADNI-3 and ADNI-11, with three and 11 features, respectively, which will be evaluated in the following sections.

5.1.3 Feature selection in the OASIS database

The OASIS clinical database, described in Chapter 3 and originally consisting of seven features, 217 instances, three classes (referred to as OASIS-7), was submitted to the proposed feature selection algorithm (FMFS) and optimized for the kNN and SVM classification models.

The results are shown in Table 5.3.

Table 5.3. Reduced OASIS dataset: selected features

Model	Dataset	Features
kNN	OASIS-5	Age Educ SES MMSE nWBV
SVM	OASIS-5	Age Educ SES MMSE nWBV

Source: Own authorship.

The procedure resulted in a single new OASIS-5 dataset, with five equal features for both classification models kNN and SVM, which will be evaluated in the following sections.

5.1.4 Performance evaluation for three classes

Classification performance was evaluated for all datasets and results were obtained for the complete and reduced datasets. In this evaluation, the AIBL and ADNI databases were classified using the CN, MCI and AD classes, while the OASIS database was classified using the CDR 0, CDR 0.5 and CDR 12 classes.

5.1.4.1 ADNI evaluation

Table 5.4 summarizes the performance results obtained by evaluating the three classes in each ADNI dataset and for each classification model. A 95% confidence interval was used for the evaluation of the B_{ca} metric. Differences between the B_{ca} results for the complete and reduced datasets were also tested for statistical significance.

Under these conditions, the training times for the ADNI-3 and ADNI-11 datasets were 55% and 7% lower than for the complete databases.

With the kNN classifier, B_{ca} was 10.3% higher (significant at $p < 0,05$) for the ADNI-3 dataset than for the ADNI-21 dataset (Table 5.4).

Table 5.4. ADNI dataset results - three classes

Model	Dataset	$B_{ca} \pm IC$	A_{cy}	P_{rn}	S_{ny}	S_{py}	F1	MAUC
kNN	ADNI-21	81,66 \pm 3,39	76,29	79,34	76,89	87,23	77,41	0,878
	ADNI-3	90,06* \pm 2,26	89,45	90,32	86,52	93,60	87,97	0,960
SVM	ADNI-21	92,60 \pm 2,06	91,97	92,52	90,09	95,11	91,07	0,969
	ADNI-11	92,61 \pm 2,29	91,77	92,34	90,20	95,02	90,90	0,969

IC - confidence interval for 95% confidence level

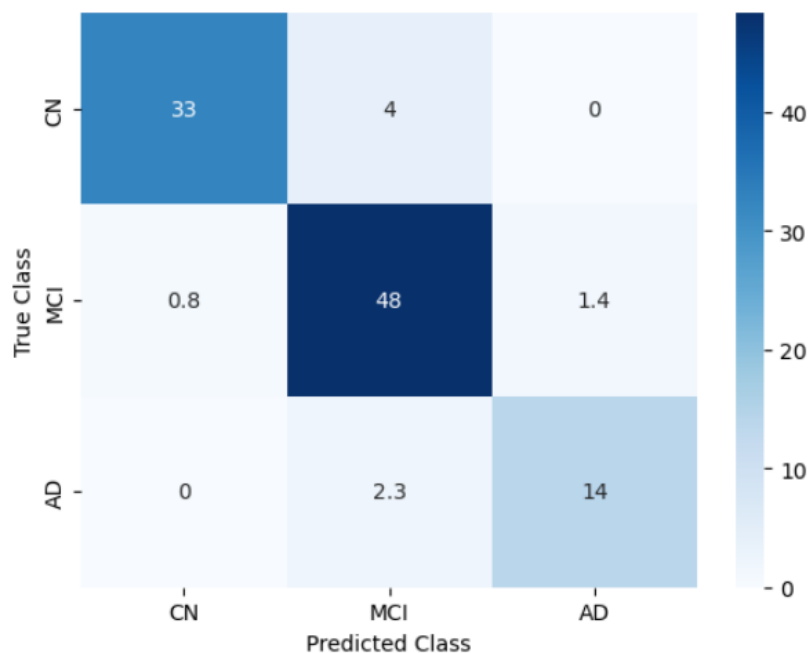
(*) Statistically significant result ($p < 0,05$)

Except for MAUC, every numerical result is expressed as a percentage.

Source: Own authorship.

The B_{ca} of the ADNI-11 dataset reached 92.61% with the SVM classifier, which was not statistically superior to that of the ADNI-21 dataset. Figure 5.1 shows the confusion matrix for this case.

Figure 5.1. ADNI-11 evaluated with the SVM classifier - average confusion matrix.



Source: Own authorship.

5.1.4.2 AIBL evaluation

Table 5.5 summarizes the performance results obtained by evaluating the three classes in each AIBL dataset, for each classification model. The B_{ca} metric was evaluated at the 95% confidence interval. Differences between the B_{ca} results for the complete and reduced datasets were also tested for statistical significance.

Consequent training times were 18% and 90% lower for AIBL-6 and AIBL-4, respectively, compared to the complete databases.

Table 5.5. AIBL dataset results - three classes

Model	Dataset	$B_{ca} \pm IC$	A_{cy}	P_{rn}	S_{ny}	S_{py}	F1	MAUC
kNN	AIBL-18	74,15 \pm 2,08	82,28	73,85	63,61	84,69	66,50	0,912
	AIBL-6	90,44* \pm 3,89	91,32	87,37	86,08	94,81	85,81	0,950
SVM	AIBL-18	91,96 \pm 2,25	91,72	86,97	88,32	95,61	87,09	0,967
	AIBL-4	92,52 \pm 2,50	92,22	88,25	89,06	95,98	87,74	0,959

IC - confidence interval for 95% confidence level

(*) Statistically significant result ($p < 0,05$)

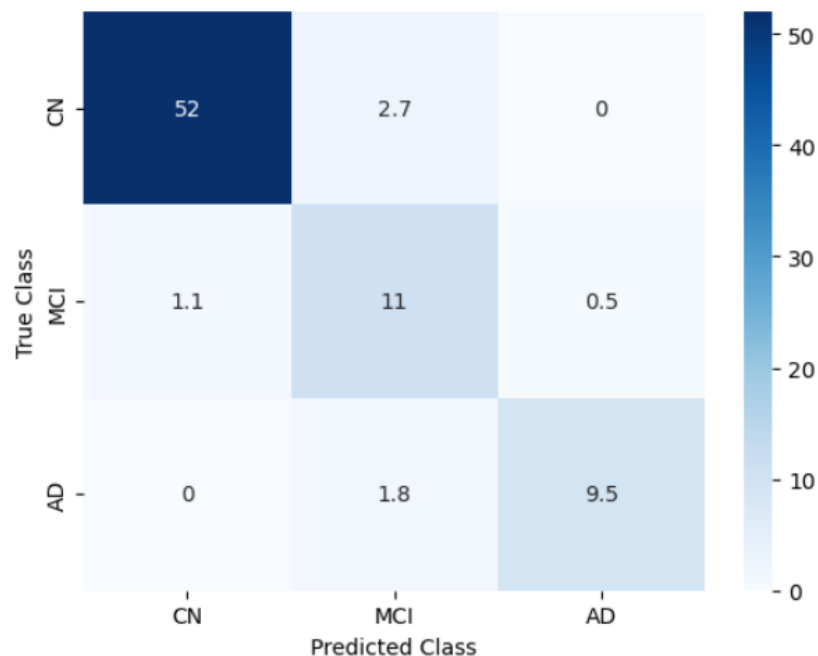
Except for MAUC, every numerical result is expressed as a percentage.

Source: Own authorship.

The B_{ca} metric was 22% superior (significant at $p < 0,05$) in the AIBL-6 dataset than in the AIBL-18 dataset, using the kNN classifier (Table 5.5).

The AIBL-4 dataset, evaluated by the SVM classifier, showed the best global performance. Figure 5.2 shows the confusion matrix for this case

Figure 5.2. AIBL-4 evaluated with the SVM classifier - average confusion matrix.



Source: Own authorship.

The B_{ca} metric of the AIBL-4 dataset (92.52%) was not statistically better than that of the AIBL-18 dataset when evaluated by the SVM classifier.

5.1.4.3 OASIS evaluation

Table 5.6 summarizes the performance results obtained by classifying each OASIS dataset into three classes, using each classification model. The 95% confidence interval was used for the B_{ca} metric.

Table 5.6. OASIS dataset results - three classes

Model	Dataset	$B_{ca} \pm IC$	A_{cy}	P_{rn}	S_{ny}	S_{py}	F1	MAUC
SVM	OASIS-7	72.13 \pm 4.90	72.29	66.13	61.95	82.31	59.09	0.858
	OASIS-5	72.66 \pm 4.46	73.64	70.85	61.75	83.57	58.33	0.871
kNN	OASIS-7	67.71 \pm 3.39	69.94	62.79	55.10	80.32	53.51	0.800
	OASIS-5	68.17 \pm 6.29	71.36	64.80	54.53	81.82	51.40	0.804

IC - confidence interval for 95% confidence level

Except for MAUC, every numerical result is expressed as a percentage.

Source: Own authorship.

The OASIS-5 dataset showed the best global performance when evaluated by the SVM classifier. Figure 5.3 shows the confusion matrix for this case.

Figure 5.3. OASIS-5 evaluated with the SVM classifier - average confusion matrix.



Source: Own authorship.

5.1.5 Performance evaluation for two classes

The ADNI-3, ADNI-11, AIBL-4 and AIBL-6 datasets were also evaluated by the SVM and kNN classifiers but arranged in two classes: CN v/s AD, MCI v/s AD and CN v/s MCI. Similarly, the OASIS-5 dataset was evaluated by the SVM and kNN classifiers with the following class combinations: C00 v/s C12, C05 v/s C12 and C00 v/s C05. Here, C00 represents the CDR 0 class, C05 represents CRD 0.5 and C12 represents the combination of CDR classes 1 and 2.

5.1.5.1 ADNI evaluation

Table 5.7 summarizes the performance results obtained by classifying each ADNI dataset into combinations of two classes, using each classification model.

Table 5.7. ADNI dataset results - two classes

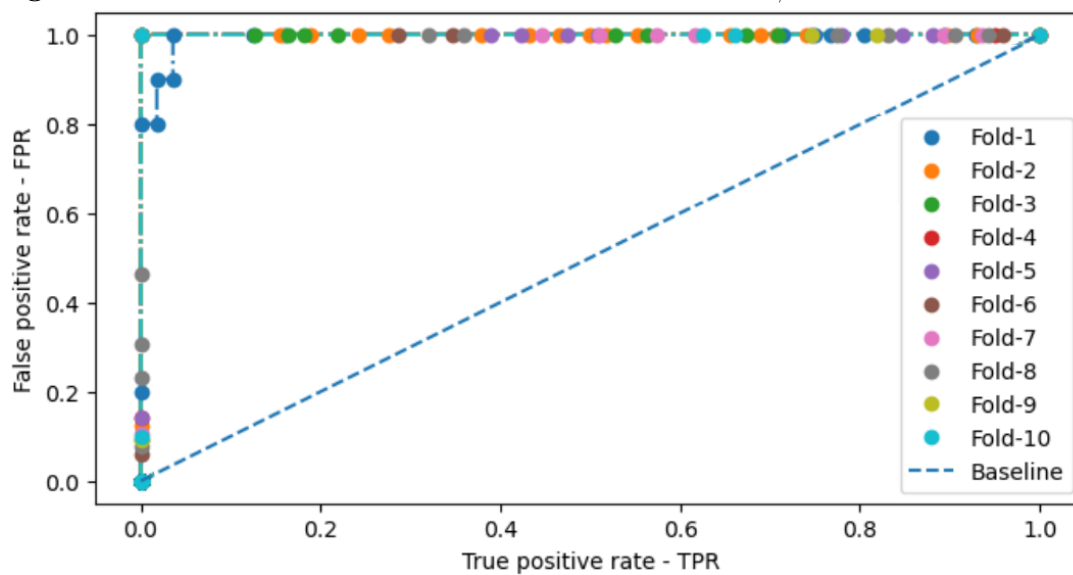
Model / Dataset	Type	B_{ca}	A_{cy}	P_{rn}	S_{ny}	S_{py}	F1	AUC
SVM / ADNI-11	CN v/s AD	99.88%	99.81%	99.55%	99.88%	99.88%	99.70%	1.000
	MCI v/s AD	91.12%	94.45%	92.93%	91.12%	91.12%	91.63%	0.986
	CN v/s MCI	93.98%	94.72%	95.33%	93.98%	93.98%	94.49%	0.975
kNN / ADNI-3	CN v/s AD	99.67%	99.81%	99.87%	99.67%	99.67%	99.76%	0.997
	MCI v/s AD	86.95%	91.89%	89.65%	86.95%	86.95%	87.82%	0.975
	CN v/s MCI	92.79%	93.58%	94.25%	92.79%	92.79%	93.28%	0.967

Source: Own authorship.

The best performance was achieved for the CN v/s AD combination in both datasets and using both combinations of classifiers. The B_{ca} metric reached 99.88% and mean AUC reached 1.000 (accurate to three decimal places).

Figure 5.4 shows the ROC curves for the CN v/s AD classes in the ADNI-11 dataset, evaluated by the SVM classifier. ROC curves (different colors) were calculated for each of the folds of the nested cross-validation process and show nearly ideal behavior.

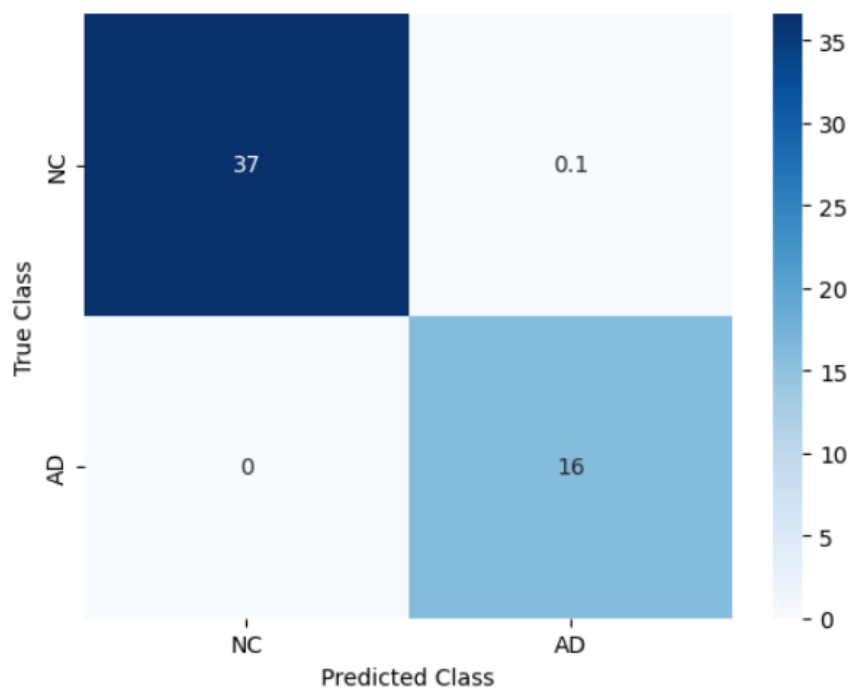
Figure 5.4. ADNI-11 avaliated with the SVM classifier over CN v/s AD classes - ROC curves.



Source: Own authorship.

Figure 5.5 shows the confusion matrix for the CN v/s AD evaluation (ADNI-11) using the SVM classifier.

Figure 5.5. ADNI-11 avaliated with the SVM classifier - CN v/s AD - average confusion matrix.



Source: Own authorship.

5.1.5.2 AIBL evaluation

Table 5.8 summarizes the performance obtained by classifying each AIBL dataset into combinations of two classes, using each classification model.

The best performance was achieved for the CN v/s AD combination in both datasets and for both combinations of classifiers, with a B_{ca} metric reaching 96.76%.

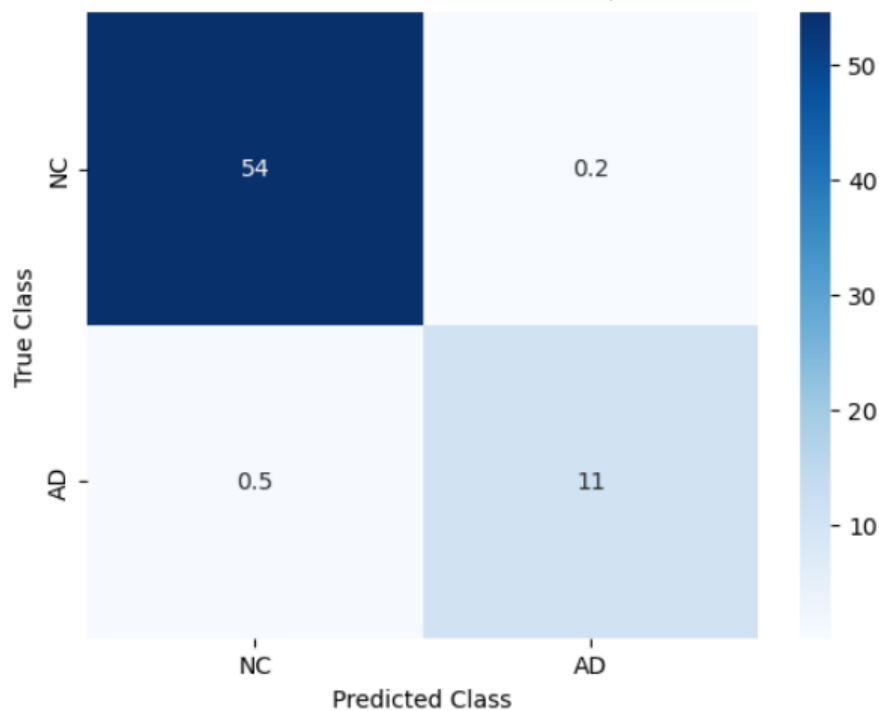
Table 5.8. AIBL dataset results - two classes

Model/ Dataset	Type	B_{ca}	A_{cy}	P_{rn}	S_{ny}	S_{py}	F1	AUC
SVM / AIBL-4	CN v/s AD	96.76%	98.94%	98.66%	96.76%	96.76%	97.55%	0.999
	MCI v/s AD	90.90%	91.16%	91.48%	90.90%	90.90%	90.82%	0.953
	CN v/s MCI	93.99%	94.51%	89.80%	93.99%	93.99%	91.55%	0.936
kNN / AIBL-6	CN v/s AD	95.80%	98.48%	97.67%	95.80%	95.80%	96.67%	0.986
	MCI v/s AD	87.69%	88.70%	89.83%	87.69%	87.69%	87.84%	0.910
	CN v/s MCI	93.72%	94.65%	90.19%	93.72%	93.72%	91.69%	0.941

Source: Own authorship.

Figure 5.6 shows the confusion matrix for the evaluation of the CN v/s AD class (AIBL-4) using the SVM classifier.

Figure 5.6. AIBL-4 evaluated with the SVM classifier - CN v/s AD - average confusion matrix.



Source: Own authorship.

5.1.5.3 OASIS evaluation

Table 5.9 summarizes the performance results obtained by classifying each OASIS dataset into two classes, using each classification model.

The best performance was achieved for the C00 v/s C12 classes in both datasets and in both combinations of classifiers. Here, the B_{ca} metric reached 89.17%.

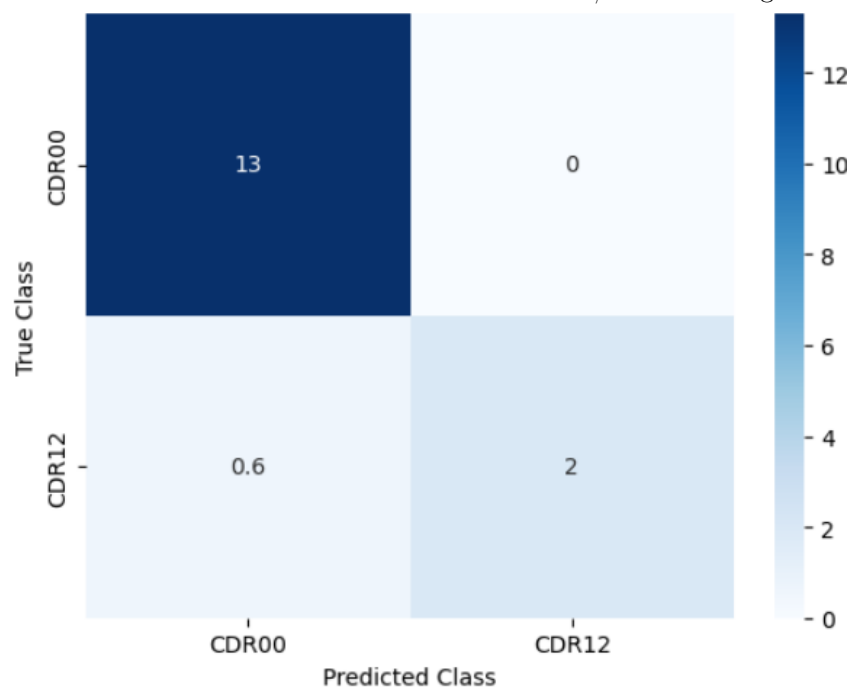
Table 5.9. OASIS dataset results - two classes

Model / Dataset	Type	B_{ca}	A_{cy}	P_{rn}	S_{ny}	S_{py}	F1	AUC
SVM / OASIS-5	C00 v/s C12	88.04%	94.29%	92.85%	88.04%	88.04%	88.66%	0.978
	C05 v/s C12	65.82%	72.50%	71.70%	65.82%	65.82%	63.54%	0.783
	C00 v/s C05	72.77%	81.58%	81.59%	72.77%	72.77%	72.71%	0.846
kNN / OASIS-5	C00 v/s C12	89.17%	96.17%	97.93%	89.17%	89.17%	91.23%	0.919
	C05 v/s C12	63.82%	71.25%	63.50%	63.82%	63.82%	62.18%	0.733
	C00 v/s C05	65.85%	78.42%	75.77%	65.85%	65.85%	64.92%	0.778

Source: Own authorship.

Figure 5.7 shows the confusion matrix for the evaluation of C00 v/s C12 in the OASIS-5 dataset using the kNN classifier.

Figure 5.7. OASIS-5 avalidated with the kNN classifier - C00 v/s C12 - average confusion matrix.



Source: Own authorship.

5.1.6 Performance evaluation with other classifiers

The datasets that performed best in the previous evaluations (ADNI-11, AIBL-4, and OASIS-5) were also evaluated using other classifiers and the following metrics: balanced accuracy, precision, sensitivity, specificity, f1 score and the multi-class area under the curve (MAUC). In this evaluation, the AIBL-4 and ADNI-11 datasets were classified into CN, MCI and AD classes, while the OASIS-5 dataset was classified into the CDR 0, CDR 0.5, and CDR 12 classes.

In addition to the SVM and kNN classifiers, evaluations were also performed with the Random Forest, Logistic Regression, Naive Bayes, Decision Tree, Ada Boost, Linear Discriminant Analysis (LDA), Gradient Boosting, Extreme Gradient Boosting and Histogram Gradient Boosting classifiers. As in the other cases, the assessments were performed using nested cross-validation.

Table 5.10 shows that the Gradient Boosting classifier achieved the best performance with the AIBL-4 dataset, reaching a B_{ca} of 92.88%.

Table 5.10. AIBL-4 - performance evaluation

Model	B_{ca}	A_{cy}	P_{rn}	S_{yn}	S_{py}	F1	MAUC
SVM	92,52%	92,22%	88,25%	89,06%	95,98%	87,74%	0,959
kNN	91.21%	91.46%	87.02%	87.12%	95.30%	86.23%	0.960
Random Forest	87.79%	89.29%	84.26%	82.77%	92.82%	82.83%	0.958
Logistic Regression	91.60%	91.71%	88.01%	88.03%	95.17%	87.28%	0.975
Naive Bayes	91.80%	91.72%	87.72%	88.08%	95.52%	87.10%	0.974
Decision Tree	91.96%	91.46%	87.73%	88.36%	95.57%	87.03%	0.959
Ada Boost	84.15%	87.77%	80.85%	77.22%	91.08%	78.27%	0.949
LDA	88.15%	90.18%	86.47%	82.93%	93.37%	83.86%	0.974
Gradient Boosting	92.88%	92.61%	88.94%	89.63%	96.13%	88.60%	0.974
Extreme GB	91.62%	91.96%	88.38%	87.87%	95.36%	87.38%	0.967
Histogram GB	87.97%	89.54%	84.38%	82.71%	93.24%	82.69%	0.961

Source: Own authorship.

Table 5.11 presents performance of various classifiers with the ADNI-11 dataset. These results show that the SVM classifier performed best, reaching a B_{ca} value of 92.61%.

Table 5.11. ADNI-11 - performance evaluation

Model	B_{ca}	A_{cy}	P_{rn}	S_{yn}	S_{py}	F1	MAUC
SVM	92,61%	91,77%	92,34%	90,20%	95,02%	90,90%	0,969
kNN	83.56%	78.13%	80.61%	79.54%	87.58%	79.57%	0.891
Random Forest	92.36%	91.39%	91.26%	89.83%	94.90%	90.34%	0.968
Logistic Regression	92.46%	91.10%	91.21%	90.14%	94.78%	90.54%	0.973
Naive Bayes	88.42%	83.17%	82.39%	85.71%	91.13%	83.35%	0.933
Decision Tree	91.85%	90.90%	91.12%	89.11%	94.59%	89.86%	0.952
Ada Boost	90.10%	88.10%	87.76%	87.04%	93.16%	86.84%	0.939
LDA	85.20%	80.94%	82.93%	81.33%	89.07%	81.84%	0.918
Gradient Boosting	92.05%	91.49%	92.33%	89.28%	94.81%	90.44%	0.967
Extreme GB	91.29%	90.23%	90.71%	88.42%	94.17%	89.33%	0.966
Histogram GB	92.00%	90.71%	90.65%	89.41%	94.60%	89.79%	0.963

Source: Own authorship.

Table 5.12 presents the performance of various classifiers with the OASIS-5 dataset. Here, the Decision Tree classifier provided the best performance with a B_{ca} of 76.25%.

Table 5.12. OASIS-5 - performance evaluation

Model	B_{ca}	A_{cy}	P_{rn}	S_{yn}	S_{py}	F1	MAUC
SVM	72.66%	73.64%	70.85%	61.75%	83.57%	58.33%	0.872
kNN	68.17%	71.36%	64.80%	54.53%	81.82%	51.40%	0.804
Random Forest	75.30%	76.41%	72.49%	65.05%	85.56%	62.50%	0.881
Logistic Regression	73.74%	74.55%	72.58%	63.10%	84.39%	60.34%	0.876
Naive Bayes	73.86%	75.00%	70.04%	63.05%	84.66%	62.38%	0.876
Decision Tree	76.25%	74.57%	67.98%	68.10%	84.40%	64.92%	0.842
Ada Boost	65.58%	64.20%	53.02%	50.71%	80.46%	49.48%	0.811
LDA	73.87%	75.00%	70.88%	64.34%	83.41%	63.00%	0.868
Gradient Boosting	73.54%	74.98%	71.42%	62.95%	84.13%	61.66%	0.890
Extreme GB	73.49%	75.02%	68.72%	62.67%	84.31%	61.57%	0.863
Histogram GB	70.36%	70.30%	63.24%	57.39%	83.34%	55.13%	0.881

Source: Own authorship.

5.1.7 Evaluation with other feature selection methods

The ADNI, AIBL and OASIS datasets were also subjected to other feature selection methods, from the three groups of techniques referenced in Chapter 3. As with the FMFS algorithm, each of the techniques was evaluated for the kNN and SVM classifiers, and the number of features that resulted in the best accuracy was found.

Table 5.13 shows the value found for the number of optimal features and accuracy, evaluated on the ADNI dataset, for each of the classifiers.

Table 5.13. ADNI - feature selection method evaluation

Feature selection method	kNN		SVM	
	Features	B _{ca}	Features	B _{ca}
FMFS	3	90.06%	11	92.61%
Chi-square	2	89.68%	17	92.81%
Mutual Information	1	90.31%	5	92.93%
Anova F-value	1	90.31%	4	92.93%
Variance Threshold	6	61.50%	11	74.66%
Fisher Score	13	83.82%	17	92.62%
MultiSURF	6	86.70%	6	92.07%
RFE (SVM)	1	90.31%	3	92.84%
Permutation Importance (LR)	2	91.72%	4	92.93%
SHAP (XGboost)	2	91.72%	3	92.97%
Boruta (RF)	3	90.54%	5	92.93%
Embedded Random Forest	1	90.31%	4	92.93%
Embedded LightGBM	1	90.31%	5	92.83%

Source: Own authorship.

Table 5.14 shows the value found for the number of optimal features and accuracy, evaluated on the AIBL dataset, for each of the classifiers.

Table 5.14. AIBL - feature selection method evaluation

Feature selection method	kNN		SVM	
	Features	B _{ca}	Features	B _{ca}
FMFS	6	90.44%	4	92.52%
Chi-square	4	91.21%	4	92.52%
Mutual Information	1	88.30%	5	92.43%
Anova F-value	1	88.30%	7	92.27%
Variance Threshold	1	50.25%	10	80.90%
Fisher Score	1	88.30%	12	92.11%
MultiSURF	4	87.66%	11	92.93%
RFE (SVM)	3	91.86%	6	92.54%
Permutation Importance (LR)	1	88.30%	12	92.29%
SHAP (XGboost)	1	50.25%	9	92.49%
Boruta (RF)	2	90.82%	4	92.52%
Embedded Random Forest	1	88.30%	3	92.79%
Embedded LightGBM	1	88.30%	7	92.76%

Source: Own authorship.

Table 5.15 shows the value found for the number of optimal features and accuracy, evaluated on the OASIS dataset, for each of the classifiers.

Table 5.15. OASIS - feature selection method evaluation

Feature selection method	kNN		SVM	
	Features	B _{ca}	Features	B _{ca}
FMFS	5	68.17%	5	72.66%
Chi-square	2	77.59%	6	72.14%
Mutual Information	2	77.59%	4	72.74%
Anova F-value	2	77.59%	3	75.30%
Variance Threshold	1	50.72%	4	70.03%
Fisher Score	5	67.63%	7	72.13%
MultiSURF	1	72.91%	6	71.97%
RFE (SVM)	1	72.91%	6	71.97%
Permutation Importance (LR)	1	72.91%	4	72.57%
SHAP (XGboost)	3	74.60%	7	72.13%
Boruta (RF)	2	77.59%	3	75.30%
Embedded Random Forest	2	77.59%	3	75.30%
Embedded LightGBM	4	68.14%	5	70.22%

Source: Own authorship.

5.2 IMAGE ANALYSIS RESULTS

The practical image analysis procedures were performed in Python version 3.9.0, with numpy libraries version 1.19.2, scikit-learn version 1.1.1, tensorflow version 2.6.0, cuda toolkit version 11.3.1 and a 12th Gen Intel(R) Core(TM) i7-12700H 2.30 GHz processor, with 16 GB of RAM. An NVIDIA GeForce RTX 3060 Graphics Processing Unit (GPU) card with 6 GB of RAM was also used.

The GPU was especially helpful with DL operations, greatly reducing processing times and increasing productivity.

Performance was evaluated using the following metrics: balanced accuracy, accuracy, precision, sensitivity, specificity, f1 score and area under the curve (AUC). Finally, classification performance was compared among the generated image datasets and with results obtained in other studies.

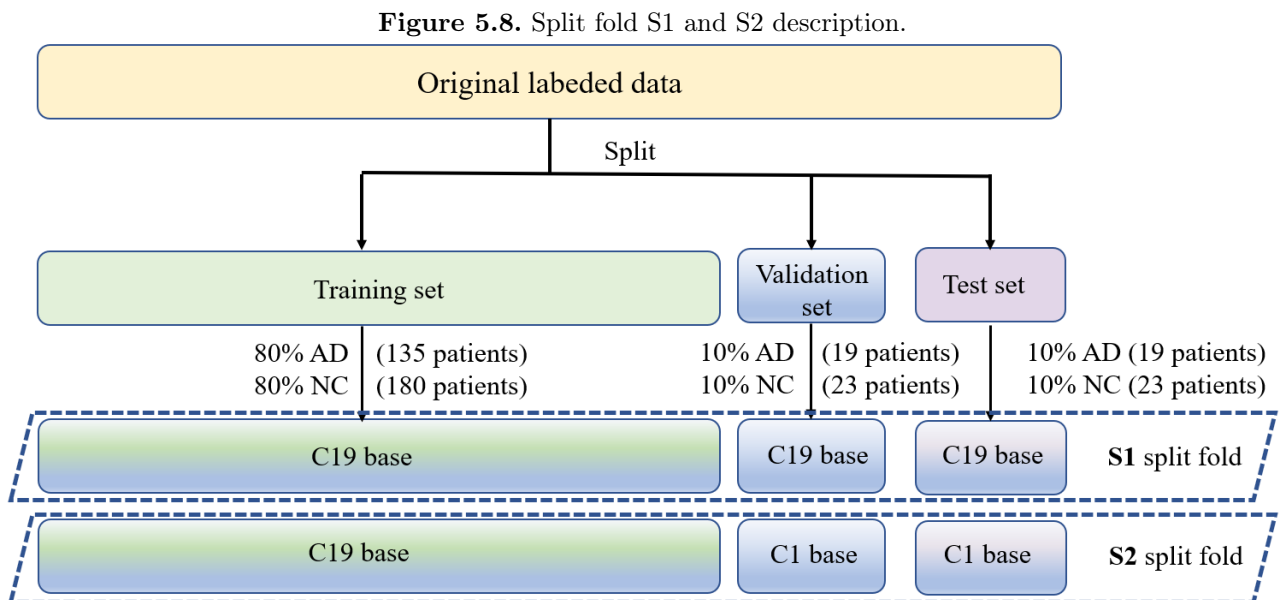
5.2.1 Analysis of the coronal image dataset

As described in Chapter 4, two versions of the coronal image dataset were generated, one consisting of central slices (Figure 4.20) - dataset C1, and another with central slices plus nine

anterior and nine posterior slices (Figure 4.21) - dataset C19. The C1 dataset was very small, consisting of images generated from 226 CN patients and 173 AD patients. The limited size of this dataset made training convolutional networks unfeasible. The C19 dataset consisted of 4,294 CN images and 3,287 AD images. These additional images are anterior and posterior slices that augment the dataset and enable convolutional network training.

5.2.1.1 Partition setup alternatives

As mentioned in Chapter 3, CNNs need to be trained with a set of three partitions (Figure 3.14) that eliminate the risk of data leakage between training and test partitions. Thus, the C1 and C19 image datasets were used to build partitions S1 and S2 (Figure 5.8).



The S1 partition consisted of images from 315 patients (C19 dataset) for the training set, images from 42 other patients (C19) for the validation set, and images from an additional 42 patients (C19) for the test set.

The S2 partition also consisted of images from 315 patients (C19 dataset) for the training set, images from 42 other patients (C1) for the validation set, and images from another 42 patients (C1) for the test set.

Table 5.16 shows the final compositions of the partitions.

Table 5.16. Split fold composition S1 and S2

Split fold	Train	Validation	Test
S1	2.565 AD images	361 AD images	361 AD images
	3.420 CN images	437 CN images	437 CN images
S2	2.565 AD images	19 AD images	19 AD images
	3.420 CN images	23 CN images	23 CN images

Source: Own authorship.

5.2.1.2 Evaluation of the S1 partition

Performance of the S1 partition was measured using the proposed DL models (CNN-1, CNN-2, CNN-3, CNN-4 and CNN-5), and the pre-trained models (VGG-19, VGG-16, ResNet-50 and Inception V3). The results can be seen in Table 5.17.

Table 5.17. S1 split fold - performance evaluation

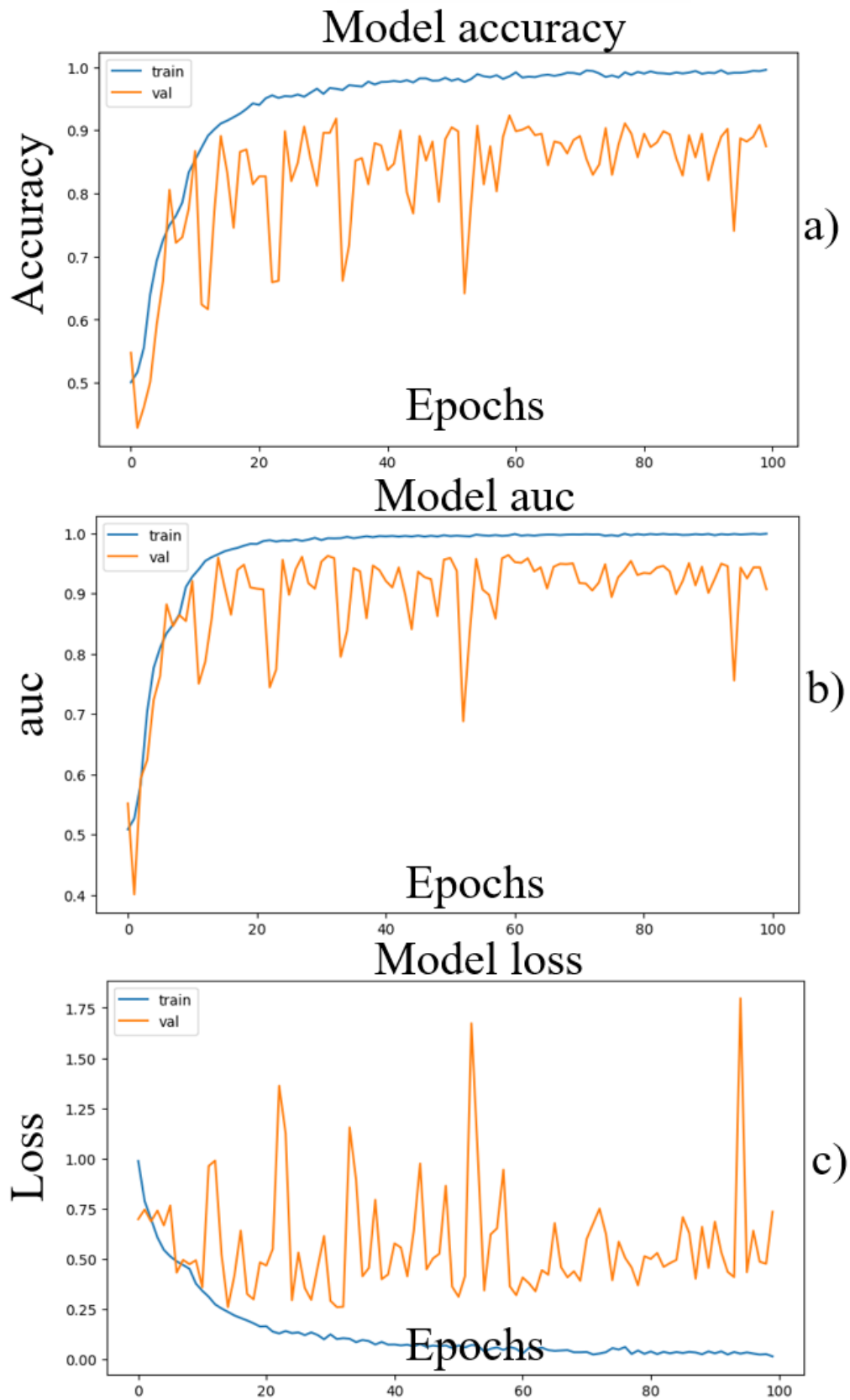
Model	B_{ca}	A_{cy}	P_{rn}	S_{yn}	S_{py}	F1	AUC
CNN-1	94.25%	94.36%	94.34%	95.42%	93.07%	94.88%	0.942
CNN-2	90.50%	90.23%	94.10%	87.64%	93.35%	90.76%	0.905
CNN-3	83.20%	84.71%	78.58%	99.08%	67.31%	87.65%	0.832
CNN-4	93.62%	93.36%	96.83%	90.85%	96.40%	93.74%	0.936
CNN-5	94.79%	94.74%	96.04%	94.28%	95.29%	95.15%	0.948
VGG-19	86.22%	86.59%	86.03%	90.16%	82.27%	88.04%	0.862
VGG-16	84.09%	84.71%	83.02%	90.62%	77.56%	86.65%	0.841
Resnet-50	83.13%	84.21%	80.19%	94.51%	71.75%	86.76%	0.831
Inception V3	81.33%	81.95%	80.84%	87.87%	74.79%	84.21%	0.813

Source: Own authorship.

In the training phase for this and other cases, all models were trained using the *Adam* optimizer with a learning rate of 0.001, 100 epochs and a batch size set to 32 based on the GPU’s memory (6 GB).

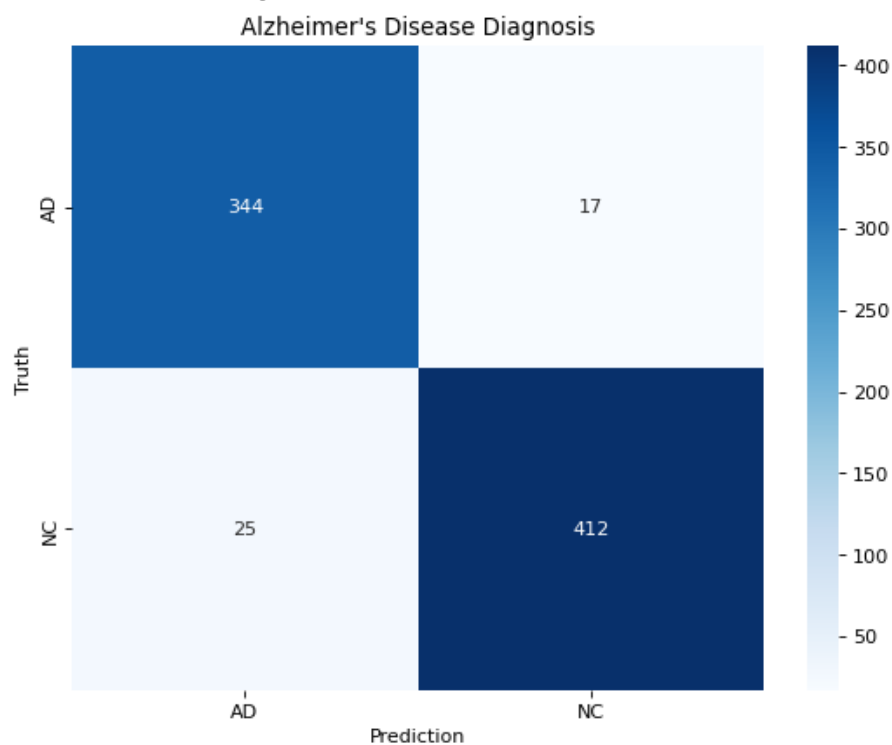
The pre-trained models, VGG-19, VGG-16, ResNet-50 and Inception V3, were initialized with weights from the training carried out on *dataset* ImageNet and had the first layers frozen for training, up to the penultimate convolutional layer. The proposed models, CNN-1, CNN-2, CNN-3, CNN-4 and CNN-5, were trained in their entirety.

A learning curve was calculated for CNN-5 and evaluated over 100 epochs (Figure 5.9).

Figure 5.9. CNN-5 model learning curves. a) Accuracy model. b) AUC model. c) Loss model

Source: Own authorship.

Additionally, a confusion matrix was also evaluated for CNN-5 (Figure [5.10](#)).

Figure 5.10. CNN-5 confusion matrix.

Source: Own authorship.

5.2.1.3 Evaluation of the S2 partition

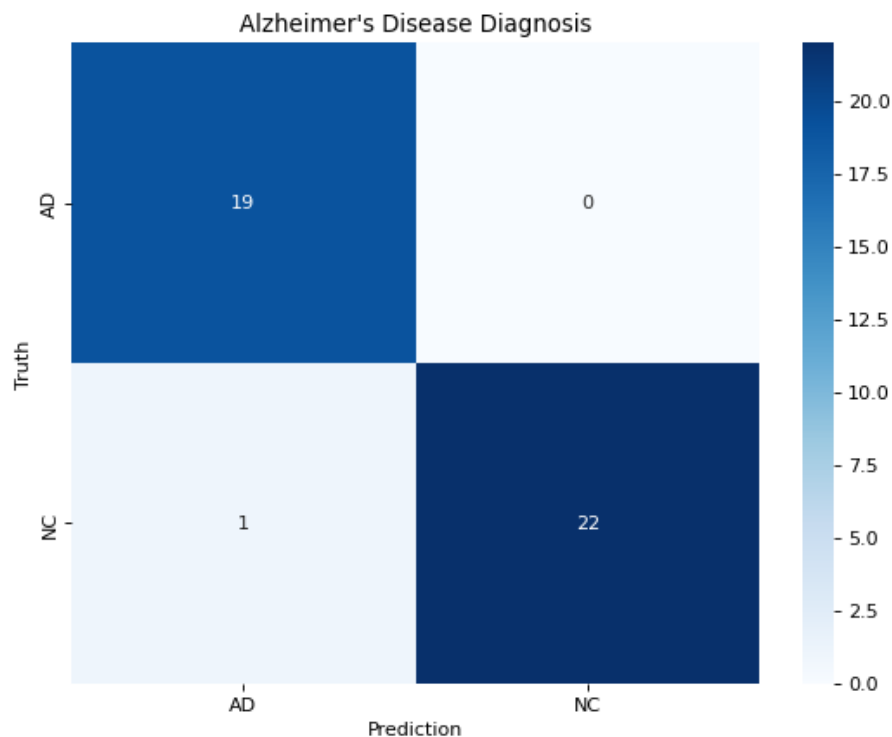
As in the S1 partition, S2 performance was measured using the proposed DL models (CNN-1, CNN-2, CNN-3, CNN-4 and CNN-5). The results can be seen in Table [5.18](#).

Table 5.18. S2 split fold - performance evaluation

Model	B_{ca}	A_{cy}	P_{rn}	S_{yn}	S_{py}	F1	AUC
CNN-1	95.65%	95.24%	100.0%	91.30%	100.0%	95.45%	0.957
CNN-2	91.30%	90.48%	100.0%	82.61%	100.0%	90.48%	0.913
CNN-3	92.11%	92.86%	88.46%	100.0%	84.21%	93.88%	0.921
CNN-4	97.83%	97.62%	100.0%	95.65%	100.0%	97.78%	0.978
CNN-5	89.93%	90.48	88.00%	95.65%	84.21%	91.67%	0.899

Source: Own authorship.

The confusion matrix was also evaluated with CNN-4 (Figure [5.11](#)).

Figure 5.11. CNN-4 confusion matrix.

Source: Own authorship.

5.2.1.4 S2 reevaluation with augmented dataset

In addition to using the C19 dataset (19 slices per patient) for S2 partition training, the data was also doubled using a random zoom variation range [0.99, 1.01], a random brightness variation range [0.8, 1.2], and a random horizontal flip. This is a classical strategy for augmenting image data. The consequent performance results are shown in Table [5.19](#).

Table 5.19. Data augmentation in S2 training dataset - performance evaluation

Model	B_{ca}	A_{cy}	P_{rn}	S_{yn}	S_{py}	F1	AUC
CNN-1	89.93%	90.48%	88.00%	95.65%	84.21%	91.67%	0.899
CNN-2	92.11%	92.86%	88.46%	100.0%	84.21%	93.88%	0.921
CNN-3	89.47%	90.48%	85.19%	100.0%	78.95%	92.00%	0.895
CNN-4	92.56%	92.86%	91.67%	95.65%	89.47%	93.62%	0.926
CNN-5	82.15%	80.95%	94.12%	69.57%	94.74%	80.00%	0.822

Source: Own authorship.

5.2.1.5 S1 reevaluation with majority voting classifier

The performance of the S1 partition was again measured using the proposed DL models (CNN-1, CNN-2, CNN-3, CNN-4 and CNN-5) and the results were compared between the decision made on each slice individually (single instance) and the decision made by the majority (multiple instance decision). The results can be found in Table 5.20.

Table 5.20. Customized cnn results

Model	B_{ca}	A_{cy}	P_{rn}	S_{yn}	S_{py}	F1	AUC
CNN-1 (<i>S</i>)	96.1%	95.9%	97.9%	94.7%	97.4%	96.3%	0.96
CNN-1 (<i>M</i>)	97.8%	97.6%	100%	95.7%	100%	97.8%	0.98
CNN-2 (<i>S</i>)	96.6%	96.5%	97.5%	96.3%	96.8%	96.9%	0.97
CNN-2 (<i>M</i>)	97.8%	97.6%	100%	95.7%	100%	97.8%	0.98
CNN-3 (<i>S</i>)	94.7%	94.9%	95.0%	95.9%	93.6%	95.4%	0.95
CNN-3 (<i>M</i>)	95.1%	95.1%	95.7%	95.7%	94.4%	95.7%	0.95
CNN-4 (<i>S</i>)	95.3%	95.3%	96.5%	95.0%	95.6%	95.7%	0.95
CNN-4 (<i>M</i>)	100%	100%	100%	100%	100%	100%	1.00
CNN-5 (<i>S</i>)	98.2%	98.2%	98.9%	97.9%	98.5%	98.4%	0.98
CNN-5 (<i>M</i>)	100%	100%	100%	100%	100%	100%	1.00

S - Single decision.

M - Majority voting decision.

Source: Own authorship.

Likewise, the performance of the S1 partition was again measured using the pre-trained models (VGG-19, VGG-16, ResNet-50 and Inception V3) and the results were compared between the decision made on each slice individually (single instance) and the decision made by the majority (multiple instance decision). The results can be found in Table 5.21.

Table 5.21. Pretrained models results

Model	B_{ca}	A_{cy}	P_{rn}	S_{yn}	S_{py}	F1	AUC
VGG-16 (<i>S</i>)	92.9%	92.9%	93.8%	93.4%	92.1%	93.7%	0.93
VGG-16 (<i>M</i>)	97.2%	97.6%	95.8%	100%	94.4%	97.9%	0.97
VGG-19 (<i>S</i>)	95.7%	95.8%	96.1%	96.3%	95.0%	96.2%	0.96
VGG-19 (<i>M</i>)	97.2%	97.6%	95.8%	100%	94.4%	97.9%	0.97
ResNet (<i>S</i>)	97.5%	97.3%	99.5%	95.7%	99.4%	97.6%	0.98
ResNet (<i>M</i>)	97.8%	97.6%	100%	95.6%	100%	97.8%	0.98
Inception (<i>S</i>)	95.9%	95.9%	96.6%	96.1%	95.6%	96.3%	0.96
Inception (<i>M</i>)	97.8%	97.6%	100%	95.6%	100%	97.8%	0.98

S - Single decision.

M - Majority voting decision.

Source: Own authorship.

The pre-trained CNNs were initialized with weights from training on the ImageNet dataset,

finalized with 3 dense layers and trained in full. ADAM was used as an optimizer and training was carried out over 100 epochs, with a stopping criterion for validation accuracy greater than 0.999.

It can be seen in Table 5.20 that CNN-4 achieved 100% performance in all objective metrics used. To investigate the performance of CNN-4 concerning the quiescent point (local minimum) where the training algorithm converges, the network was retrained 10 times with random starting points and all trained models were saved. Using the models obtained in the training step, the network’s performance was evaluated again for each case. The performance results are shown in Table 5.22.

Table 5.22. CNN-4 training results

Train	B_{ca}	A_{cy}	P_{rn}	S_{yn}	S_{py}	F1	AUC
1	100%	100%	100%	100%	100%	100%	1.00
2	100%	100%	100%	100%	100%	100%	1.00
3	97.8%	97.6%	100%	95.7%	100%	97.8%	0.98
4	97.2%	97.6%	95.8%	100%	94.4%	97.8%	0.97
5	100%	100%	100%	100%	100%	100%	1.00
6	100%	100%	100%	100%	100%	100%	1.00
7	100%	100%	100%	100%	100%	100%	1.00
8	97.2%	97.6%	95.8%	100%	94.4%	97.8%	0.97
9	100%	100%	100%	100%	100%	100%	1.00
10	100%	100%	100%	100%	100%	100%	1.00
Mean	99.2%	99.3%	99.2%	99.6%	98.9%	99.3%	0.99
Std	1.3%	1.2%	1.8%	1.4%	2.4%	1.1%	0.01

Mean - statistical average.

Std - standard deviation.

Source: Own authorship.

The Table 5.22 reveals that CNN-4 achieved an average accuracy value of 99.3% in distinguishing the diagnosis between CN and AD.

5.2.2 Analysis of the axial image dataset

As described in Chapter 4, an axial image dataset was generated with a central slice (as in the coronal strategy), and four caudal and four cranial slices (Figure 4.15). The axial dataset was generated with the three classes CN, MCI and AD. However, to compare the performance of the axial and coronal dataset, only the CN and AD classes were used. Table 5.23 shows the resulting numbers of images in the training, validation, and test partitions.

Table 5.23. Split axial dataset

Split	Training	Validation	Test
Axial	1.494 AD images	81 AD images	81 AD images
	1.728 CN images	99 CN images	90 CN images

Source: Own authorship.

The performance of the axial image dataset was evaluated with the proposed DL models (CNN-1, CNN-2, CNN-3, CNN-4 and CNN-5). These metrics are shown in Table 5.24.

Table 5.24. Axial plan - performance evaluation

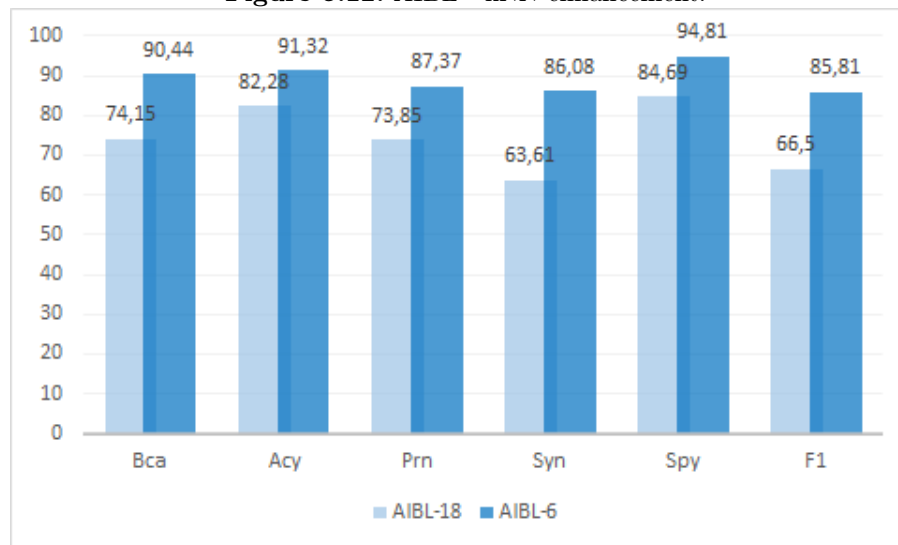
Model	B_{ca}	A_{cy}	P_{rn}	S_{yn}	S_{py}	F1	AUC
CNN-1	55.99%	56.73%	57.27%	70.00%	41.98%	63.00%	0.560
CNN-2	42.41%	43.27%	46.90%	58.89%	25.93%	52.22%	0.424
CNN-3	60.93%	59.06%	88.46%	25.56%	96.3%	39.66%	0.609
CNN-4	59.94%	59.06%	67.24%	43.33%	76.54%	52.70%	0.599
CNN-5	59.01%	59.06%	61.36%	60.00%	58.02%	60.67%	0.590

Source: Own authorship.

5.3 DISCUSSION OF THE RESULTS OF THE CLINICAL DATA ANALYSIS

The kNN model improved performance with the AIBL-18 and AIBL-6 datasets by roughly 22% (B_{ca}). This result is statistically significant ($p < 0,05$).

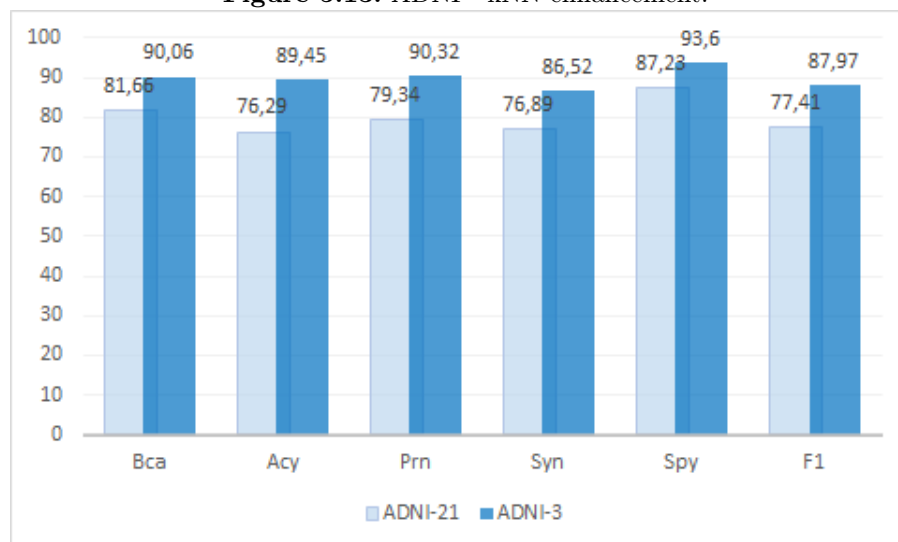
Figure 5.12 shows the impact of the FMFS feature selection algorithm on improving the kNN evaluation metrics (Table 5.5) with the AIBL database.

Figure 5.12. AIBL - kNN enhancement.

Source: Own authorship.

Similarly, the kNN model improved performance with the ADNI-21 and ADNI-3 datasets by 10% (\mathbf{B}_{ca} metric) (significant at $p < 0,05$).

Figure 5.13 shows the impact of the FMFS feature selection algorithm on improving the performance of the kNN model (Table 5.4) with the ADNI database.

Figure 5.13. ADNI - kNN enhancement.

Source: Own authorship.

Therefore, the kNN model improved the performance of diagnosing Alzheimer's disease in the AIBL-6 and ADNI-3 datasets by reducing dimensionality, highlighting key features, and reducing training time – all desirable objectives for the proposed feature selection method.

The SVM model did not significantly improve performance (\mathbf{B}_{ca} metric) relative to AIBL-18 and AIBL-4 or ADNI-21 and ADNI-11. Nevertheless, the model did reduce dimensionality, enhance key features, and reduce training times.

The FMFS algorithm helped eliminate features that did not significantly contribute to distinguishing database classes. This strategy tends to facilitate the performance of the kNN algorithm by eliminating what are essentially noise in the calculation of nearest neighbors. The result of this combination was an improvement of 10.29% in the accuracy of the ADNI base evaluation and 21.97% in the accuracy of the AIBL base evaluation.

However, the FMFS algorithm did not have a significant impact on the SVM model, probably because the kernel function has a regularization parameter (C) that finds an optimal balance between high margins and a smaller number of incorrectly classified data points. Controlling C then amounts to controlling outliers - low C values imply more outliers while high C values imply fewer outliers. The hyper-parameters were fit within the inner loop of the nested cross-validation for all the classification models used in this study.

Evaluation of the datasets using 12 other feature selection methods produced accuracy results that were both higher and lower than those obtained with the proposed FMFS algorithm. The simplicity of the FMFS calculation and the quality of the results obtained show that FMFS is a viable option for evaluating the datasets dealt with in this thesis.

5.3.1 Comparisons with other studies

The best results obtained in this study were compared with the best results of other studies that worked with AIBL, ADNI and OASIS (Tables 5.25, 5.26 and 5.27). These studies were carried out between 2018 and 2021 and most used the ADNI database (KHAN *et al.*, 2021). For the AIBL and ADNI databases, performance comparisons were based on the CN, MCI, and AD classes, while the CDR 0, CDR 0.5, and CDR 12 classes were used for OASIS.

This multi-study comparison only evaluated the \mathbf{B}_{ca} metric, the MAUC metrics, and the machine learning techniques and models used in each study.

The \mathbf{B}_{ca} achieved by (NIYAS; P, 2021b) was nearly equal to that of the present study ($\mathbf{B}_{ca} = 92.52\%$) for the AIBL database and the SVM classifier (Table 5.25). When the Gradient Boosting classifier was used, the \mathbf{B}_{ca} of the present study was slightly higher (92.88%).

Table 5.25. Comparison with previous study (AIBL dataset)

Reference	ML/DL	Model	\mathbf{B}_{ca}	MAUC
		GB	92,88	0,974
Present study	ML	SVM	92,52	0,959
(NIYAS; P, 2021b)	ML	kNN	82,50	NA
		SVM	92,50	NA

NA - not available. GB - Gradient Boosting.

Except for MAUC, all numerical results are expressed as percentages.

Source: Own authorship.

For the AIBL database, the best performance boost for the kNN classifier was achieved using the FMFS algorithm. Here, \mathbf{B}_{ca} increased from 74.15% to 90.44% (Table 5.5).

Table 5.26 compares our results with the ADNI database to those of (NIYAS; P, 2021a), (NIYAS; P, 2021b), (NGUYEN M., 2020), (LIU J LI, 2020), (ALBRIGHT, 2019), (IDDI *et al.*, 2019), (MOORE *et al.*, 2019), (GHAZI *et al.*, 2019) and (NGUYEN *et al.*, 2018). In the present study, \mathbf{B}_{ca} reached 92.61% for the SVM classifier, which was 2.9% higher than the best result reported by the other studies. The logistic regression classifier also reached a \mathbf{B}_{ca} of 92.46%.

Table 5.26. Comparison with previous studies (ADNI dataset)

Reference	ML/DL	Model	B_{ca}	MAUC
		LR	92,46	0,973
Present study	ML	SVM	92,61	0,969
(NIYAS; P, 2021a)	ML	RF	87	NA
		kNN	90	0,96
(NIYAS; P, 2021b)	ML	SVM	88	0,95
(NGUYEN M., 2020)	DL	RNN	88,7	0,944
(LIU J LI, 2020)	DL	RNN	87,93	0.9439
(ALBRIGHT, 2019)	DL	RNN	NA	0,866
(IDDI <i>et al.</i> , 2019)	ML	RF	86	NA
(MOORE <i>et al.</i> , 2019)	ML	RF	73	0,82
(GHAZI <i>et al.</i> , 2019)	ML	LDA	NA	0,75
(NGUYEN <i>et al.</i> , 2018)	ML	SVM	79	0,86

NA - not available. LDA-*linear discriminant analysis*. RNN-*recurrent neural network*. RF-*random forest*. LR-*logistic regression*.

Except for MAUC, all numerical results are expressed as percentages.

Source: Own authorship.

Our results for diagnosing AD from the clinical data in the ADNI and AIBL databases were approved and presented at the IX Latin American Congress of Biomedical Engineering (CLAIB 2022) and at the XXVIII Brazilian Congress of Biomedical Engineering (CBEB 2022) via the article "Simplified feature selection in Alzheimer's disease diagnosis via machine learning" (Seleção simplificada de atributos para auxílio ao diagnóstico da doença de Alzheimer utilizando aprendizado de máquina) (SARAIVA JR. *et al.*, 2022). The article presented the FMFS feature selection algorithm, its impact on improving performance metrics, and a comparison with previous studies.

Few studies have focused on the OASIS and AIBL databases (KHAN *et al.*, 2021). Nevertheless, it was possible to compare our results with those of (PAIVA; ESCOVEDO, 2021). Table 5.27 shows that the B_{ca} of the present study was 4.5% better than the best result obtained by (PAIVA; ESCOVEDO, 2021). The random forest classifier also performed better, reaching a B_{ca} of 75.30%.

Table 5.27. Comparison with previous study (OASIS dataset)

Reference	ML/DL	Model	B_{ca}	MAUC
		RF	75,30	0,881
Present study	ML	DT	76,25	0,842
(PAIVA; ESCOVEDO, 2021)	ML	RF	64,30	NA
		LR	73,00	NA

NA - not available. RF-*random forest*. DT-*decision tree*. LR-*logistic regression*.

Except for MAUC, all numerical results are expressed as percentages.

Source: Own authorship.

5.3.2 Other considerations regarding clinical data analysis

The FMFS algorithm reduced the features from 18 to four features ($4.5\times$ reduction) in the AIBL database and from 21 to three features ($7\times$ reduction) in the ADNI database. However, in the OASIS database, the reduction was only $1.4\times$. This partly explains why the benchmarking metrics (Table 5.6) did not improve significantly.

While the AIBL-18 dataset had 785 instances and ADNI-21 had 1033 instances, OASIS-7 had only 217 instances. This can explain why the classification models struggled to generalize predictions for the OASIS-7 dataset when distinguishing between either two or three classes.

The performance evaluation for two classes showed AD diagnosis accuracy of nearly 100%. A particularly high B_{ca} of 99.88% and an AUC of 1,000 were achieved when the CN x AD classes were evaluated in the ADNI-11 dataset, using the SVM classifier. Under the same evaluation conditions, B_{ca} values of 96.76% and 88.04% were achieved for the AIBL-4 and OASIS-5 datasets, respectively.

Tables 5.10 and 5.11 show that the SVM classifier achieved consistently better results than other classifiers, with the exception of Gradient Boosting. This is partly because the AIBL-4 and ADNI-11 datasets were generated by the FMFS algorithm and fit to the SVM classifier. Conversely, the decision tree classifier produced the best results for the OASIS-5 dataset.

The FMFS algorithm showed comparable results with 12 other traditional feature selection methods, as shown in the results presented in Table 5.13, Table 5.14 and Table 5.15.

5.4 DISCUSSION OF THE IMAGE ANALYSIS RESULTS

The primary ADNI1: Screening 1.5T database consists of 307 patients diagnosed as CN and 243 diagnosed with AD (Table 4.1). Clearly, this quantity of MRI images is insufficient for image-based diagnosis based on DL and deep neural network training.

MRI data were available in NIFTI 3D format, with images in the axial, coronal and sagittal planes. The medial temporal lobe is the region of interest for AD diagnosis. The mesial temporal lobe is home to the hippocampus, which is mirrored across the midline of the body, responsible for memory, and one of the structures that is most affected by AD. Images in the sagittal plane cannot be used since they show only the left or right side of the hippocampus. Thus, the coronal and axial planes were chosen for analysis.

All coronal and axial image datasets generated in this study were used to train the proposed networks (CNN-1, CNN-2, CNN-3, CNN-4 and CNN-5). The CNN-1 network (Figure 4.24) has ten convolutional layers, five max pooling layers, seven batch normalization layers, five dropout layers, one flatten layer, and four dense layers. CNN-2 (Figure 4.25) is a version of CNN-1 with reduced layer dimensionality while CNN-3 (Figure 4.26) is version of CNN-1 with greater layer dimensionality. CNN-4 (Figure 4.27) has more layers than CNN-1, which increases depth, while CNN-5 (Figure 4.28) has the same layers as CNN-4, but with reduced layer dimensionality. The pre-trained network models VGG-16, VGG-19, ResNet-50 and Inception V3 were also used to evaluate the image datasets.

5.4.1 Coronal plane

An image dataset with 226 CN patients and 173 AD patients was generated for the coronal plane. Therefore, the proposed image selection strategy used 73.6% of the CN patients and 71.2% of the AD patients in the primary ADNI1: Screening 1.5T database. More complete use of the database was impossible because of missing data or handling errors in the MRI files. This meant that an already small dataset was made even smaller, necessitating the use of data augmentation strategies.

Since image selection was based on detecting the hippocampus within a given slice, the first

data augmentation strategy was to identify a central image and then select adjacent slices, given that the hippocampus spans a few centimeters. This strategy was used both in the coronal and axial planes.

This process produced two datasets: C1, consisting of only central slices from 226 CN images and 173 AD images, and C2, which added nine anterior and nine posterior slices, resulting in 4,294 CN images and 3,287 AD images.

Additionally, two different partitions, S1 and S2, were generated for image processing by combining the C1 and C2 datasets (Table 5.23).

Evaluating the S1 coronal partition (Table 5.23) with different DL models (Table 5.17) showed that CNN-5 produced the highest \mathbf{B}_{ca} (94.79%). Greater network depth and reduced dimensionality may have contributed to this outcome. The same evaluation was also carried out with pre-trained networks (VGG-19, VGG-16, ResNet-50 and Inception V3), yielding results that were lower than that of CNN-5. The number of images in both S1 and S2 was insufficient to train networks as deep as the pre-trained networks.

The S2 coronal partition was evaluated with the CNN-1, CNN-2, CNN-3, CNN-4 and CNN-5 models. CNN-4 showed the highest \mathbf{B}_{ca} result (97.83%), which was the best performance for image-based AD classification. Like CNN-5, CNN-4 has greater depth, but with increased dimensionality.

The data in the S2 coronal partition was also augmented via classical treatments that consisted of $2\times$ data magnification of the training partition using a random zoom range [0.99, 1.01], random brightness [0.8, 1.2], and random horizontal inversion. These treatments did not improve the results. Thus, for this dataset and under these evaluation conditions, data augmentation via lateral slices was more efficient than using classical techniques based on zoom, brightness, and horizontal inversion.

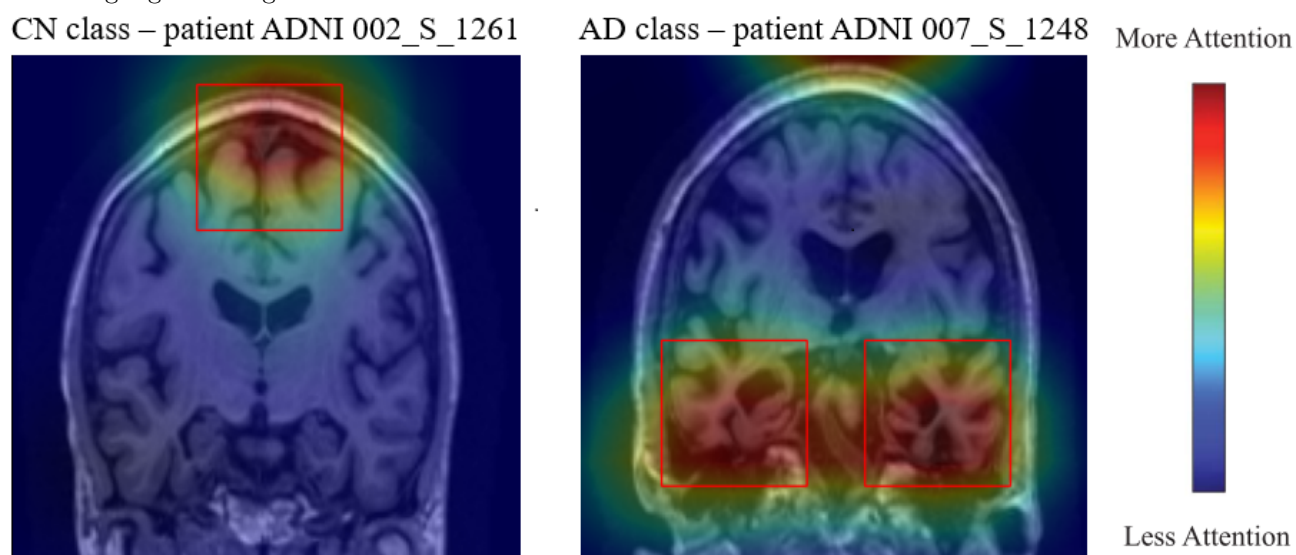
In (LIN *et al.*, 2013) a new deep network called "Network In Network" (NIN) was proposed for classification tasks. This new structure consists of using a global average pooling (GAP) layer as a replacement for the fully connected layers in conventional CNNs. According to the authors, the GAP layer acts as a structural regularizer that prevents global overfitting of the network. In addition, the visualization of the class activation maps (CAM) of the last

convolutional layer of the CNN makes it possible to detect objects in the input image.

The use of the GAP layer for object localization was carried out in (ZHOU *et al.*, 2015), which concluded that the network was able to localize the discriminative regions of the image. CAM highlights specific regions in the image according to the class decided.

Using these concepts, we created a VGG-16-GAP network by adding the GAP layer to a VGG-16 network, trained it using images from the S1 partition of the coronal dataset, and began evaluating the regions highlighted by the CAM as a function of each class decided by the network. Figure 5.14 shows the heat map of two different patients diagnosed with CN and AD.

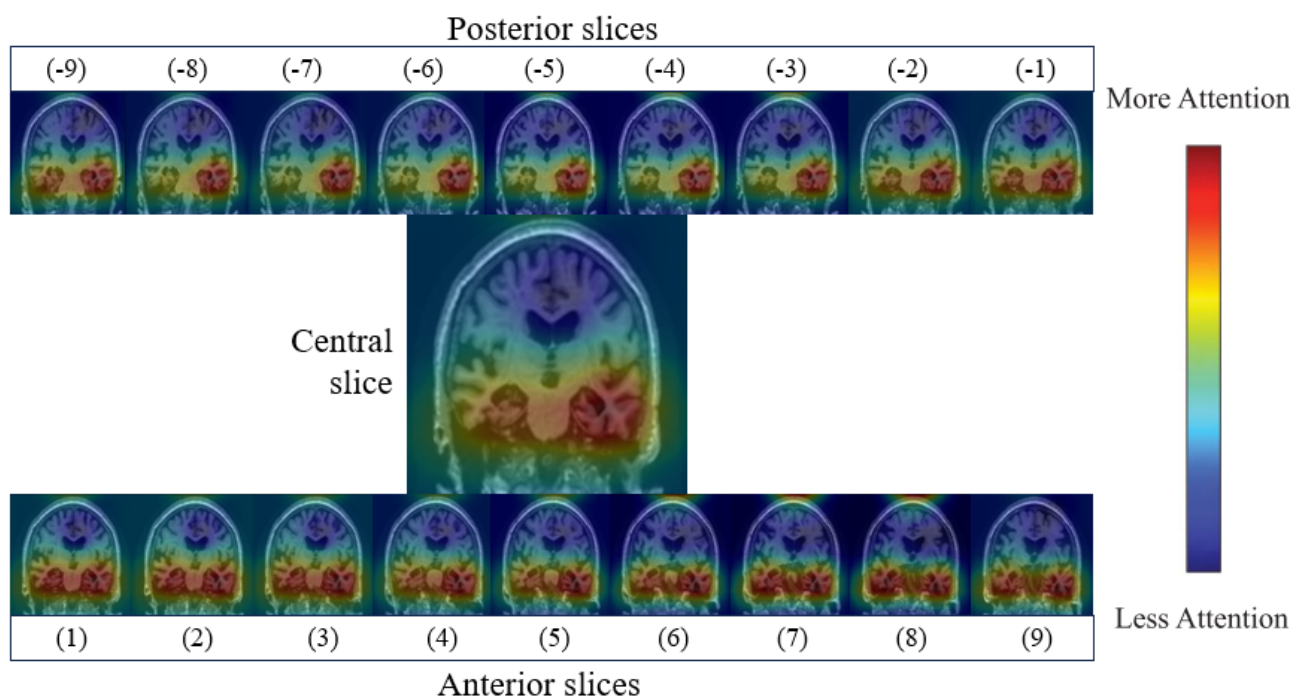
Figure 5.14. CN ADNI 002_S_1261 patient heat map and AD ADNI 007_S_1248 patient heat map. Red boxes highlight the region of interest.



Source: Own authorship.

In these cases, shown in Figure 5.14, it is clear that the network highlights the hippocampus region in the patient diagnosed with AD, which is not the case for the patient diagnosed with CN. Cortical atrophy is a non-specific alteration of AD, as it is present in several neurodegenerative diseases (TEIPEL *et al.*, 2008). It can be observed in several slices of the AD group, but was not taken into account by the neural network. This demonstrates the neural network's ability to identify more specific patterns of dementia associated with AD.

The heat maps for each of the 19 images of ADNI patient 007_S_1248 were also evaluated, as shown in Figure 5.15.

Figure 5.15. Heat map of the 19 images of patient ADNI 007_S_1248 diagnosed with AD.

Source: Own authorship.

It can be seen that the heat map of the 19 slices chosen for ADNI patient 007_S_1248 have a certain similarity, although it is noticeable that there is greater prominence in the region of the right hippocampus in the posterior slices, moving towards balance as one moves towards the anterior slices. In this case, all slices were individually decided to be diagnosed with AD. The choice of 19 slices per patient seems to have been correct, given that the network was able to locate the hippocampus region in all the slices of ADNI patient 007_S_1248.

The results presented by Table 5.20 and Table 5.21 show that the strategy of using multiple instances and a majority classifier is a good alternative to a single instance decision. In addition to achieving better metric values, majority decision showed greater robustness in choosing the CNN model.

5.4.2 Axial plane

An image dataset was generated for the axial plane with data from 213 CN and 184 AD patients. Thus, our proposed selection strategy used 69.4% of the CN patients and 75.7% of the AD patients that had been diagnosed in the primary ADNI1: Screening 1.5T database.

More complete use of the database was not possible because of numerous cases of missing data or errors in handling the MRI files.

The resulting axial plane dataset consisted of 1,656 AD images and 1,917 CN images (Table 5.23). These were set up in a single partition. Unlike the coronal plane, where the dataset was expanded with 19 slices, data augmentation was achieved with only 9 additional slices per patient in the axial plane.

B_{ca} was lower than 61% when the axial dataset was evaluated with the CNN-1, CNN-2, CNN-2, CNN-3, CNN-4 and CNN-5 networks. Thus, in the current study, images in the coronal plane yielded much better results than those of the axial plane.

The axial image dataset in this study consisted of only 1656 AD images and 1917 CN images, while the S2 partition of the coronal image dataset had 2600 AD images and 3466 CN images. The smaller dimensionality of the axial image dataset may have contributed to its inferior performance. Moreover, slice selection in the axial plane may have provided inferior options for identifying areas of interest.

5.4.3 Comparison with other studies

The best results from this study were compared with those from other studies (Table 5.28).

Table 5.28. Comparison with previous DL studies

Reference	Dataset	B_{ca}	A_{cy}	P_{rn}	S_{yn}	S_{py}	F1
Present study	ADNI	99.2%	99.3%	99.2%	99.6%	98.9%	99.3%
(HAN <i>et al.</i> , 2020)	ADNI	NA	98,67%	99,66%	98,18%	99,27%	99,46%
(LI <i>et al.</i> , 2022)	ADNI	NA	93,16%	75,00%	83,18%	78,59%	77,38%
(RAZZAK <i>et al.</i> , 2022)	ADNI	NA	87,92%	NA	NA	NA	NA
(HAZARIKA <i>et al.</i> , 2022)	ADNI	NA	97%	96%	NA	NA	97%
(WU <i>et al.</i> , 2022)	ADNI	NA	91,31%	93,8%	88,3%	94,2	91,0%
(SHARMA <i>et al.</i> , 2022)	ADNI	NA	97,33%	NA	100%	97%	NA
(ZHANG <i>et al.</i> , 2023)	AIBL	NA	93,4%	NA	NA	NA	NA
(GAO <i>et al.</i> , 2023a)	ADNI	NA	90,5%	NA	84,6%	95,0%	NA
(GAO <i>et al.</i> , 2023b)	ADNI	NA	90,4%	NA	90,3%	90,5%	NA

NA - not available.

Source: Own authorship.

The accuracy achieved by this study was better than that of the other studies from 2020 to 2023 presented in Table 5.28. All the studies presented attempted to distinguish between CN and AD, used MR images from the ADNI and AIBL datasets, and employed DL techniques. It should be noted that these compared studies may have selected datasets from the many options available in the ADNI database that differed from this study, making direct comparison difficult.

5.5 SUGGESTED APPLICATION FOR NEW TECHNOLOGIES

As seen in Chapter-1, the Alzheimer's Disease International report (Gauthier S. *et al.*, 2021) states that 75% of people with dementia are not being diagnosed for some reason. From a technological point of view, it can be said that the techniques presented in this study can add up to a useful technological toolkit to help diagnose AD, with the potential to help millions of people who may be under suspicion of this disease.

In addition to academic studies, some of which have been reviewed in this paper, there are also technological platforms that use state-of-the-art ML technology to build innovative tools for automatic quantitative analysis of medical images, such as the "InnerEye" project by Microsoft Research, which is supported by the National Health System in England.

5.5.1 Brazilian Unified Health System - SUS

According to a definition on the Brazilian Ministry of Health's website, the Unified Health System (SUS) is one of the largest and most complex public health systems in the world, covering everything from a simple blood pressure assessment through Primary Care to organ transplants, guaranteeing full, universal and free access for the entire population of the country. With its creation, the SUS provided universal access to the public health system, without discrimination. Comprehensive health care, and not just healthcare, became the right of all Brazilians, from pregnancy onwards and throughout their lives, with a focus on health and quality of life, aimed at prevention and health promotion.

¹<https://www.microsoft.com/en-us/research/project/medical-image-analysis/>

²<https://www.gov.br/saude/pt-br/assuntos/saude-de-a-a-z/s/sus>

5.5.2 Biomedical Engineering Innovation Award for SUS 2023

Figure 5.16. Biomedical Engineering Innovation Award for SUS 2023.



Source: Adapted from <https://www.bostonscientific.com/pt-BR/inovacao/premio-de-inovacao-para-o-sus.html>.

This competition aims to recognize and reward technical-scientific work in the area of Biomedical Engineering, with a theme related to solutions for the public health system, proposing practical solutions to the problems faced by the Unified Health System - SUS.

According to the call for proposals³, the proposals will be evaluated by members of:

1. Brazilian Society of Biomedical Engineering - SBEB;
2. Ministry of Health - MS;
3. National Health Surveillance Agency - ANVISA;
4. Pan American Health Organization - PAHO;
5. International Federation of Medical and Biological Engineering - IFMBE.

5.5.3 Winning proposal

This work was submitted to the Award and was recognized as a winner, having been evaluated under the following aspects:

1. INNOVATION - Potential innovation of the project for the SUS;
2. APPLICABILITY TO SUS - The extent to which the project can be applied and impact the SUS, considering its policies, basic pillars and the demands of the population;

³https://sbeb.org.br/wp-content/uploads/2023/06/EDITAL_SBEB_PREMIO_2023_V2-1.pdf

3. QUALITY OF LIFE AND POPULATION EXPANSION - Impact of the project on the quality of life of the SUS user population and the increase in the population served;
4. TIME TO BENEFIT - Degree of maturity of the project. Time needed for the benefits to the SUS to be achieved;
5. COST/BENEFIT - Quantity and complexity of resources required (technological, financial, human, time) to achieve/execute the project in view of its the benefits for SUS and society;
6. IMPACTS: SOCIAL, TECHNOLOGICAL, SCIENTIFIC - From the perspective of Brazilian society, the potential of the project in terms of generating new scientific advances, future job creation, exports and the strengthening of the Brazilian Industrial Complex, reducing the dependence of the SUS on the foreign market (imports).

This chapter presented the results of the clinical analyses and those obtained from evaluating magnetic resonance images. It also presented a discussion about the results and a comparison with results achieved by other researchers. The next chapter will provide conclusions and present the contributions of the present study.

CONCLUSIONS

The application of computational intelligence to clinical data proved to be an efficient aid to Alzheimer's disease diagnosis. We used clinical data from three renowned studies, carried out several exploratory experiments and proposed new technology-based procedures that provided the best fit between data and diagnosis. Our results reached a predictive accuracy of 99.81% for AD, using data from the Alzheimer's Disease Neuroimaging Initiative study, the most frequently used database.

We developed a promising new algorithm for processing clinical data that simplified the feature selection process and produced superior accuracy results in deciding the diagnosis of Alzheimer's compared to other studies.

The proposed feature selection algorithm also helped reduce database dimensionality and training time for learning models and assisted in identifying the features that most influence the diagnosis of Alzheimer's disease, whose precise diagnosis continues to challenge science.

Computational intelligence techniques were also applied to image data obtained from magnetic resonance imaging. Strategies were proposed for building datasets with images in the coronal and axial planes. Disease prediction accuracy of 99.3% was reached by performing image treatments on magnetic resonance data from the Alzheimer's Disease Neuroimaging Initiative study (ADNI1: Screening 1.5T).

Given the large volumes of data needed for learning models and the paucity of available MRI images, a data augmentation strategy was developed and compared to a classical augmentation approach. Under the specific conditions of the current study, this proposed data augmentation strategy proved to be superior, and images in the coronal plane provided better results than those in the axial plane.

Convolutional network architectures were also proposed by varying layer dimensionality and depth. This was done to find computational tools that could help improve diagnostic indicators

of Alzheimer's disease.

Our results indicate that the approach taken in this study is promising and contributes to improving Alzheimer's diagnosis based on computational intelligence. Further research could be carried out to automate the selection of the central slice in the coronal plane, instead of the manual selection carried out in this study.

BIBLIOGRAPHY

ALBRIGHT, J. *Forecasting the Progression of Alzheimer's Disease Using Neural Networks and a Novel Pre-Processing Algorithm*. arXiv, 2019. Disponível em: <https://arxiv.org/abs/1903.07510>. Cited 5 times in pages 15, 16, 18, 103, and 104.

ALTMANN, A.; TOLOŞI, L.; SANDER, O.; LENGAUER, T. Permutation importance: a corrected feature importance measure. *Bioinformatics*, Oxford University Press, v. 26, n. 10, p. 1340–1347, 2010. Cited in page 28.

BALAKRISHNAMA, S.; GANAPATHIRAJU, A. Linear discriminant analysis - a brief tutorial. *Institute for Signal and Information Processing*, 1998. Cited in page 48.

BALESTRIERI J., N. M. B. G. L. . P. M. N. Structural volume of hippocampus and alzheimer's disease. *Revista da Associacao Medica Brasileira*, Vol. 66(4), p. 512–515, 2020. Cited 2 times in pages 2 and 3.

BLEICH, H. L. Computer evaluation of acid-base disorders. *The Journal of clinical investigation*, Vol. 48,9, p. 1689–1696, 1969. Cited in page 14.

BREIMAN, L. Random forests. *Machine Learning*, Kluwer Academic Publishers, v. 45, n. 1, p. 5–32, 2001. ISSN 0885-6125. Disponível em: <http://dx.doi.org/10.1023/A%3A1010933404324>. Cited in page 48.

BRITO, E. N. D.; FIGUEIREDO, B. Q. d.; SOUTO, D. N.; NOGUEIRA, J. F.; MELO, A. L. d. C.; SILVA, I. T. d.; OLIVEIRA, I. P.; ALMEIDA, M. G. d. Artificial intelligence in the diagnosis of neurodegenerative diseases: a systematic literature review. *Research, Society and Development*, v. 10, n. 11, p. e482101120004, Sep. 2021. Disponível em: <https://rsdjournal.org/index.php/rsd/article/view/20004>. Cited in page 15.

CAWLEY, G. C.; TALBOT, N. L. On over-fitting in model selection and subsequent selection bias in performance evaluation. *J. Mach. Learn. Res.*, JMLR.org, v. 11, p. 2079–2107, aug 2010. ISSN 1532-4435. Cited in page 42.

CENTRAL, D. C. *Stages of Alzheimer's Dementia: Durations Scales Used to Measure Progression (GDS, FAST CDR)*. 2022. URL: <https://www.dementiacarecentral.com/aboutdementia/facts/stages/>. Cited in page 5.

CHAWLA, N. V.; BOWYER, K. W.; HALL, L. O.; KEGELMEYER, W. P. SMOTE: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, AI Access Foundation, v. 16, p. 321–357, jun 2002. Disponível em: <https://doi.org/10.1613%2Fjair.953>. Cited in page 26.

CHEN, T.; GUESTIN, C. XGBoost. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 2016. Disponível em: <https://doi.org/10.1145%2F2939672.2939785>. Cited in page 48.

- CORTES, C.; VAPNIK, V. Support vector networks. *Machine Learning*, v. 20, p. 273–297, 1995. Cited 2 times in pages [44](#) and [45](#).
- COVER, T.; HART, P. Nearest neighbor pattern classification. *IEEE Transactions on Information Theory*, v. 13, n. 1, p. 21–27, 1967. Cited in page [47](#).
- CUI, Y.; JIA, M.; LIN, T.-Y.; SONG, Y.; BELONGIE, S. *Class-Balanced Loss Based on Effective Number of Samples*. 2019. Cited in page [26](#).
- DEKEYZER, S.; BOSSCHE, S. V. "unforgettable" - a pictorial essay on anatomy and pathology of the hippocampus. *Insights into Imaging*, v. 8, 01 2017. Cited in page [58](#).
- DHAL, P.; AZAD, C. A comprehensive survey on feature selection in the various fields of machine learning. *Applied Intelligence*, Springer, p. 1–39, 2022. Cited in page [27](#).
- DING, H.; FENG, P.-M.; CHEN, W.; LIN, H. Identification of bacteriophage virion proteins by the anova feature selection and analysis. *Molecular BioSystems*, Royal Society of Chemistry, v. 10, n. 8, p. 2229–2235, 2014. Cited in page [28](#).
- ELLIS, K.; BUSH, A.; DARBY, D.; FAZIO, D.; FOSTER, J.; HUDSON, P.; LAUTENSCHLAGER, N.; LENZO, N.; MARTINS, R.; MARUFF, P.; MASTERS, C.; MILNER, A.; PIKE, K.; ROWE, C.; SAVAGE, G.; SZOEKE, C.; TADDEI, K.; VILLEMAGNE, V.; WOODWARD, M.; AMES, D. The australian imaging, biomarkers and lifestyle (aibl) study of aging: Methodology and baseline characteristics of 1112 individuals recruited for a longitudinal study of alzheimer's disease. *International psychogeriatrics / IPA*, v. 21, p. 672–87, 06 2009. Cited 3 times in pages [4](#), [21](#), and [23](#).
- ESTUDIOS circulo de. *Brajula Metropolitana*. 2017. URL: <https://circulodeestudios-centrohistorico.blogspot.com/2015/09/sigue-comiendo-chatarra-y-se-te-va.html>. Cited in page [49](#).
- FEDOROV, A. Y.; BEICHEL, R. R.; KALPATHY-CRAMER, J.; FINET, J.; FILLION-ROBIN, J.-C.; PUJOL, S.; BAUER, C.; JENNINGS, D. L.; FENNESSY, F. M.; SONKA, M.; BUATTI, J. M.; AYLWARD, S. R.; MILLER, J. V.; PIEPER, S. D.; KIKINIS, R. 3d slicer as an image computing platform for the quantitative imaging network. *Magnetic resonance imaging*, v. 30 9, p. 1323–41, 2012. Cited in page [55](#).
- FREUND, Y.; MASON, L. The alternating decision tree learning algorithm. In: *ICML*. [s.n.], 1999. p. 124–133. Disponível em: <http://www.lsmason.com/papers/ICML99-AlternatingTrees.pdf>. Cited in page [48](#).
- FRIEDMAN, J. H. Greedy function approximation: A gradient boosting machine. *The Annals of Statistics*, Institute of Mathematical Statistics, v. 29, n. 5, p. 1189–1232, 2001. ISSN 00905364. Disponível em: <http://www.jstor.org/stable/2699986>. Cited in page [48](#).
- GAO, X.; CAI, H.; LIU, M. A hybrid multi-scale attention convolution and aging transformer network for alzheimer's disease diagnosis. *IEEE Journal of Biomedical and Health Informatics*, IEEE, 2023. Cited 3 times in pages [16](#), [19](#), and [110](#).
- GAO, X.; LIU, H.; SHI, F.; SHEN, D.; LIU, M. Brain status transferring generative adversarial network for decoding individualized atrophy in alzheimer's disease. *IEEE Journal of Biomedical and Health Informatics*, IEEE, 2023. Cited 3 times in pages [16](#), [19](#), and [110](#).

- Gauthier S.; P., R.-N.; JA, M.; C., W. World alzheimer report 2021. *Journey through the diagnosis of dementia*. London, England: Alzheimer's Disease International, 2021. Cited 2 times in pages [9](#) and [111](#).
- GHAZI, M. M.; NIELSEN, M.; PAI, A.; CARDOSO, M. J.; MODAT, M.; OURSELIN, S.; SØRENSEN, L. Training recurrent neural networks robust to incomplete data: Application to alzheimer's disease progression modeling. *Medical Image Analysis*, Elsevier BV, v. 53, p. 39–46, apr 2019. Disponível em: <https://doi.org/10.1016%2Fj.media.2019.01.004>. Cited 5 times in pages [15](#), [16](#), [18](#), [103](#), and [104](#).
- GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. *Deep Learning*. [S.l.]: MIT Press, 2016. <http://www.deeplearningbook.org>. Cited 2 times in pages [13](#) and [14](#).
- GOUBRAN, M.; NTIRI, E. E.; AKHAVEIN, H.; HOLMES, M.; NESTOR, S.; RAMIREZ, J.; ADAMO, S.; OZZOUDE, M.; SCOTT, C.; GAO, F. *et al. Hippocampal segmentation for brains with extensive atrophy using three-dimensional convolutional neural networks*. [S.l.], 2020. Cited in page [51](#).
- HAN, R.; CHEN, C. L. P.; LIU, Z. A novel convolutional variation of broad learning system for alzheimer's disease diagnosis by using mri images. *IEEE Access*, v. 8, p. 214646–214657, 2020. Cited 3 times in pages [16](#), [18](#), and [110](#).
- HART, P. E.; STORK, D. G.; DUDA, R. O. *Pattern classification*. [S.l.]: Wiley Hoboken, 2000. Cited 2 times in pages [28](#) and [29](#).
- HAZARIKA, R. A.; KANDAR, D.; MAJI, A. K. A deep convolutional neural networks based approach for alzheimer's disease and mild cognitive impairment classification using brain images. *IEEE Access*, v. 10, p. 99066–99076, 2022. Cited 3 times in pages [16](#), [19](#), and [110](#).
- HE, H.; BAI, Y.; GARCIA, E. A.; LI, S. Adasyn: Adaptive synthetic sampling approach for imbalanced learning. p. 1322–1328, 2008. Cited in page [26](#).
- HE, K.; ZHANG, X.; REN, S.; SUN, J. *Deep Residual Learning for Image Recognition*. arXiv, 2015. Disponível em: <https://arxiv.org/abs/1512.03385>. Cited in page [71](#).
- HOQUE, N.; BHATTACHARYYA, D.; KALITA, J. Mifs-nd: A mutual information-based feature selection method. *Expert Systems with Applications*, v. 41, n. 14, p. 6371–6385, 2014. ISSN 0957-4174. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0957417414002164>. Cited in page [28](#).
- HUGHES C. P., B. L. D. W. L. C. L. A. . M. R. L. A new clinical scale for the staging of dementia. *The British journal of psychiatry : the journal of mental science*, Vol. 140, p. 566–572, 1982. Cited 2 times in pages [4](#) and [20](#).
- IDDI, S.; D, L.; PS, A.; MS, R.; WK, T.; MC, D. Predicting the course of alzheimer's progression. *Brain Informatics*, Vol. 6(1), 2019. Cited 5 times in pages [15](#), [16](#), [17](#), [103](#), and [104](#).
- JIN, X.; XU, A.; BIE, R.; GUO, P. Machine learning techniques and chi-square feature selection for cancer classification using sage gene expression profiles. In: SPRINGER. *Data Mining for Biomedical Applications: PAKDD 2006 Workshop, BioDM 2006, Singapore, April 9, 2006. Proceedings*. [S.l.], 2006. p. 106–115. Cited in page [28](#).

KHAN, P.; KADER, M. F.; ISLAM, S. M. R.; RAHMAN, A. B.; KAMAL, M. S.; TOHA, M. U.; KWAK, K.-S. Machine learning and deep learning approaches for brain disease diagnosis: Principles and recent advances. *IEEE Access*, v. 9, p. 37622–37655, 2021. Cited 6 times in pages [15](#), [16](#), [17](#), [20](#), [102](#), and [104](#).

KNOPMAN D. S., P. R. C. . J. C. R. J. A brief history of "alzheimer disease". *Neurology*, Vol. 92(22), p. 1053–1059, 2019. Cited in page [6](#).

KRIZHEVSKY, A.; SUTSKEVER, I.; HINTON, G. E. Imagenet classification with deep convolutional neural networks. In: PEREIRA, F.; BURGESS, C.; BOTTOU, L.; WEINBERGER, K. (Ed.). *Advances in Neural Information Processing Systems*. Curran Associates, Inc., 2012. v. 25. Disponível em: <https://proceedings.neurips.cc/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf>. Cited in page [71](#).

KURSA, M. B.; RUDNICKI, W. R. Feature selection with the boruta package. *Journal of statistical software*, v. 36, p. 1–13, 2010. Cited in page [28](#).

LAROBINA, M.; MURINO, L. Medical image file formats. *Journal of digital imaging*, 12 2013. Cited in page [55](#).

LAVALLEY, M. Logistic regression. *Circulation*, v. 117, p. 2395–9, 06 2008. Cited in page [48](#).

LECUN, Y.; BENGIO, Y.; HINTON, G. Deep learning. *Nature*, v. 521, p. 436–44, 05 2015. Cited in page [70](#).

LECUN, Y.; BOTTOU, L.; BENGIO, Y.; HAFFNER, P. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, v. 86, n. 11, p. 2278–2324, 1998. Cited in page [70](#).

LEDIG, C.; SCHUH, A.; GUERRERO, R.; HECKEMANN, R. A.; RUECKERT, D. Structural brain imaging in alzheimer's disease and mild cognitive impairment: biomarker analysis and shared morphometry database. *Scientific reports*, Nature Publishing Group, v. 8, n. 1, p. 1–16, 2018. Cited in page [50](#).

LI, J.; WEI, Y.; WANG, C.; HU, Q.; LIU, Y.; XU, L. 3-d cnn-based multichannel contrastive learning for alzheimer's disease automatic diagnosis. *IEEE Transactions on Instrumentation and Measurement*, v. 71, p. 1–11, 2022. Cited 3 times in pages [16](#), [18](#), and [110](#).

LI, X.; MORGAN, P. S.; ASHBURNER, J.; SMITH, J.; RORDEN, C. The first step for neuroimaging data analysis: Dicom to nifti conversion. *Journal of Neuroscience Methods*, v. 264, p. 47–56, 2016. ISSN 0165-0270. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0165027016300073>. Cited in page [55](#).

LIN, M.; CHEN, Q.; YAN, S. Network in network. *CoRR*, abs/1312.4400, 2013. Disponível em: <https://api.semanticscholar.org/CorpusID:16636683>. Cited in page [107](#).

LIU J LI, P. C. X. Modeling disease progression with deep neural networks. *The Fourth International Symposium on Image Computing and Digital Medicine (ISICDM 2020)*, p. 32–34, 2020. Cited 4 times in pages [16](#), [18](#), [103](#), and [104](#).

- LUNDBERG, S. M.; LEE, S.-I. A unified approach to interpreting model predictions. *Advances in neural information processing systems*, v. 30, 2017. Cited in page [28](#).
- MARCUS, D. S.; WANG, T. H.; PARKER, J.; CSERNANSKY, J. G.; MORRIS, J. C.; BUCKNER, R. L. Open Access Series of Imaging Studies (OASIS): Cross-sectional MRI Data in Young, Middle Aged, Nondemented, and Demented Older Adults. *Journal of Cognitive Neuroscience*, v. 19, n. 9, p. 1498–1507, 09 2007. ISSN 0898-929X. Disponível em: <https://doi.org/10.1162/jocn.2007.19.9.1498>. Cited 3 times in pages [4](#), [20](#), and [23](#).
- MARINESCU, R. V.; OXTOBY, N. P.; YOUNG, A. L.; BRON, E. E.; TOGA, A. W.; WEINER, M. W.; BARKHOF, F.; FOX, N. C.; KLEIN, S.; ALEXANDER, D. C.; CONSORTIUM, t. E. Tadpole challenge: Prediction of longitudinal evolution in alzheimer's disease. *F1000Research*, 2018. Cited 4 times in pages [4](#), [22](#), [23](#), and [56](#).
- MELBOURNE, T. U. of. *QUANTUM BOOST FOR MEDICAL IMAGING*. 2018. URL: <https://pursuit.unimelb.edu.au/articles/quantum-boost-for-medical-imaging>. Cited 2 times in pages [53](#) and [54](#).
- MILLER-THOMAS, M. M.; SIPE, A. L.; BENZINGER, T. L. S.; MCCONATHY, J.; CONNOLLY, S.; SCHWETYE, K. E. Multimodality review of amyloid-related diseases of the central nervous system. *RadioGraphics*, v. 36, n. 4, p. 1147–1163, 2016. PMID: 27399239. Disponível em: <https://doi.org/10.1148/rg.2016150172>. Cited 2 times in pages [49](#) and [58](#).
- MOORE, P. J.; LYONS, T. J.; GALLACHER, J.; INITIATIVE, A. D. N. Random forest prediction of alzheimer's disease using pairwise selection from time series data. *Plos One*, Vol. 14(2), 2019. Cited 5 times in pages [15](#), [16](#), [17](#), [103](#), and [104](#).
- NGUYEN, M.; SUN, N.; ALEXANDER, D. C.; FENG, J.; YEO, B. T. Modeling alzheimer's disease progression using deep recurrent neural networks. p. 1–4, 2018. Cited 6 times in pages [15](#), [16](#), [18](#), [44](#), [103](#), and [104](#).
- NGUYEN M., H. T. A. L. A. D. C. F. J. Y. B. . A. D. N. I. Predicting alzheimer's disease progression using deep recurrent neural networks. *NeuroImage*, Vol. 222, 2020. Cited 5 times in pages [15](#), [16](#), [17](#), [103](#), and [104](#).
- NIYAS, M.; P, T. Alzheimer's classification using dynamic ensemble of classifiers selection algorithms: A performance analysis. *Biomedical Signal Processing and Control*, v. 68, p. 102729, 2021. ISSN 1746-8094. Disponível em: <https://www.sciencedirect.com/science/article/pii/S1746809421003268>. Cited 5 times in pages [15](#), [16](#), [17](#), [103](#), and [104](#).
- NIYAS, M.; P, T. Feature selection using efficient fusion of fisher score and greedy searching for alzheimer's classification. *Journal of King Saud University - Computer and Information Sciences*, 2021. Cited 10 times in pages [15](#), [16](#), [17](#), [20](#), [23](#), [24](#), [29](#), [44](#), [103](#), and [104](#).
- NORFRAY, J. F.; PROVENZALE, J. M. Alzheimer's disease: Neuropathologic findings and recent advances in imaging. *American Journal of Roentgenology*, v. 182, n. 1, p. 3–13, 2004. PMID: 14684506. Disponível em: <https://doi.org/10.2214/ajr.182.1.1820003>. Cited in page [49](#).
- ONU. World population prospects 2022: Summary of results. *United Nations Department of Economic and Social Affairs, Population Division*, UN DESA/POP/2022/TR/NO. 3, 2022. Cited in page [7](#).

- PAIVA, N.; ESCOVEDO, T. Detecção precoce de alzheimer usando machine learning. *XV Congresso Brasileiro de Inteligência Computacional*, 2021. Disponível em: https://sbic.org.br/eventos/cbic_2021/. Cited 3 times in pages 24, 104, and 105.
- PARMERA, J. B.; NITRINI, R. Demências: da investigação ao diagnóstico. *Revista de Medicina*, v. 94, n. 3, p. 179–184, dez. 2015. Disponível em: <https://www.revistas.usp.br/revistadc/article/view/108748>. Cited in page 8.
- PARVANDEH, S.; YEH, H.-W.; PAULUS, M.; MCKINNEY, B. Consensus features nested cross-validation. *Bioinformatics (Oxford, England)*, v. 36, 01 2020. Cited 2 times in pages 42 and 43.
- PATEL, K. P.; WYMER, D. T.; BHATIA, V. K.; DUARA, R.; RAJADHYAKSHA, C. D. Multimodality imaging of dementia: Clinical importance and role of integrated anatomic and molecular imaging. *RadioGraphics*, v. 40, n. 1, p. 200–222, 2020. PMID: 31917652. Disponível em: <https://doi.org/10.1148/rg.2020190070>. Cited in page 49.
- RASCHKA, S. *STAT 479: Machine Learning - Lecture Notes*. [S.l.]: Department of Statistics - University of Wisconsin, 2018. <http://stat.wisc.edu/~sraschka/teaching/stat479-fs2018/>. Cited in page 47.
- RAZZAK, I.; NAZ, S.; ALINEJAD-ROKNY, H.; NGUYEN, T.; KHALIFA, F. A cascaded multiresolution ensemble deep learning framework for large scale alzheimer’s disease detection using brain mris. *IEEE/ACM transactions on computational biology and bioinformatics*, PP, 11 2022. Cited 3 times in pages 16, 19, and 110.
- REPORT, A. A. Alzheimer’s disease facts and figures. *Alzheimer’s dementia : the journal of the Alzheimer’s Association*, Vol. 17(3), p. 327–406, 2021. Cited in page 7.
- RISH, I. An empirical study of the naïve bayes classifier. *IJCAI 2001 Work Empir Methods Artif Intell*, v. 3, 01 2001. Cited in page 48.
- ROGACHESKI, ; MAZER, S.; RODRIGUES, D.; BITTENCOURT, P. R. M. D. Análises visual e volumétrica por ressonância magnética das formações hipocampais em um grupo de pacientes com diagnóstico clínico de epilepsia do lobo temporal. *Arquivos de Neuro-Psiquiatria*, Academia Brasileira de Neurologia - ABNEURO, v. 56, n. Arq. Neuro-Psiquiatr., 1998 56(3A), Sep 1998. ISSN 0004-282X. Disponível em: <https://doi.org/10.1590/S0004-282X1998000300011>. Cited in page 50.
- SAEYS, Y.; ABEEL, T.; PEER, Y. Van de. Robust feature selection using ensemble feature selection techniques. In: SPRINGER. *Machine Learning and Knowledge Discovery in Databases: European Conference, ECML PKDD 2008, Antwerp, Belgium, September 15-19, 2008, Proceedings, Part II 19*. [S.l.], 2008. p. 313–325. Cited in page 29.
- SALIH, Y.; MD-ESA, W. bt.; MALIK, A. S.; SAAD, N. Tone mapping of hdr images: A review. In: *2012 4th International Conference on Intelligent and Advanced Systems (ICIAS2012)*. [S.l.: s.n.], 2012. v. 1, p. 368–373. Cited in page 65.
- SARAIVA JR., R. G.; CARVALHO, J. L. A.; PEIXOTO, E.; NASCIMENTO, F. A. de O. Seleção simplificada de atributos para auxílio ao diagnóstico da doença de alzheimer utilizando aprendizado de máquina. *XXVIII Congresso Brasileiro de Engenharia de Engenharia Biomédica (CBEB 2022)*, 2022. Cited in page 104.

SCHACHTER MD; KENNETH L. DAVIS, M. A. S. Alzheimer's disease. *Dialogues in Clinical Neuroscience*, Vol. 2(2), p. 91–100, 2000. Cited in page [6](#).

SCHAPIRE, R. E. Explaining adaboost. In: *Empirical inference*. [S.l.]: Springer, 2013. p. 37–52. Cited in page [48](#).

SEIJO-PARDO, B.; BOLÓN-CANEDO, V.; ALONSO-BETANZOS, A. On developing an automatic threshold applied to feature selection ensembles. *Information Fusion*, Elsevier, v. 45, p. 227–245, 2019. Cited in page [28](#).

SHAFER, J. C.; AGRAWAL, R.; MEHTA, M. Sprint: A scalable parallel classifier for data mining. In: *Proceedings of the 22th International Conference on Very Large Data Bases*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1996. (VLDB '96), p. 544–555. ISBN 1558603824. Cited in page [48](#).

SHARMA, R.; GOEL, T.; TANVEER, M.; SUGANTHAN, P.; RAZZAK, I.; MURUGAN, R. Conv-ervfl: Convolutional neural network based ensemble rvfl classifier for alzheimer's disease diagnosis. *IEEE Journal of Biomedical and Health Informatics*, IEEE, 2022. Cited 3 times in pages [16](#), [19](#), and [110](#).

SIMONYAN, K.; ZISSERMAN, A. *Very Deep Convolutional Networks for Large-Scale Image Recognition*. arXiv, 2014. Disponível em: <https://arxiv.org/abs/1409.1556>. Cited in page [71](#).

SWAFFE, K. *What the hell happened to my brain? Living Beyond Dementia*. London: Jessica Kingsley Publishers, 2016. 391 p. ISBN 978-1-84905-6-083. Cited in page [1](#).

SZEGEDY, C.; VANHOUCKE, V.; IOFFE, S.; SHLENS, J.; WOJNA, Z. *Rethinking the Inception Architecture for Computer Vision*. arXiv, 2015. Disponível em: <https://arxiv.org/abs/1512.00567>. Cited in page [71](#).

TAHIRA, A. C.; VERJOVSKI-ALMEIDA, S.; FERREIRA, S. T. Dementia is an age-independent risk factor for severity and death in covid-19 inpatients. *Alzheimer's & Dementia*, v. 17, n. 11, p. 1818–1831, 2021. Disponível em: <https://alz-journals.onlinelibrary.wiley.com/doi/abs/10.1002/alz.12352>. Cited in page [9](#).

TAN, M.; LE, Q. *EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks*. 2019. Cited in page [71](#).

TEIPEL, S. J.; MEINDL, T.; GRINBERG, L.; HEINSEN, H.; HAMPEL, H. Novel mri techniques in the assessment of dementia. *European journal of nuclear medicine and molecular imaging*, Springer, v. 35, p. 58–69, 2008. Cited in page [108](#).

URBANOWICZ, R. J.; MEEKER, M.; CAVA, W. L.; OLSON, R. S.; MOORE, J. H. Relief-based feature selection: Introduction and review. *Journal of biomedical informatics*, Elsevier, v. 85, p. 189–203, 2018. Cited in page [28](#).

VARMA, S.; SIMON, R. Bias in error estimation when using cross-validation for model selection. *BMC Bioinformatics*, v. 7, 02 2006. Cited in page [43](#).

WELLER J., B. A. Current understanding of alzheimer's disease diagnosis and treatment. *F1000Research*, Vol. 7, 2018. Cited in page [1](#).

WU, Y.; ZHOU, Y.; ZENG, W.; QIAN, Q.; SONG, M. An attention-based 3d cnn with multi-scale integration block for alzheimer's disease classification. *IEEE Journal of Biomedical and Health Informatics*, IEEE, v. 26, n. 11, p. 5665–5673, 2022. Cited 3 times in pages [16](#), [19](#), and [110](#).

YAMASHITA, R.; NISHIO, M.; DO, R.; TOGASHI, K. Convolutional neural networks: an overview and application in radiology. *Insights into Imaging*, v. 9, 06 2018. Cited in page [70](#).

YANG H. D., K. D. H. L. S. B. . Y. L. D. History of alzheimer's disease. *Dementia and neurocognitive disorders*, Vol. 15(4), p. 115–121, 2016. Cited in page [5](#).

ZHANG, D.; SHEN, D.; INITIATIVE, A. D. N. Multi-modal multitask learning for joint prediction of multiple regression and classification variables in alzheimer's disease. *NeuroImage*, Vol. 59(2), p. 895–907, 2011. Cited 2 times in pages [15](#) and [16](#).

ZHANG, Z.-C.; ZHAO, X.; DONG, G.; ZHAO, X.-M. Improving alzheimer's disease diagnosis with multi-modal pet embedding features by a 3d multi-task mlp-mixer neural network. *IEEE Journal of Biomedical and Health Informatics*, IEEE, 2023. Cited 3 times in pages [16](#), [19](#), and [110](#).

ZHOU, B.; KHOSLA, A.; LAPEDRIZA, A.; OLIVA, A.; TORRALBA, A. *Learning Deep Features for Discriminative Localization*. 2015. Cited in page [108](#).

APPENDIX A

PORTUGUESE THESIS

TESE DE DOUTORADO

"INTELIGÊNCIA COMPUTACIONAL NO APOIO AO DIAGNÓSTICO DA
DOENÇA DE ALZHEIMER"

** Tabelas e figuras foram mantidas com o texto em inglês **



**INTELIGÊNCIA COMPUTACIONAL NO
APOIO AO DIAGNÓSTICO DA DOENÇA
DE ALZHEIMER**

RAIMUNDO GUIMARÃES SARAIVA JUNIOR

TESE DE DOUTORADO EM ENGENHARIA ELÉTRICA

DEPARTAMENTO DE ENGENHARIA ELÉTRICA

FACULDADE DE TECNOLOGIA

UNIVERSIDADE DE BRASÍLIA

Universidade de Brasília
Faculdade de Tecnologia
Departamento de Engenharia Elétrica

Inteligência computacional no apoio ao diagnóstico da doença de
Alzheimer

Raimundo Guimarães Saraiva Junior

TESE DE DOUTORADO EM ENGENHARIA ELÉTRICA SUBMETIDA AO
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA DA UNI-
VERSIDADE DE BRASÍLIA COMO PARTE DOS REQUISITOS NECESSÁRIOS
PARA A OBTENÇÃO DO GRAU DE DOUTOR.

APROVADA POR:

Prof. Francisco A. O. Nascimento, D.Sc. (ENE-UnB)
(Orientador)

Prof. Henrique Sarmiento Malvar, Ph.D. (University of Washington)
(Examinador externo)

Prof. José Antonio Apolinário Jr., D.Sc. (IME)
(Examinador externo)

Prof. Daniel Guerreiro e Silva, D.Sc. (ENE-UnB)
(Examinador interno)

Brasília/DF, dezembro de 2023.

FICHA CATALOGRÁFICA

SARAIVA-JUNIOR, RAIMUNDO GUIMARÃES

Inteligência computacional no apoio ao diagnóstico da doença de Alzheimer. [Distrito Federal] 2023.

xxx, nnp., 210 x 297 mm (ENE/FT/UnB, Doutor, Tese de doutorado, 2023).

Universidade de Brasília, Faculdade de Tecnologia, Departamento de Engenharia Elétrica.

Departamento de Engenharia Elétrica

- | | |
|---------------------------|----------------------------|
| 1. Alzheimer | 2. Demência |
| 3. Diagnóstico | 4. Inteligência Artificial |
| 5. Aprendizado de máquina | 6. Aprendizado profundo |
| 7. Ressonância magnética | |
| I. ENE/FT/UnB | II. Título (série) |

REFERÊNCIA BIBLIOGRÁFICA

SARAIVA-JUNIOR, R. G. (2023). Inteligência computacional no apoio ao diagnóstico da doença de Alzheimer. Tese de doutorado em Engenharia Elétrica, Publicação PPGEE.202/2023, Departamento de Engenharia Elétrica, Universidade de Brasília, Brasília, DF, xxxxp.

CESSÃO DE DIREITOS

AUTOR: Raimundo Guimarães Saraiva Junior

TÍTULO: Inteligência computacional no apoio ao diagnóstico da doença de Alzheimer.

GRAU: Doutor ANO: 2023

É concedida à Universidade de Brasília permissão para reproduzir cópias desta tese e para emprestar ou vender tais cópias somente para propósitos acadêmicos e científicos. O autor reserva outros direitos de publicação e nenhuma parte desta tese pode ser reproduzida sem autorização por escrito do autor.

Raimundo Guimarães Saraiva Junior

Universidade de Brasília (UnB)

Campus Darcy Ribeiro

Faculdade de Tecnologia - FT

Departamento de Engenharia Elétrica (ENE)

Brasília - DF - CEP: 70919-970

Dedico esta tese especialmente ao meu irmão José Flávio Sombra Saraiva, professor titular aposentado da UnB, que recebeu tão precocemente o diagnóstico de Alzheimer.

Ao meu filho, Eduardo, aos meus irmãos, Gilson e Ana Gláucia, à minha mãe, Maria Alice, ao meu pai, Raimundo Saraiva (in memoriam), e a Kaline Kelle, uma pessoa especial.

AGRADECIMENTOS

Agradeço ao meu orientador, Francisco Assis de Oliveira Nascimento, pelos conselhos, apoio e confiança depositada. À coordenação e demais colegas do PPGEE/UnB, pelas pequenas e grandes contribuições, particularmente ao colega Tony Alexandre Medeiros.

Com relação aos temas da medicina, esta tese contou com os preciosos aconselhamentos do médico e pesquisador Yuri Almeida de Oliveira, a quem agradeço pela atenção e por todas as informações prestadas.

Também agradeço aos colegas do Controle e Processos Industriais, do Instituto Federal de Educação, Ciência e Tecnologia do Ceará - Limoeiro do Norte, em particular ao coordenador, José Gesival da Macena, em nome do qual cumprimento todos os colegas pelo apoio recebido.

Aos professores Geovany Araújo Borges e Lucas Vieira Barros, pelo incentivo e apoio ao processo de ingresso no PPGEE/UnB.

Aos colegas da confraria Caríssimos, por acompanharem sempre de perto as nossas intenções, com muita atenção e carinho.

Agradecimento todo especial a Kaline Kelle Lima de Oliveira, pelo apoio inestimável prestado em todos os momentos da elaboração desta tese.

Finalmente, agradeço a disponibilização de dados primários pelos estudos *Australian Imaging, Biomarker & Lifestyle Flagship Study of Ageing* (AIBL), a *Alzheimers Disease Neuroimaging Initiative* (ADNI), a *Open Access Series of Imaging Studies* (OASIS) e a *University of Southern California* pela autorização de uso da plataforma *Image Data Archive* (IDA).

RESUMO

O diagnóstico preciso da doença de Alzheimer ainda é um campo aberto para a ciência, tendo em vista se tratar de uma doença multifatorial e com difícil identificação das causas. As demências estão cada vez mais presentes na população mundial, que experimenta um aumento significativo na expectativa de vida, o que explica, em parte, o aumento do diagnóstico de Alzheimer, que é o tipo de demência mais frequente. A decadência neurológica faz parte do processo natural do envelhecimento, mas a doença de Alzheimer acelera o processo de degeneração neurológica, trazendo grandes transtornos ao paciente e seus familiares, onerando os serviços públicos de saúde e causando óbitos precoces.

Nesse contexto, este trabalho investiga técnicas de inteligência artificial que busquem auxiliar no diagnóstico da doença de Alzheimer, utilizando dados de natureza clínica e imagens de ressonância magnética. Mediante licenciamento para utilização, foram buscadas informações primárias nos estudos científicos *Australian Imaging, Biomarker & Lifestyle Flagship Study of Ageing* (AIBL), *Alzheimers Disease Neuroimaging Initiative* (ADNI) e no *Open Access Series of Imaging Studies* (OASIS). Para tratamento dos dados clínicos, foi proposto um novo método de seleção de atributos, foram criadas bases de dados reduzidas e fez-se a avaliação por modelos de classificação de aprendizado de máquina, tendo atingido a acurácia de 99,81% no diagnóstico do Alzheimer para os dados ADNI. Também foi investigado o uso de imagens obtidas em exames de ressonância magnética. Foram propostas diferentes estratégias de construção de um banco de imagens adequadas para a aplicação de técnicas de aprendizado profundo, tendo atingido a acurácia de 99,3% no diagnóstico do Alzheimer para os dados ADNI. Métricas de avaliação foram computadas sobre as estratégias propostas, e resultados de diagnóstico foram obtidos para os casos de indivíduo normal e com Alzheimer estabelecido. Também foi avaliado o caso de indivíduos com comprometimento cognitivo leve. Como parte da avaliação de desempenho, os resultados alcançados foram comparados com os resultados obtidos em outros trabalhos de pesquisa disponíveis na literatura.

Palavras-chave: Alzheimer; Aprendizado de máquina; Seleção de atributos; Aprendizado profundo; Múltiplas instâncias.

ABSTRACT

Accurate diagnosis of Alzheimers disease remains difficult because of the complexity of the disease and its similarity to other forms of dementia. The prevalence of dementia, and its most common form Alzheimers, is increasing worldwide, partly due to increasing longevity. While neurological deterioration is a normal part of aging, Alzheimers accelerates this process, straining public health services, leading to early mortality, and generating tremendous challenges for patients and their families.

We used clinical data and magnetic resonance images to investigate artificial intelligence techniques that could assist in Alzheimers disease diagnosis. Primary information was obtained from the Australian Imaging, the Biomarker & Lifestyle Flagship Study of Ageing (AIBL), the Alzheimers Disease Neuroimaging Initiative (ADNI), and the Open Access Series of Imaging Studies (OASIS). To process these clinical data, we developed a new feature selection method, created reduced databases, evaluated the data via machine learning classification models and achieved 99.81% accuracy in diagnosing Alzheimer's for ADNI data. MRI images were also investigated. Different strategies were proposed for building an image dataset suitable for deep learning techniques and achieved 99.3% accuracy in diagnosing Alzheimer's for ADNI data. Evaluation metrics were computed for the proposed strategies, and diagnostic results were generated for normal and established Alzheimers cases. Individuals with mild cognitive impairment were also evaluated. Finally, the results of the current study were compared to those from other studies.

Keywords: Alzheimers disease, Machine learning, Feature selection, Deep learning, Multiple instances.

SUMÁRIO

Tabela de conteúdos	i
Lista de figuras	v
Lista de tabelas	ix
Glossário	xii
Capítulo 1 – Introdução	iii
1.1 A doença de Alzheimer	iii
1.1.1 Estágios da doença	3
1.1.2 Breve histórico da doença	6
1.1.3 O diagnóstico da doença	8
1.1.4 Alzheimer no Brasil	9
1.1.5 Alzheimer e a Covid-19	9
1.2 Justificativa para investigação do tema	10
1.3 Objetivos	10
1.4 Organização da tese	12
Capítulo 2 – Fundamentos teóricos	13
2.1 Inteligência artificial no diagnóstico de doenças neurodegenerativas	13
2.2 Auxílio computacional no diagnóstico da doença de Alzheimer	15
2.3 Revisão da literatura	17
Capítulo 3 – Análise baseada em dados clínicos	21
3.1 Detalhamento das bases de dados utilizadas	21
3.1.1 Estudo OASIS	21
3.1.2 Estudo AIBL	22
3.1.3 Estudo ADNI	23
3.1.4 Pré-processamento	24
3.2 Análise dos dados	27

3.3	Seleção de atributos	28
3.3.1	Métodos de filtro	29
3.3.2	Métodos de invólucro	30
3.3.3	Métodos incorporados	30
3.3.4	Algoritmo proposto	31
3.3.4.1	O algoritmo FMFS	32
3.3.4.2	Aplicação do FMFS às bases de dados	36
3.4	Métricas de avaliação de desempenho	40
3.4.0.1	Acurácia - A_{cy}	40
3.4.0.2	Precisão - P_{rn}	41
3.4.0.3	Sensibilidade - S_{ny}	41
3.4.0.4	Especificidade - S_{py}	41
3.4.0.5	Acurácia balanceada - B_{ca}	42
3.4.0.6	Pontuação f1 - $F1$	42
3.4.0.7	Área sobre a curva ROC - AUC	42
3.5	Validação cruzada aninhada	43
3.6	Modelos de ML utilizados	46
3.6.1	SVM : support vector machine	46
3.6.2	kNN : k-nearest neighbors	49
3.6.3	Outros modelos de ML	51
Capítulo 4 – Análise baseada em imagens		52
4.1	Diagnóstico de AD baseado em imagens de MRI	52
4.2	Imagens de ressonância magnética	56
4.2.1	Formato digital das imagens de ressonância magnética	57
4.2.1.1	Formato NIFTI	58
4.3	Detalhamento da base de dados utilizada	59
4.3.1	Estudo ADNI	59
4.3.1.1	MRI no ADNI	60
4.3.1.2	Escolha dos <i>slíces</i> para análise	61
4.4	Avaliação no plano axial	62
4.4.1	Algoritmo de seleção de <i>slíces</i> proposto	63
4.4.1.1	O algoritmo SMR	63
4.4.1.2	Aplicação do SMR na base de dados	66
4.4.2	Banco de imagens resultante	67
4.5	Avaliação no plano coronal	68
4.5.1	Estratégia para construção dos bancos de imagens	68

4.5.2	Avaliação por múltiplas instâncias	70
4.5.3	Bancos de imagens resultantes	72
4.6	Modelos de DL utilizados	74
4.6.1	Modelos clássicos de CNNs	75
4.6.2	Modelos propostos de CNNs	76

Capítulo 5 – Resultados e discussão **83**

5.1	Resultados de análise por dados clínicos	83
5.1.1	Seleção de atributos para a base AIBL	84
5.1.2	Seleção de atributos para a base ADNI	84
5.1.3	Seleção de atributos para a base OASIS	85
5.1.4	Avaliação do desempenho para três classes	85
5.1.4.1	Avaliação ADNI	85
5.1.4.2	Avaliação AIBL	86
5.1.4.3	Avaliação OASIS	88
5.1.5	Avaliação do desempenho para duas classes	88
5.1.5.1	Avaliação ADNI	89
5.1.5.2	Avaliação AIBL	90
5.1.5.3	Avaliação OASIS	91
5.1.6	Avaliação do desempenho com outros classificadores	92
5.1.7	Avaliação com outros métodos de seleção de atributos	94
5.2	Resultados da análise por imagem	96
5.2.1	Análise do banco de imagens coronal	96
5.2.1.1	Alternativas de montagem das partições	96
5.2.1.2	Avaliação da partição S1	98
5.2.1.3	Avaliação da partição S2	100
5.2.1.4	Reavaliação da partição S2 com aumento de dados	101
5.2.1.5	Reavaliação de S1 com classificador de votação por majoritária	101
5.2.2	Análise do banco de imagens axial	103
5.3	Discussão dos resultados da análise sobre dados clínicos	104
5.3.1	Comparação com outras pesquisas	106
5.3.2	Outras considerações sobre a análise em dados clínicos	109
5.4	Discussão dos resultados da análise sobre as imagens	110
5.4.1	Plano coronal	111
5.4.2	Plano axial	114
5.4.3	Comparação com outras pesquisas	115
5.5	Sugestão de aplicação das novas tecnologias	116

5.5.1	Sistema Único de Saúde brasileiro - SUS	117
5.5.2	Premio Inovação Engenharia Biomédica para o SUS 2023	117
5.5.3	Proposta vitoriosa	118
Capítulo 6 – Conclusões		120
Referências		122

LISTA DE FIGURAS

1.1	Types of dementia.	0
1.2	Images with several medial temporal atrophy scores (MTA).	3
1.3	Evolution of AD diagnosis in Olmsted County, Minnesota - USA.	6
1.4	Life expectancy at birth: country estimates, 1950-2021, and projections, 2022-2050	7
1.5	PubMed portal search results with the word "Alzheimer's"from 1980 to September 2022.	9
2.1	A Venn diagram showing how deep learning is a kind of representation learning, which is in turn a kind of machine learning that is used for many but not all approaches to AI.	14
2.2	Article distributions with respect to ML / DL.	16
2.3	Databases used in studies on AD diagnosis.	16
3.1	Feature selection methods.	29
3.2	Class histograms for the LIMMTOTAL(a) and AXT117(b) features in the AIBL dataset.	32
3.3	Nomenclature definition.	33
3.4	Features with different class means.	34
3.5	FMFS flowchart.	35
3.6	AIBL feature ranking - <i>dca</i> values.	37

3.7	AIBL performance measurement. a) The orange and blue curves represent the evolution of SVM and kNN accuracy, respectively, with indicated maximums. b) The green curve represents the evolution of the number of selected features. The orange and blue vertical lines represent the point of maximum accuracy for SVM and kNN respectively.	37
3.8	ADNI feature ranking - <i>dca</i> values.	38
3.9	ADNI performance measurement. a) The orange and blue curves represent the evolution of SVM and kNN accuracy, respectively, with indicated maximums. b) The green curve represents the evolution of the number of selected features. The orange and blue vertical lines represent the point of maximum accuracy for SVM and kNN respectively.	38
3.10	OASIS feature ranking - <i>dca</i> values.	39
3.11	OASIS performance measurement. a) The orange and blue curves represent the evolution of SVM and kNN accuracy, respectively, with indicated maximums. b) The green curve represents the evolution of the number of selected features. The orange and blue vertical lines represent the point of maximum accuracy for SVM and kNN respectively.	39
3.12	Confusion matrix.	40
3.13	Receiver Operating Characteristics (ROC) curve.	43
3.14	Train, validation and test data split.	44
3.15	Nested cross validation.	45
3.16	Generic machine learning process.	46
3.17	SVM algorithm - the support vectors, marked with grey squares, define the margin of largest separation between the two classes.	48
3.18	kNN algorithm - example with three classes and $k = 5$	50
4.1	Illustration of the hippocampus in the brain.	53
4.2	Three examples of MRI images (brain-extracted) of subjects from the ADNI (coronal slices).	54

4.3	Proposed hippocampus extraction scheme.	54
4.4	Extracted hippocampus.	54
4.5	Visualization of the hippocampus from various angles.. . . .	55
4.6	Maximum dimensions of the hippocampus in the 3 planes. Right hippocampus rotated 300 degrees.	56
4.7	MRI equipment (creative commons license BY-ND).	57
4.8	Orange lines representing axial MRI slices.	57
4.9	Axial (red bar), coronal (green bar) and sagittal (yellow bar) slice views of an MRI in NIFTI format.	59
4.10	Anatomy of the hippocampal formation: 1 = hippocampal head, 2 = hippocampal body, 3 = hippocampal tail, 4 = mesencephalon.	62
4.11	Axial slice showing the hippocampus.	63
4.12	Image matching.	64
4.13	Example SMR result. The red line in (a) shows the position of the selected slice with respect to the patients head. The range of interest on the mt curve (b) for the slice search is outlined in yellow. The image of the selected slice is shown in (c).	66
4.14	Template.	67
4.15	CN, MCI and AD image examples.	67
4.16	Red circles identify the left and right structures of the hippocampus in the coronal plane.	68
4.17	Preprocessing scheme	69
4.18	Coronal slices with 0.9442 mm spacing. Central, anterior, and posterior slices are highlighted.	70
4.19	Multiple instance decision scheme. In this case, a patient was evaluated and 11 instances indicated AD and 08 instances indicated CN. The majority decision is that the patient's diagnosis is AD.	71

4.20	Coronal image dataset of central slices.	72
4.21	Coronal dataset with a center slice, 9 anterior slices, and 9 posterior slices. The slice ending in -10 is the central slice chosen.	73
4.22	Generic deep learning process.	74
4.23	CNN Evolution - examples.	75
4.24	CNNs used.	76
4.25	CNN 1 model - total parameters: 4,534,866. Initial model.	77
4.26	CNN 2 model - total parameters: 717,034. Reduced dimensionality with respect to CNN-1.	78
4.27	CNN 3 model - total parameters: 18,421,666. Increased dimensionality with respect to CNN-1.	79
4.28	CNN 4 model - total parameters: 7,426,642. Increased depth compared to CNN-1.	80
4.29	CNN 5 model - total parameters: 1,786,090. Reduced dimensionality with respect to CNN-4.	81
5.1	ADNI-11 avaliated with the SVM classifier - average confusion matrix.	86
5.2	AIBL-4 avaliated with the SVM classifier - average confusion matrix.	87
5.3	OASIS-5 avaliated with the SVM classifier - average confusion matrix.	88
5.4	ADNI-11 avaliated with the SVM classifier over CN v/s AD classes - ROC curves.	89
5.5	ADNI-11 avaliated with the SVM classifier - CN v/s AD - average confusion matrix.	90
5.6	AIBL-4 avaliated with the SVM classifier - CN v/s AD - average confusion matrix.	91
5.7	OASIS-5 avaliated with the kNN classifier - C00 v/s C12 - average confusion matrix.	92
5.8	Split fold S1 and S2 description.	97
5.9	CNN-5 model learning curves. a) Accuracy model. b) AUC model. c) Loss model.	99
5.10	CNN-5 confusion matrix.	100

5.11	CNN-4 confusion matrix.	101
5.12	AIBL - kNN enhancement.	105
5.13	ADNI - kNN enhancement.	105
5.14	CN ADNI 002_S_1261 patient heat map and AD ADNI 007_S_1248 patient heat map. Red boxes highlight the region of interest.	113
5.15	Heat map of the 19 images of patient ADNI 007_S_1248 diagnosed with AD.	114
5.16	Biomedical Engineering Innovation Award for SUS 2023.	117

LISTA DE TABELAS

1.1	Clinical Dementia Rating (CDR) Scale	5
3.1	OASIS: Features descriptions	22
3.2	AIBL: Features descriptions	23
3.3	ADNI: Features descriptions	24
3.4	AIBL-18: dataset composition	25
3.5	ADNI-21: dataset composition	25
3.6	OASIS-8: dataset composition	25
3.7	AIBL-18: instance distribution by class	27
3.8	ADNI-21: instance distribution by class	27
3.9	OASIS-7: instance distribution by class	27
4.1	ADNI1:Screening 1.5T - Number of patients per class	61
5.1	Reduced AIBL dataset: selected features	84
5.2	Reduced ADNI dataset: selected features	84
5.3	Reduced OASIS dataset: selected features	85
5.4	ADNI dataset results - three classes	86
5.5	AIBL dataset results - three classes	87
5.6	OASIS dataset results - three classes	88
5.7	ADNI dataset results - two classes	89
5.8	AIBL dataset results - two classes	90

5.9	OASIS dataset results - two classes	91
5.10	AIBL-4 - performance evaluation	93
5.11	ADNI-11 - performance evaluation	93
5.12	OASIS-5 - performance evaluation	94
5.13	ADNI - feature selection method evaluation	94
5.14	AIBL - feature selection method evaluation	95
5.15	OASIS - feature selection method evaluation	95
5.16	Split fold composition S1 and S2	97
5.17	S1 split fold - performance evaluation	98
5.18	S2 split fold - performance evaluation	100
5.19	Data augmentation in S2 training dataset - performance evaluation	101
5.20	Customized cnn results	102
5.21	Pretrained models results	102
5.22	CNN-4 training results	103
5.23	Split axial dataset	104
5.24	Axial plan - performance evaluation	104
5.25	Comparison with previous study (AIBL dataset)	107
5.26	Comparison with previous studies (ADNI dataset)	108
5.27	Comparison with previous study (OASIS dataset)	109
5.28	Comparison with previous DL studies	116

GLOSSÁRIO

AD	Alzheimer's disease
AI	Artificial intelligence
AIBL	Australian Imaging, Biomarker & Lifestyle Flagship Study of Ageing
ADNI	Alzheimers Disease Neuroimaging Initiative
ANALYZE	Analyze 7.5
BCA	Balanced classification accuracy
CN	Cognitively normal
CSV	Comma-separated values
CDR	Clinical dementia rating
CNN	Convolutional Neural Networks
DL	Deep learning
DICOM	Digital Imaging and Communications in Medicine
GPU	Graphics Processing Unit
kNN	K-nearest neighbors

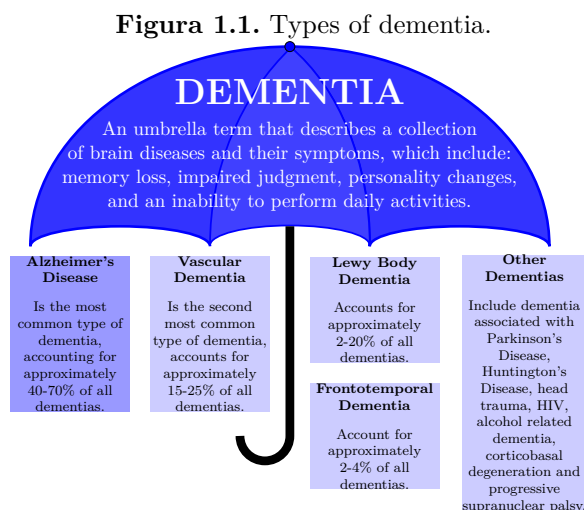
LSTM	Long short-term memory
MCI	Mild cognitive impairment
ML	Machine learning
MRI	Magnetic resonance imaging
MINC	Montreal Neurological Institute
MTA	Medial temporal lobe atrophy
NIFTI	Neuroimaging Informatics Technology Initiative
NRRD	Nearly Raw Raster Data
PET	Positron Emission Tomography
PAR/REC	Philips MRI scanner formats
RNN	Recurrent neural networks
RepL	Representation Learning
SVM	Support vector machine

Neste capítulo, é feita uma revisão histórica, contextualização e justificativa do problema que foi estudado nesta tese. Também são definidos os objetivos, a contribuição científica e apresentada a forma como este trabalho está organizado.

1.1 A DOENÇA DE ALZHEIMER

A doença de Alzheimer (AD, do inglês *Alzheimer's disease*) é compreendida como uma demência cognitiva progressiva, passando por um estágio de prejuízo cognitivo leve e atingindo a fase mais aguda, quando vai retirando do paciente a sua condição de independência e capacidade funcional (WELLER, J., 2018). A palavra demência vem do latim *dementia*, que significa falta de razão.

A AD é o tipo de demência mais frequente no conjunto de todas as demências conhecidas, chegando a alcançar 70% do total dos pacientes diagnosticados (SWAFFEE, 2016). A Figura 1.1 mostra o conjunto de demências mais conhecidas e a sua predominância em todo o espectro das demências.



Source: Adapted from (SWAFFEE, 2016).

Porém, há mais de 100 tipos diferentes de demência conhecidas que, pela prevalência da AD, acabam sendo popularmente chamadas de Alzheimer. Quando um idoso apresenta desorientação, perda de memória, lentidão nas respostas, as pessoas já julgam que está com AD. Dos outros tipos de demência, podemos destacar:

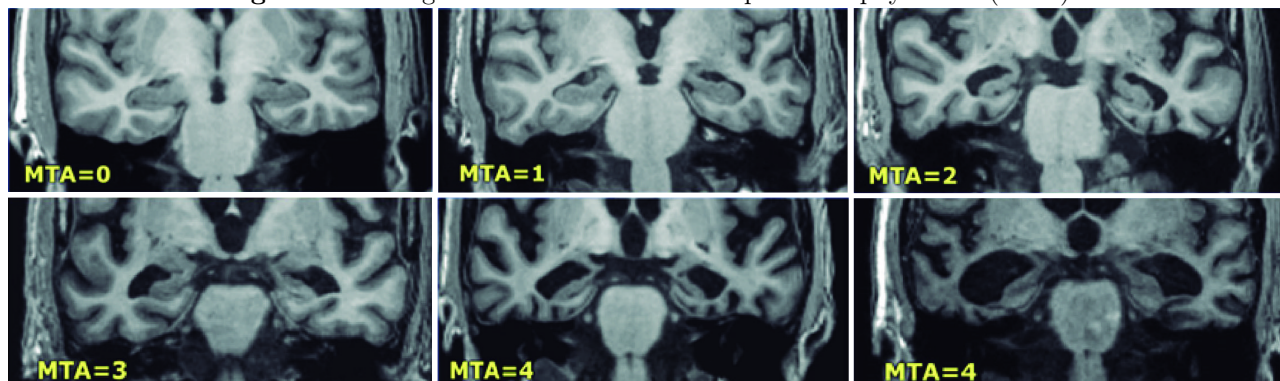
1. Demência vascular: pode aparecer depois de um acidente vascular cerebral que comprometa o funcionamento de uma área importante do cérebro. Também pode aparecer devido a pressão alta e diabetes não tratados ao longo dos anos, podendo causar pequenas isquemias que progressivamente afetam o cérebro.
2. Demência de corpos de Lewy: o paciente tem tremores, rigidez muscular, lentidão dos movimentos, também presentes no mal de Parkinson. O paciente flutua entre dias em melhor condição e dias piores. Também pode haver alucinações e perda de memória.
3. Demência frontotemporal: a doença afeta os lobos frontais e temporais do cérebro, que têm entre suas funções a regulação do humor e do comportamento. O paciente muda do estado de apatia para o estado de excitação com facilidade.
4. Demência de Parkinson: proteínas tóxicas começam a ser depositadas em áreas responsáveis pelo movimento. As demências de corpos de Lewy e da doença de Parkinson são difíceis de diferenciar.

Na AD há uma perda irrecuperável e real de células nervosas - neurônios e massa encefálica. É uma mudança mais rápida e precoce do que a que acontece no processo de envelhecimento natural, onde há uma perda discreta e é considerada normal. O próprio cérebro começa a se atacar e os cientistas ainda estudam para descobrir os motivos. O hipocampo é a principal parte do cérebro afetada. Ele é responsável por guardar novas informações, então no início da doença o paciente ainda tem memória de anos atrás, mas esquece do que fez ontem, por exemplo.

Estudos sobre as alterações das estruturas hipocampais atestam que as alterações na plasticidade neuronal, mielinização e conectividade interneuronal são uma doença importante (BALESTRIERI *et al.*, 2020). O escore de atrofia medial temporal (MTA) é utilizado pela neuroimagem para avaliar essas estruturas. Na Figura 1.2, pode-se visualizar cortes cerebrais de

ressonância magnética com MTA variando de 0 até 4. Em pacientes com 75 anos ou mais, um escore MTA acima de 3 é considerado anormal.

Figura 1.2. Images with several medial temporal atrophy scores (MTA).



Source: Adapted from ([BALESTRIERI, J., 2020](#)).

ESCALA MTA

O escore é calculado com base na análise visual da espessura da fissura coróideia, espessura do corno temporal e altura da formação hipocampal no corte coronal da sequência T1 da ressonância magnética do crânio.

- MTA 0: sem atrofia;
- MTA 1: apenas alargamento da fissura coróideia;
- MTA 2: alargamento da fissura coróideia e do corno temporal do ventrículo lateral;
- MTA 3: perda moderada do volume hipocampal (redução da sua altura);
- MTA 4: perda severa do volume hipocampal.

1.1.1 Estágios da doença

Os profissionais de saúde geralmente discutem a demência em estágios, que se referem ao quanto a demência de uma pessoa ou a AD progrediu. Definir o estágio ajuda os médicos a determinar os melhores tratamentos e auxilia a comunicação entre médicos e cuidadores.

Existem diversas escalas para a classificação de demência. Essas escalas ajudam a entender melhor os diferentes estágios da AD com base em quão bem uma pessoa pensa (declínio cognitivo) e funções (habilidades físicas). Como exemplo de escalas pode-se citar:

1. Global Deterioration Scale / Reisberg Scale (GDS) - A GDS se divide em sete estágios com base na quantidade de declínio cognitivo. Esse teste é mais relevante para pessoas com AD porque alguns outros tipos de demência nem sempre incluem perda de memória.
2. Functional Assessment Staging Test (FAST) - Assim como a escala GDS, a FAST é um sistema de sete estágios baseado mais no nível da habilidade funcional e na capacidade de realizar atividades da vida diária, do que no declínio cognitivo.
3. Clinical Dementia Rating (CDR) - A CDR usa um sistema de cinco pontos baseado em habilidades cognitivas (pensamento) e quão bem uma pessoa funciona. Alguém com suspeita de demência é avaliado em seis áreas: memória, orientação, julgamento e resolução de problemas, assuntos da comunidade, lar e hobbies e cuidados pessoais. A partir de entrevistas com o sujeito e outros, inclusive familiares, é atribuída uma das cinco etapas possíveis (HUGHES C P, 1982).

Na Tabela 1.1, estão descritas as características de cada estágio da escala CDR e o tempo médio esperado para a duração de cada estágio.

Tabela 1.1. Clinical Dementia Rating (CDR) Scale

Stage	Description	Expected duration of stage
CDR-0	No Dementia	N/A
CDR-0.5	Very Mild Dementia Memory problems are slight, but consistent. - Some difficulty with time and problem-solving - Daily life is slightly impaired - Individuals can perform personal care activities	Average duration is a few years up to 7 years.
CDR-1	Mild Dementia Memory loss is moderate, especially for recent events, and interferes with daily activities. - Moderate difficulty with solving problems - Cannot function independently in community affairs - Difficulty with daily activities and hobbies, especially complex ones	Average duration is 2 years.
CDR-2	Moderate Dementia More profound memory loss, only retaining highly learned material. - Disorientation with respect to time and place - Impaired judgment, with difficulty handling problems - Little to no independent function at home - Can only do simple chores - Fewer interests	Average duration is just under 2 years to 4 years.
CDR-3	Severe Dementia Severe memory loss. - Disorientation with respect to time or place - No judgment or problem-solving abilities - Cannot participate in community affairs outside home - Requires help with all tasks of daily living - Requires help with most personal care - Frequent incontinence	Average duration is 1 year to 2.5 years.

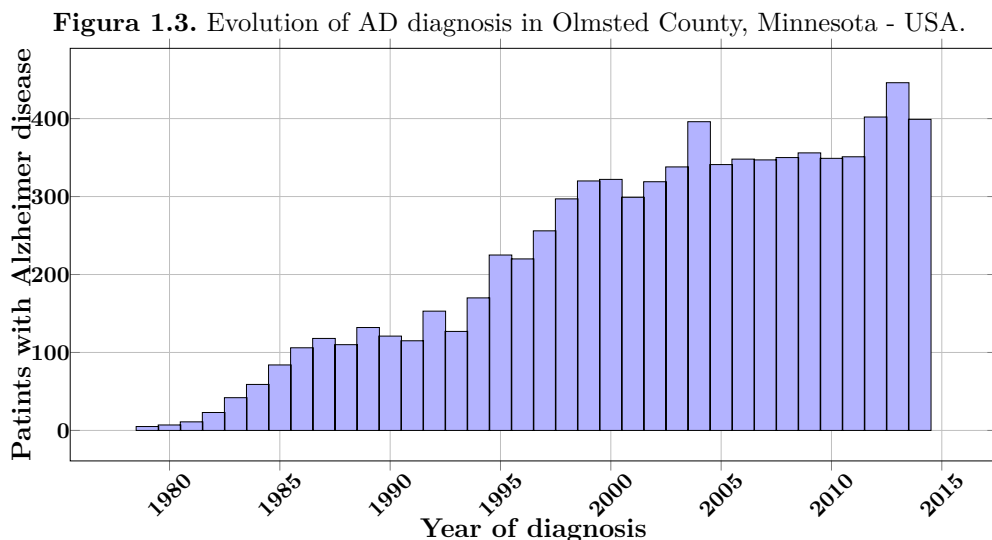
Source: Adapted from (CENTRAT, 2022).

Alguns estudos utilizam outros tipos de classificação do estágio da AD, como é o caso do *Australian Imaging, Biomarker & Lifestyle Flagship Study of Ageing* (AIBL) (ELLIS *et al.*, 2009), que é resultante de um estudo australiano com centenas de pacientes diagnosticados em três estágios de progressão da doença: voluntários com cognição normal (CN - *cognitively normal*), com prejuízo cognitivo leve (MCI - *mild cognitive impairment*) e com doença de Alzheimer (AD - *probable Alzheimers disease*) estabelecida. Outro caso é o *Alzheimers Disease Neuroimaging Initiative* (ADNI) (MARINESCU *et al.*, 2018), um estudo norte-americano, que também classifica o estágio da progressão da doença nas três categorias: CN, MCI e AD. Dados desses dois estudos serão utilizados neste trabalho. O estudo *Open Access Series of Imaging Studies* (OASIS) (MARCUS *et al.*, 2007) utiliza a escala CDR.

1.1.2 Breve histórico da doença

A AD somente passou a ser assim conhecida quando, em 1906, o médico alemão Alois Alzheimer diagnosticou a demência em uma paciente de 55 anos, após ter realizado biopsia em seu cérebro e relacionado as sequelas cerebrais com os registros médicos e sintomas da paciente. Alois Alzheimer anunciou esses resultados na 37^a Conferência de Psiquiatria, realizada em Tübingen, Alemanha (YANG H. D., 2016). A AD foi posteriormente batizada pelo professor Emil Kraepelin, em seu livro de psiquiatria de 1910. Apesar de ser conhecida por AD a partir do início do século XX, a doença marca o histórico passado distante da civilização, carregada de preconceitos e falta de compreensão, tornando-se um estigma social (YANG H. D., 2016).

Somente na segunda metade dos anos 1970 a AD recebeu um código internacional de identificação de doenças (CID - *International Classification of Disease*) e passou assim a ser diagnosticada. Utilizando os dados do *Rochester Epidemiology Project*, foi possível identificar a evolução no número de diagnósticos de AD, em pacientes com idade ≥ 65 , no condado de Olmsted - Minnesota, Estados Unidos, de 1980 a 2014 (KNOPMAN D. S., 2019). Na Figura 1.3, o eixo horizontal representa o ano e o eixo vertical, o número de pessoas únicas atribuídas nesse ano com um diagnóstico de AD. Os dados revelam um evidente aumento no número de casos, provavelmente pela melhor compreensão da doença por parte da classe médica, ou pelo aumento dos casos resultantes do envelhecimento da população, pela melhoria nas ferramentas de diagnóstico, por esses e outros fatores.

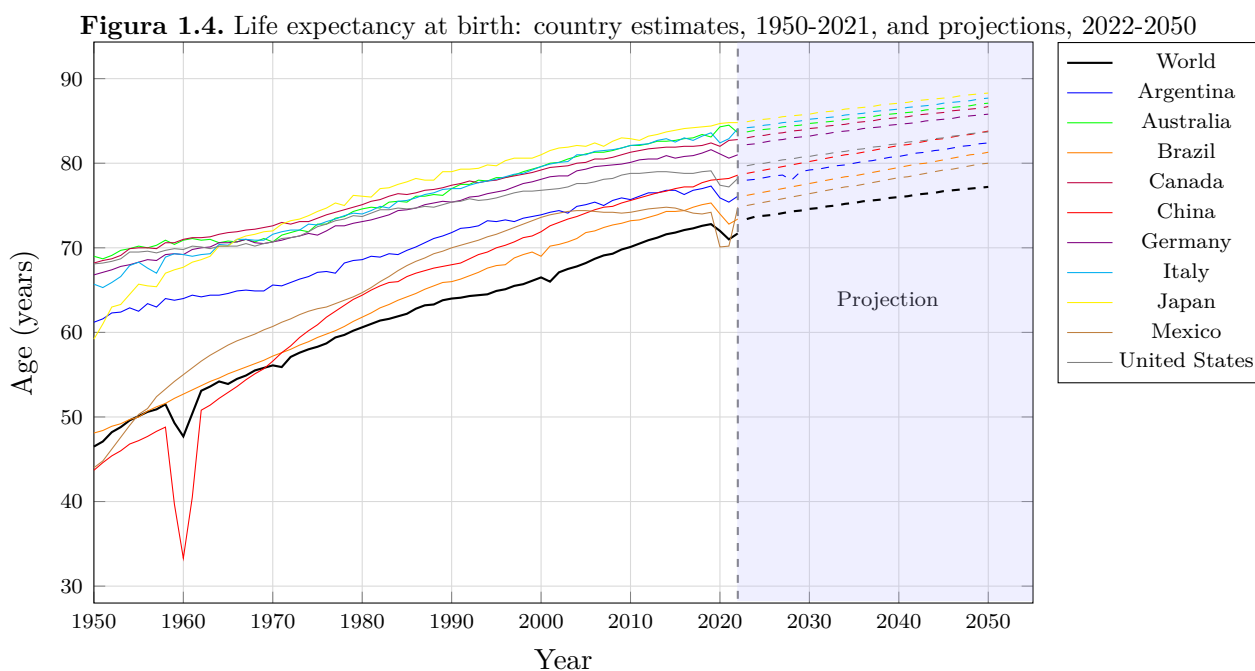


Source: Adapted from (KNOPMAN D. S., 2019).

Por estar associada principalmente com o processo de envelhecimento cognitivo, a AD deve se tornar ainda mais presente com o progressivo envelhecimento da população mundial. O aumento significativo da expectativa de vida em todo o planeta tem tornado a AD um importante problema de saúde pública com crescentes custos de assistência médica. Nos Estados Unidos, no ano de 2020, esse custo anual de tratamento foi estimado em U\$ 50 bilhões (SCHACHTER MD; KENNETH L. DAVIS, 2000).

A Organização das Nações Unidas (ONU), no seu relatório *World Population Prospects* de 2022, estima que a expectativa de vida ao nascer vai aumentar em todos os continentes nos próximos 30 anos. Até 2050, a expectativa de vida ao nascer é projetada para atingir 77,2 anos em média em todo o mundo, com uma diferença de 31,8 anos entre os países com menores e maiores índices. À medida que os países com menores índices completam a transição demográfica e atingem níveis historicamente baixos de mortalidade, novas reduções nas taxas de mortalidade tornam-se mais difíceis de alcançar e o aumento da esperança de vida se estabiliza, provocando uma tendência para a convergência entre países e regiões a longo prazo (ONU, 2022).

Na Figura 1.4, baseada em dados da ONU (ONU, 2022), tem-se a Austrália e o Japão com as maiores expectativas de vida. No outro extremo, não mostrado na figura, e abaixo da média mundial, a África subsariana possui os menores índices.



Source: Adapted from (ONU, 2022).

1.1.3 O diagnóstico da doença

O diagnóstico de AD é atualmente feito utilizando um variado ferramental que envolve diversas especialidades como neurologistas, neuropsicólogos, geriatras e psiquiatras geriátricos (REPORTI, 2021), e tem como base diferentes grupos de exames, dentre os quais pode-se destacar: histórico médico, avaliação de mudança comportamental, testes cognitivos, exames físicos e neurológicos, exames de sangue, exames de imagem e determinação de níveis de beta-amiloide.

Na avaliação em um paciente com comprometimento cognitivo e com uma possível síndrome demencial, a anamnese clínica e o exame neurológico são de extrema importância, e os exames complementares. Com o avanço dos biomarcadores, novas técnicas de diagnóstico evoluíram, proporcionando maior acurácia na avaliação, contudo a indicação destas ainda está restrita a casos específicos e no âmbito das pesquisas científicas (PARMERA; NITRINI, 2015).

Após uma anamnese satisfatória, o próximo passo constitui o exame neurológico geral e cognitivo aprofundados. A Academia Americana de Neurologia recomenda testagem de vitamina B12, hormônios tireoidianos, creatinina e hemograma. Alguma imagem estrutural como tomografia de crânio ou ressonância de crânio é igualmente recomendada (PARMERA; NITRINI, 2015).

Ainda assim, diagnosticar a AD não é tarefa fácil para os médicos especialistas, abrindo espaço para uso de outras estratégias que possam contribuir com o diagnóstico. Por ser uma doença multifatorial e com difícil identificação das causas, podendo também ser confundida com outros tipos de demência, o diagnóstico preciso ainda é um campo aberto para a ciência.

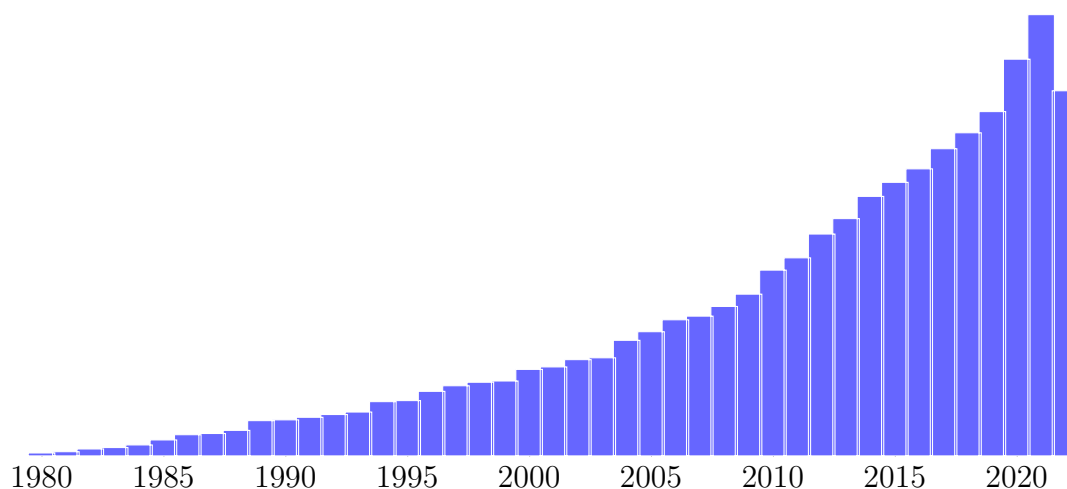
Dado que a AD evolui progressivamente do estágio da menor para o de maior comprometimento cognitivo, também é desejável diagnosticar pacientes que estão no estágio de prejuízo cognitivo leve, o que permitirá adoção inicial de tratamento médico com a intenção de atrasar a evolução da demência ou de reduzir os impactos financeiros do tratamento.

As pesquisas em torno do tema AD crescem a cada ano. Uma pesquisa feita no portal PubMed¹, usando a palavra "Alzheimer's", no período entre 1980 e setembro de 2022, revelou mais de 200 mil artigos de pesquisa cadastrados sobre o assunto. A Figura 1.5 mostra distribuição dos artigos científicos em cada ano e fica evidenciado o interesse crescente da comunidade

¹<https://pubmed.ncbi.nlm.nih.gov/>

científica pelo tema da AD.

Figura 1.5. PubMed portal search results with the word "Alzheimer's" from 1980 to September 2022.



Source: Own authorship.

1.1.4 Alzheimer no Brasil

Segundo o Instituto Brasileiro de Geografia e Estatística (IBGE), no seu site², o Brasil terá a população de 232.933.276 habitantes em 2050. Desses 232 milhões de habitantes, cerca de 34,88% serão de idosos com 65 anos ou mais, totalizando aproximadamente 81 milhões de habitantes.

A Sociedade Brasileira de Geriatria e Gerontologia (SBGG), no seu site³, estima que 6,8% da população acima de 60 anos tem chance de desenvolver algum tipo de demência, sendo de 40% a 60% do tipo Alzheimer.

Assim, pode-se estimar que cerca de 2,2 a 3,3 milhões de brasileiros poderão desenvolver a **AD** em torno do ano de 2050.

1.1.5 Alzheimer e a Covid-19

Um recente estudo([LAHIRA et al., 2021](#)) de pesquisadores do Instituto Butantã chegou à conclusão que a AD é um fator de risco para quem contrai Covid-19, independentemente

²<https://www.ibge.gov.br/apps/populacao/projecao/>

³<https://sbgg.org.br/>

da idade. Segundo a pesquisa, a AD não aumentou o risco de internações ao ser comparada com outras comorbidades. Mas quando um paciente era hospitalizado e já tinha a doença, o risco de desenvolver um quadro mais grave por conta do SARS-CoV-2 foi três vezes maior na comparação com quem não tinha a demência, e ainda seis vezes maior se o paciente tivesse mais de 80 anos.

1.2 JUSTIFICATIVA PARA INVESTIGAÇÃO DO TEMA

Segundo o relatório de 2019 da Alzheimer’s Disease International ([Alzheimers Disease International, 2019](#)), a estimativa do número de pessoas com demência no mundo será de 78 milhões em 2030 e de 139 milhões em 2050.

Segundo o relatório de 2021 da Alzheimer’s Disease International ([Gauthier S *et al*, 2021](#)), 75% das pessoas com demência não estão sendo diagnosticadas. Também nesse relatório, faz-se a estimativa de que 46% das pessoas que vivem com demência e cuidadores identificaram o medo do diagnóstico e o estigma como barreiras ao diagnóstico.

O relatório de 2019 da Alzheimer’s Disease International ([Alzheimers Disease International, 2019](#)) ainda relata que o total de custos com a demência em 2015 foi de US\$ 818 bilhões de dólares, em 2018 de US\$ 1 trilhão de dólares e estima para 2030 o valor de US\$ 2 trilhões de dólares.

Desse modo, justifica-se a presente tese por contribuir para tornar o mundo mais próximo do diagnóstico da AD, propondo construir ferramentas computacionais que possam auxiliar os profissionais de saúde a realizar o diagnóstico da AD, num esforço complementar às competências médicas.

1.3 OBJETIVOS

O objetivo geral desta tese consiste em investigar estratégias de aprendizado de máquina que possam contribuir para o auxílio ao diagnóstico da AD, com base em dados clínicos e de imagens de ressonância magnética. Os dados primários utilizados foram obtidos junto a estudos científicos realizados nos Estados Unidos e na Austrália, mediante autorização de uso dos dados.

Para alcançar o objetivo geral desta tese, os seguintes objetivos específicos foram propostos:

1. Estabelecer licenciamentos para obtenção das bases de dados confiáveis de natureza clínica e de ressonância magnética;
2. Desenvolver um *framework* para leitura e tratamento dos bancos de dados e implementar técnicas de aprendizado de máquina;
3. Realizar experimentos exploratórios sobre as bases de dados clínicos e de imagens de ressonância magnética;

CENÁRIO 1: Dados clínicos

4. Investigar e propor uma técnica de extração de atributos (*features*) que se adéque ao diagnóstico baseado em dados clínicos;
5. Definir e realizar experimentos com o método de seleção de atributos, sobre dados clínicos, que apresente melhor resultado;

CENÁRIO 2: Imagens médicas

6. Avaliar as imagens no plano sagital, coronal e axial;
7. Propor algoritmo automático de seleção de imagem (*slice*) na ressonância magnética;
8. Elaborar estratégias de montagem de bancos de imagem que melhor representem o problema do diagnóstico da AD;
9. Viabilizar e realizar experimentos de classificação com base em redes convolucionais;
10. Realizar a análise e gerar resultados dos métodos propostos;
11. Discutir os resultados obtidos.

1.4 ORGANIZAÇÃO DA TESE

A presente tese está dividida em seis capítulos, a saber:

- Capítulo **III**: INTRODUÇÃO. Este capítulo contém a introdução do trabalho, a revisão da doença de Alzheimer, a contextualização e a justificativa do problema de pesquisa. Nele também estão apresentados os objetivos do estudo.
- Capítulo **IV**: FUNDAMENTOS. São revisados os conceitos de inteligência artificial, aprendizado de máquina e aprendizado profundo. Também é apresentado o problema do diagnóstico da doença de Alzheimer, com revisão da literatura e estado da arte.
- Capítulo **V**: DADOS CLÍNICOS. São apresentadas algumas das bases tecnológicas para o tratamento de dados clínicos de auxílio ao diagnóstico de Alzheimer. As bases de dados são descritas e o pré-processamento para montagem dos *datasets* finais é apresentado. Também é feita uma contextualização do problema da seleção de atributos e apresentado o algoritmo proposto nesta tese. São apresentadas as métricas de avaliação do desempenho dos modelos treinados e a metodologia de validação cruzada aplicada. Por fim, são apresentados os modelos de aprendizado de máquina utilizados.
- Capítulo **VI**: IMAGEM. São apresentadas algumas das bases tecnológicas para o tratamento de imagens de ressonância magnética de auxílio ao diagnóstico de Alzheimer. O plano axial é avaliado e proposto um algoritmo automático de seleção de imagem. Também é avaliado o plano coronal. É feita uma proposta de aumento de dados com base no uso de imagens de cortes vizinhos. Os bancos de imagem são gerados e preparados para o processamento de algoritmos de aprendizado de máquina.
- Capítulo **VII**: RESULTADOS. A aplicação no novo algoritmo de seleção de atributos gerou bases de dados clínicas reduzidas e elas são apresentadas neste capítulo. Foram gerados bancos de dados de imagem com estratégia de aumento de dados, para viabilizar a avaliação por redes convolucionais. Também é feita a avaliação dos modelos e os resultados são explicitados. Os resultados são comparados com outros resultados do estado da arte.
- Capítulo **VIII**: CONCLUSÃO. É feita uma avaliação geral das contribuições deste trabalho e também a sugestão de trabalhos futuros.

FUNDAMENTOS TEÓRICOS

Neste capítulo, são apresentados os fundamentos teóricos, com a revisão das pesquisas e das discussões de outros autores sobre o auxílio ao diagnóstico da AD por técnicas computacionais, com ênfase nas técnicas de inteligência artificial e apresentação do estado da arte.

2.1 INTELIGÊNCIA ARTIFICIAL NO DIAGNÓSTICO DE DOENÇAS NEURODEGENERATIVAS

A inteligência artificial (AI) é um próspero campo com muitas aplicações práticas e pontos de investigação. A busca é constante por sistemas em software inteligente que permitam automatizar o trabalho de rotina, compreender a fala ou as imagens, fazer diagnósticos em medicina e apoiar a investigação científica básica. Logo no início do uso das técnicas de AI, a área abordou e resolveu rapidamente problemas que são intelectualmente difíceis para os seres humanos, mas relativamente simples para computadores - problemas que podem ser descritos por uma lista de regras formais e matemáticas. O verdadeiro desafio para a AI provou ser a resolução das tarefas que são fáceis de executar, mas difíceis de descrever formalmente - problemas que resolvemos intuitivamente, quase que automaticamente, como o reconhecimento de palavras faladas ou rostos em imagens ([GOODFELLOW *et al.*, 2016](#)).

Os sistemas de AI necessitam da capacidade de adquirir os seus próprios conhecimentos, através da extração de padrões a partir de dados brutos. Essa capacidade é conhecida como aprendizagem por máquina (ML). O advento do ML permitiu aos computadores enfrentar problemas que envolvem o conhecimento do mundo real e tomar decisões que parecem subjetivas.

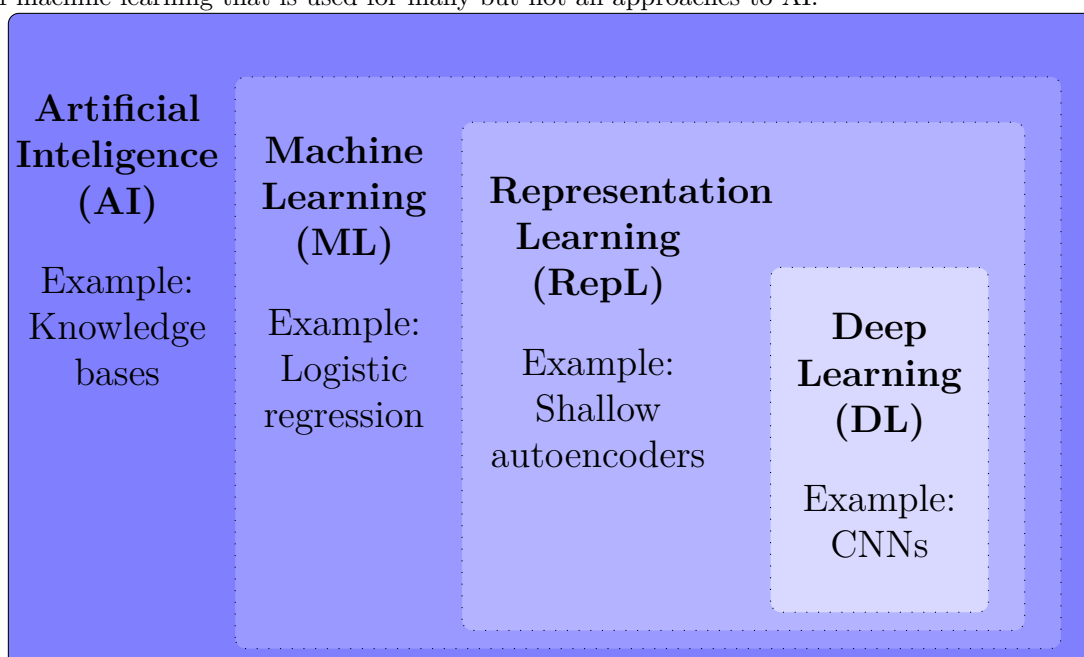
Para problemas mais complexos, contudo, é difícil saber que características devem ser extraídas. Uma solução para esse problema é utilizar ML para descobrir não só o mapeamento desde a representação até a saída, mas também a própria representação. Essa abordagem é

conhecida como aprendizagem da representação (do inglês *Representation Learning* - RepL). A RepL resulta muitas vezes num desempenho muito melhor do que aquele que pode ser obtido com representações extraídas manualmente. A RepL também permite que os sistemas de AI se adaptem rapidamente a novas tarefas, com o mínimo de intervenção humana. Um algoritmo de RepL pode descobrir um conjunto de características para uma tarefa simples em minutos, ou para uma tarefa complexa em horas. A representação manual de características para uma tarefa complexa requer muito tempo humano e esforço, podendo demorar décadas para toda uma comunidade de investigadores (GOODEFELLOW *et al.*, 2016).

Uma fonte importante de dificuldade em muitas aplicações de AI do mundo real é que muitos dos fatores influenciam cada pedaço de dados que somos capazes de observar. Muitos desses fatores de variação podem ser identificados apenas utilizando uma compreensão sofisticada, quase humana, dos dados. Nesses casos, a RepL não parece ser muito útil. A aprendizagem profunda (DL) resolve esse problema, introduzindo representações que se expressam em termos de outras representações mais simples.

De modo geral, essas técnicas de AI se relacionam conforme mostrado na Figura 2.1, que também apresenta um exemplo de algoritmo para cada caso.

Figura 2.1. A Venn diagram showing how deep learning is a kind of representation learning, which is in turn a kind of machine learning that is used for many but not all approaches to AI.



Source: Own authorship.

Desde algum tempo, busca-se desenvolver sistemas computadorizados de apoio ao diagnóstico clínico. Há mais de 50 anos, já se oferecia um sistema que, avaliando dados de um paciente, sugeria ações para restabelecer seu equilíbrio hidroeletrolítico (BLEICH, 1969).

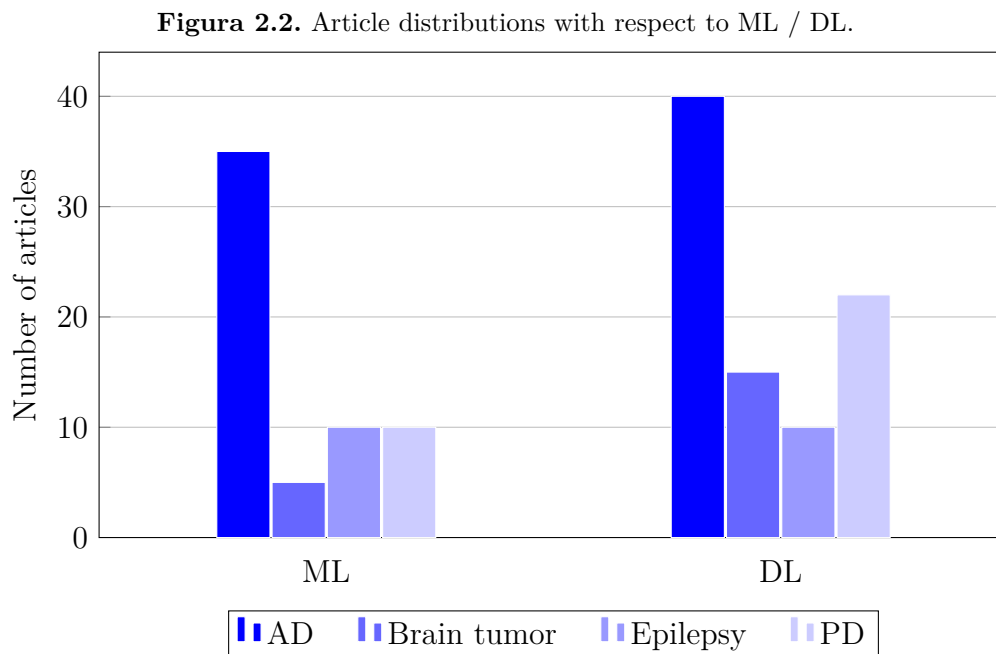
Com a AI, os distúrbios neurodegenerativos podem ser investigados em um nível mais profundo, fornecendo uma visão geral abrangente da doença e abrindo caminhos para a aplicação da medicina de precisão para essas patologias. Portanto, constitui um recurso poderoso para auxiliar no diagnóstico de doenças neurodegenerativas, uma vez que essas ferramentas revolucionam o meio médico, dando mais segurança e praticidade para o manejo dessas patologias (BRITO *et al.*, 2021).

2.2 AUXÍLIO COMPUTACIONAL NO DIAGNÓSTICO DA DOENÇA DE ALZHEIMER

As técnicas de ML e DL têm sido importantes aliadas no auxílio ao diagnóstico da AD (NIYAS; P, 2021a; NIYAS; P, 2021b; NGUYEN M, 2020; ALBRIGHT, 2019; IDDI *et al.*, 2019; MOORE *et al.*, 2019; GHAZI *et al.*, 2019; NGUYEN *et al.*, 2018; ZHANG *et al.*, 2011).

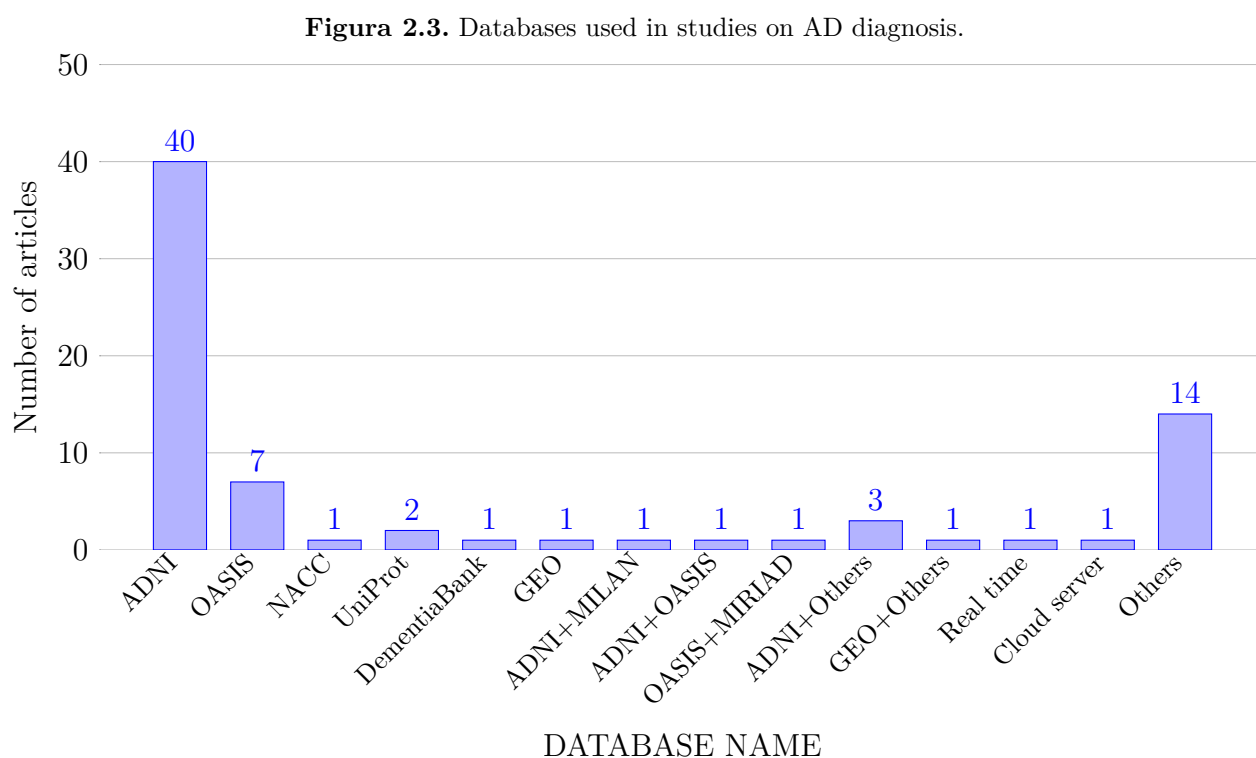
De modo geral, o uso de técnicas de inteligência artificial para melhoria da precisão do diagnóstico de doenças do cérebro tem crescido significativamente nos últimos anos (KHAN *et al.*, 2021). É possível identificar trabalhos científicos recentemente publicados que utilizam essas duas estratégias para realizar o diagnóstico de AD e outras doenças do cérebro, como: tumor cerebral, epilepsia e mal de Parkinson. Em (KHAN *et al.*, 2021), esse levantamento foi realizado utilizando as bases do IEEE Xplore, Sciencedirect e Google Scholar, entre os anos de 2018 e 2020, tendo sido encontrados 75 artigos relacionados a AD, sendo 40 fazendo uso de DL e 35 utilizando outras técnicas de ML.

Na Figura 2.2, é mostrado o número de artigos encontrados que utilizam ML ou DL para AD, tumor cerebral, epilepsia e mal de Parkinson.



Source: Adapted from (KHAN *et al.*, 2021).

Também em (KHAN *et al.*, 2021) foi avaliado qual a base de dados utilizada por cada pesquisa, na determinação do diagnóstico da AD. A Figura 2.3 mostra que 60% dos trabalhos foram feitos utilizando dados do estudo ADNI, que também é utilizado neste trabalho, juntamente dos estudos AIBL e OASIS.



Source: Adapted from (KHAN *et al.*, 2021).

Especificamente utilizando as bases de dados da AIBL, ADNI, ou ambas, os estudos (NIYAS; P, 2021a), (NIYAS; P, 2021b), (IDDI *et al.*, 2019), (MOORE *et al.*, 2019), (ZHANG *et al.*, 2011) utilizaram técnicas de ML e os estudos (NGUYEN M, 2020), (ALBRIGHT, 2019), (GHAZI *et al.*, 2019), (NGUYEN *et al.*, 2018), (LIU J, 2020), (HAN *et al.*, 2020), (LI *et al.*, 2022), (RAZAK *et al.*, 2022), (HAZARIKA *et al.*, 2022), (ZHANG *et al.*, 2023), (WU *et al.*, 2022), (GAO *et al.*, 2023a), (GAO *et al.*, 2023b), (SHARMA *et al.*, 2022), utilizaram técnicas de DL. Para efeito de avaliação com o estado da arte, os resultados obtidos nos trabalhos referenciados no último período foram comparados com os resultados obtidos neste trabalho e serão explicitados no Capítulo 5 - Resultados e discussão.

2.3 REVISÃO DA LITERATURA

Como ficou evidenciado na Figura 3.5, houve um acréscimo significativo de artigos científicos do ano de 1980 até setembro de 2022, mostrando o interesse crescente da comunidade científica em relação ao tema da AD. Também em (KHAN *et al.*, 2021) foram encontrados 75 artigos relacionados ao diagnóstico de AD utilizando técnicas de inteligência computacional, sendo deles 40 fazendo uso de DL e outros 35 utilizando outras técnicas de ML, entre os anos de 2018 e 2020. Alguns desses artigos estão revisados a seguir e também artigos publicados até o ano de 2022.

Em (NIYAS; P, 2021a), os autores procuraram comparar o desempenho do estado da arte de algoritmos de classificação utilizando *Dynamic Ensemble Selection* (DES) para o diagnóstico de participantes em três categorias: CN, MCI e AD. Um conjunto dos mais comuns classificadores de ML é utilizado como entrada do DES. Os resultados de classificação são comparados com e sem o uso de DES. Os autores concluem que o uso do DES melhora as métricas de avaliação da classificação. Nesse trabalho, foi utilizada a base de dados da ADNI.

Já em (NIYAS; P, 2021b), os autores realizaram uma seleção de atributos baseada no *Fischer score* sobre as bases de dados da ADNI e AIBL e analisaram as métricas da avaliação da classificação para as classes CN, MCI e AD utilizando os classificadores *support vector machine* (SVM) e *k-nearest neighbors* (kNN). Os autores concluíram que o método de seleção de atributos utilizado foi eficiente e melhorou os resultados da classificação. Também foram

avaliadas as métricas de classificação para as classes NC v/s AD, MCI v/s AD e MCI v/s NC. Foi proposto fazer a avaliação dos resultados utilizando a acurácia balanceada (*balanced classification accuracy* - BCA) em virtude do desbalanceamento das classes nos conjuntos de dados.

Em (IDDI *et al.*, 2019) foi proposta uma abordagem em duas fases para modelagem e previsão de medidas de cognição, funcionalidade, imagem do cérebro, fluidos biomarcadores e diagnóstico de indivíduos usando múltiplos domínios simultaneamente. Na primeira fase, foram utilizados modelos de efeitos mistos multivariados para modelar simultaneamente vários marcadores ao longo do tempo. Na segunda fase, foram utilizadas *random forests* (RF) para prever diagnósticos de CN, MCI e AD de marcadores contínuos baseados no modelo da primeira fase. A base de dados utilizada foi a ADNI.

Uma previsão de AD baseada em séries temporais foi proposta em (MOORE *et al.*, 2019). Foi utilizada uma RF para aprender a relação entre pares de pontos de dados em diferentes separações temporais. O vector de entrada foi composto por um resumo do histórico das séries cronológicas e inclui tanto variáveis demográficas como não demográficas, tais como dados genéticos. A base de dados utilizada foi a ADNI. Segundo os autores, os resultados mostram que o método foi eficaz e comparável com outros métodos.

Uma técnica de previsão da progressão da doença de Alzheimer usando *recurrent neural networks* (RNN) foi proposta em (NGUYEN M, 2020). Nesse estudo, com marcadores de AD multimodais e diagnóstico clínico de um indivíduo a partir de um ou mais pontos no tempo, procurou-se prever o diagnóstico clínico, a cognição e o diagnóstico do volume ventricular do indivíduo para cada mês no futuro. Foi utilizada a base de dados do desafio *The Alzheimer's Disease Prediction Of Longitudinal Evolution* (TADPOLE) da ADNI. Os autores declaram que essa abordagem ficou em segundo lugar no desafio TADPOLE a partir de 3 de junho de 2020.

Também a previsão da progressão da doença de Alzheimer utilizando redes neurais e um novo algoritmo de pré-processamento, chamado "All-Pairs", que envolveu a comparação de todos os pares possíveis de pontos de dados temporais para cada paciente, foi proposto em (ALBRIGHT, 2019). Segundo os autores, a rede neural treinada pode ser utilizada para identificar pacientes em fases iniciais de AD que são, portanto, bons candidatos a ensaios clínicos para a terapêutica de AD. Foi utilizada a base de dados da ADNI.

Em (GHAZI *et al.*, 2019), um algoritmo *long short-term memory* (LSTM) foi proposto e aplicado para modelar a progressão da doença de Alzheimer (AD) usando seis biomarcadores de imagem, ou seja, volumes de ventrículos, hipocampo, cérebro inteiro, fusiforme, giro temporal médio e córtex entorhinal, da base de dados de ressonância magnética. Foi utilizada a base de dados da ADNI. O trabalho concluiu que o tratamento integrado de valores em falta na formação de uma rede LSTM beneficia a aplicação de RNNs em modelagem neurodegenerativa da progressão da doença em cortes longitudinais.

Também em (NGUYEN *et al.*, 2018) foi feita a proposta de modelagem da progressão da doença de Alzheimer usando RNN. Fazendo uso da base de dados da ADNI, foram propostas estratégias para treinar a RNN utilizando dados que envolvem a utilização do RNN para preencher os dados incompletos ou pontos em falta na linha do tempo.

Em (LIU *et al.*, 2020), também usando a base de dados ADNI, foi considerado o problema de prever a progressão da doença medida pelas pontuações cognitivas e selecionar biomarcadores preditivos da progressão. A maioria dos trabalhos anteriores sobre a previsão da progressão da AD ignora a questão dos dados em falta. Os dados em falta representam uma grande dificuldade para a modelação de dados longitudinais, uma vez que a maioria dos modelos estatísticos assume dados completos.

Novos modelos de diagnóstico utilizando imagens de ressonância magnética para identificar os vários estágios da AD foram apresentados em (HAN *et al.*, 2020). Para verificar a validade dos modelos propostos, experimentos em imagens de ressonância magnética coletadas no site da ADNI foram testados e avaliados.

Em (LI *et al.*, 2022) foi utilizada uma rede neural convolucional 3D, que fez uso de toda a imagem de ressonância magnética 3D como a entrada. O estudo utilizou dados da ADNI e imagens de ressonância magnética.

Uma proposta de estrutura integrativa baseada em aprendizagem profunda para alcançar um melhor desempenho preditivo para o diagnóstico da doença de Alzheimer foi apresentado em (RAZZAK *et al.*, 2022). O estudo fez uso de dados da ADNI e imagens de ressonância magnética.

Já (HAZARIKA *et al.*, 2022) apresentaram um novo modelo baseado em redes convolucio-

nais para distinguir pacientes com AD e MCI de indivíduos cognitivamente normais. O novo modelo foi baseado na VGG-19 e na DenseNet. Os autores também utilizaram dados da ADNI.

Uma acurácia de 93,4% para o diagnóstico AD \times CN usando dados AIBL foi atingido por (ZHANG *et al.*, 2023), que apresentou um modelo DL misturador perceptron multicamada multitarefa tridimensional.

Um novo modelo CNN 3D Multi-scale baseado em atenção foi proposto em (WU *et al.*, 2022), tendo alcançado a acurácia de 91,3% sobre dados ADNI.

(GAO *et al.*, 2023a) apresentou uma rede híbrida de convolução de atenção multi-escala e uma rede transformadora de envelhecimento para diagnóstico de DA, com dados de sMRI. O estudo reporta acurácia de 90,5% no diagnóstico de AD \times CN sobre dados ADNI.

Neste artigo (GAO *et al.*, 2023b), os autores propuseram uma rede adversária generativa de transferência de estado cerebral (BrainStatTrans-GAN) para um diagnóstico da AD. O trabalho reporta acurácia de 94,9% para diagnóstico AD \times CN em dados ADNI.

Uma nova abordagem de fusão de imagens estruturais e metabólicas baseada na transformada de pacotes wavelet utilizando exames de MRI e PET é proposta em (SHARMA *et al.*, 2022). O trabalho reporta acurácia de 97,33% para diagnóstico AD \times CN para dados da ADNI.

No Capítulo 5, o desempenho de alguns dos trabalhos aqui revisados foi comparado com os resultados obtidos neste trabalho.

Este capítulo apresentou fundamentos teóricos e fez a revisão de pesquisas de outros autores que também empregaram técnicas computacionais para auxiliar no diagnóstico da AD. No próximo capítulo, serão apresentadas as propostas tecnológicas para o desenvolvimento de técnicas computacionais para o auxílio do diagnóstico de AD baseado em dados clínicos.

ANÁLISE BASEADA EM DADOS CLÍNICOS

Este capítulo faz o detalhamento dos dados clínicos utilizados e das propostas tecnológicas desenvolvidas.

Os dados clínicos de um paciente envolvem dados de anamnese, exames laboratoriais, peso, idade, imagens, frequência cardíaca, valores de pressão, medicações, histórico clínico, diagnósticos e outros, que vão se tornar mais ou menos importantes diante do diagnóstico que se tenta determinar.

Nesta tese, os dados clínicos para AD foram extraídos dos estudos detalhados a seguir.

3.1 DETALHAMENTO DAS BASES DE DADOS UTILIZADAS

Como apresentado na Figura 2.3, há diversas bases de dados relacionados com a AD. Neste trabalho, decidimos por utilizar as bases ADNI, OASIS e AIBL. As bases ADNI e OASIS estão entre as mais utilizadas em artigos científicos no período de 2018 a 2020, segundo (KHAN *et al.*, 2021). A base AIBL também tem sido recentemente referenciada (NIYAS; P, 2021b).

3.1.1 Estudo OASIS

O *Open Access Series of Imaging Studies* (OASIS) é um projeto que visa tornar os conjuntos de dados de neuroimagem do cérebro livremente disponíveis à comunidade científica. A base de dados clínica pode ser acessada no site da OASIS¹, fazendo referência à base de dados OASIS-1 (MARCUS *et al.*, 2007). Essa base é constituída por uma coleção transversal de 416 pacientes com idades compreendidas entre os 18 e 96 anos. Para cada paciente, estão incluídas três ou quatro ressonâncias individuais ponderadas em T1, obtidas em sessões de varrimento

¹<https://www.oasis-brains.org>

único. Os sujeitos são todos destros e incluem tanto homens como mulheres. Os pacientes estão diagnosticados segundo a escala CDR de avaliação da AD (HUGHES *et al.*, 1982) e está disponível em formato *comma-separated values* (CSV).

Na Tabela 3.1, estão relacionados os atributos e a descrição do seu significado para a base de dados OASIS-1, utilizada neste trabalho.

Tabela 3.1. OASIS: Features descriptions

Feature	Description
Age	Age at time of image acquisition (years)
Sex	Sex (male or female)
Education	Years of education
Socioeconomic status	Assessed by the Hollingshead Index of Social Position and classified into categories from 1 (highest status) to 5 (lowest status)
MMSE score	Ranges from 0 (worst) to 30 (best)
CDR scale	0 = no dementia, 0.5 = very mild AD, 1 = mild AD, 2 = moderate AD
Atlas scaling factor	Computed scaling factor (unitless) that transforms native-space brain and skull to the atlas target (i.e., the determinant of the transform matrix)
eTIV	Estimated total intracranial volume (cm^3)
nWBV	Expressed as the percent of all voxels in the atlas-masked image that are labeled as gray or white matter by the automated tissue segmentation process

Source: Own authorship.

3.1.2 Estudo AIBL

O *Australian Imaging, Biomarker & Lifestyle Flagship Study of Ageing* (AIBL) (ELLIS *et al.*, 2009) é um estudo para descobrir quais biomarcadores, características cognitivas e fatores de saúde e estilo de vida determinam o desenvolvimento subsequente da AD sintomática. É um estudo australiano, iniciado em 2006, com mais de 1000 pacientes diagnosticados em três estágios de progressão da doença: voluntários com cognição normal (CN), com prejuízo cognitivo leve (MCI) e com doença de Alzheimer (AD) estabelecida. Informações complementares podem ser consultadas no site² do estudo.

Na Tabela 3.2, estão relacionados os atributos e a descrição do seu significado para a base de dados AIBL, utilizada neste trabalho.

²<https://aibl.csiro.au/>

Tabela 3.2. AIBL: Features descriptions

Feature	Description
HMT3	Red Blood Cell Count
HMT7	White Blood Cell Count
HMT13	Platelets
HMT40	Hemoglobin
HMT100	Mean Cell Hemoglobin
HMT102	Mean Cell Hemoglobin Concentrate
AXT117	Thyroid Stimulate Hormone
BAT126	Vitamin 12
RCT6	Urea Nitrogen
RCT11	Serum Glucose
RCT20	Cholestrol
RCT392	Creatinine
LIMMTOTAL	Total number of story units recalled-Logical Memory Immediate Recall
LDELTOTAL	Total number of story units recalled-Partial Score of Logical Memory test
MMSCORE	Mini Mental State Examination Score
CDGLOBAL	Clinical Dementia Rating Global
APGEN1	Apoe gene 1
APGEN2	Apoe gene 2
DXCURREN	Diagnosis status

Source: Own authorship.

3.1.3 Estudo ADNI

O *Alzheimers Disease Neuroimaging Initiative* (ADNI) (MARINESCU *et al.*, 2018) é um estudo norte-americano multicêntrico e longitudinal projetado para desenvolver biomarcadores clínicos, de imagem, genéticos e bioquímicos para a detecção precoce e rastreamento da AD. Desde o seu lançamento, há mais de uma década, a parceria público-privada histórica trouxe grandes contribuições para a pesquisa de AD, permitindo o compartilhamento de dados entre pesquisadores de todo o mundo. Informações complementares podem ser encontradas no site³ do estudo. Esse estudo também classifica o estágio da progressão da doença em três categorias: CN, MCI e AD.

Na Tabela 3.3, estão relacionados os atributos e a descrição do seu significado para a base de dados ADNI, utilizada neste trabalho.

Os dados dos estudos ADNI e AIBL podem ser baixados na plataforma *Image and Data*

³<https://adni.loni.usc.edu/>

Archive (IDA)⁴, que é gerenciada pelo *Laboratory of Neuro Imaging (LONI)* do *Mark and Mary Stevens Neuroimaging and Informatics Institute*, associado à *University of Southern California*.

Tabela 3.3. ADNI: Features descriptions

Feature	Description
Ventricles	Ventricles Volume
Hippocampus	Hippocampus Volume
CDRSB	Clinical Dementia Rating Scale Box
MMSE	Mini-Mental State Examination
WholeBrain	WholeBrain Volume
Entorhinal	Entorhinal Volume
Fusiform	Fusiform Volume
MidTemp	Med Temp Volume
ICV	Intracranial Volume
ADAS11	Alzheimers Disease Assesement Cognition Scale 11
ADAS13	Alzheimers Disease Assesement Cognition Scale 13
Age	Age at baseline
PTEDUCAT	Education
RAVLT_learning	Rey Auditory Verbal Learning Test
RAVLT_immediate	Rey Auditory Verbal Learning Test (5 sum)
RAVLT_Forgetting	Rey Auditory Verbal Learning Test Forgetting
RAVLT_Perc_Forgetting_bl	Rey Auditory Verbal Learning Test Percentile Forgetting
AV45	Average AV45 SUVR of frontal, AC, precuneus, and parietal cortex relative to the cerebellum
APOE4	Apolipoprotein E4
FAQ	Functional Activities Questionnaire
FDG	Average FDG-PET of angular, temporal, and posterior cingulate
DX	Diagnosis Status

Source: Own authorship.

Um maior detalhamento sobre os atributos em cada uma das bases de dados montadas pode ser verificado nos artigos de referência ([ELLIS et al., 2009](#)), ([MARINESCU et al., 2018](#)) e ([MARCUS et al., 2007](#)).

3.1.4 Pré-processamento

O pré-processamento é uma etapa essencial na preparação de dados de ML. Envolve a transformação de dados brutos em um formato mais adequado e eficaz para a formação de modelos de ML. O pré-processamento ajuda a melhorar a qualidade e a eficiência do processo

⁴<https://ida.loni.usc.edu/login.jsp>

de aprendizagem e pode ter um impacto significativo no desempenho e na precisão do modelo resultante.

Dos dados do estudo AIBL, originalmente distribuídos em 18 arquivos, foram escolhidos 18 atributos, 16 utilizados em (NIYAS; P, 2021b) e 2 genéticos, atribuídos em três classes: AD, MCI e CN, para cada avaliação do paciente, contendo as seguintes informações: avaliações de neuropsicologia, análises de sangue e dados genéticos. Chamamos essa base de dados de AIBL-18, cuja composição está detalhada na Tabela 3.4.

Tabela 3.4. AIBL-18: dataset composition

Quantity	Features
18	CDGLOBAL AXT117 BAT126 HMT3 HMT7 HMT13 HMT40 HMT100 HMT102 RCT6 RCT11 RCT20 RCT392 MMSCORE LIMMTOTAL LDELTOTAL APGEN1 APGEN2

Source: Own authorship.

Dos dados do estudo ADNI-TADPOLE, foram escolhidos 21 atributos, utilizados em (NIYAS; P, 2021b), também atribuídos em três classes: AD, MCI e CN, para cada avaliação do paciente. Chamamos essa base de dados de ADNI-21, cuja composição está detalhada na Tabela 3.5.

Tabela 3.5. ADNI-21: dataset composition

Quantity	Features
21	Ventricles Hippocampus WholeBrain Entorhinal Fusiform MidTemp ICV CDRSB MMSE RAVLT_learning RAVLT_immediate RAVLT_forgetting RAVLT_perc_forgetting_bl ADAS11 ADAS13 FAQ FDG AV45 APOE4 AGE PTEDUCAT

Source: Own authorship.

Dos dados do estudo OASIS, foram escolhidos oito atributos, utilizados em (PAIVA; ES-COVEDO, 2021), distribuídos em quatro classes: CDR0, CDR0.5, CDR1 e CDR2, para cada avaliação do paciente. Chamamos essa base de dados de OASIS-8, cuja composição está detalhada na Tabela 3.6.

Tabela 3.6. OASIS-8: dataset composition

Quantity	Features
8	Age Sex Education Socioeconomic_status MMSE_score Atlas_scaling_factor eTIV nWBV

Source: Own authorship.

O conjunto de dados AIBL-18 foi inicialmente montado com 1639 instâncias. Porém, houve a necessidade de pré-processamento, tendo em vista que:

- existe a ausência de dados em algumas instâncias, o que pode ter sido ocasionado por esquecimento de anotar o resultado de algum exame ou mesmo não ter havido a realização daquele exame;
- em alguns casos, há até seis instâncias por paciente, resultante das avaliações feitas pelo mesmo paciente em seis diferentes datas. Naturalmente, nesses casos o diagnóstico pode até ter mudado de CN para MCI, MCI para AD ou até AD para MCI. Pacientes com muitas avaliações tendem a produzir instâncias com maior dependência estatística.

Para tratar esses casos, foram realizadas, na sequência, as seguintes estratégias:

1. Optou-se por não imputar os dados, tendo em vista que a quantidade de instâncias é satisfatória para realizar a tarefa de classificação, mas excluir por inteiro todas as instâncias que possuem pelo menos um dado ausente.
2. Selecionou-se apenas uma instância por paciente, escolhendo-se o registro mais recente daqueles disponíveis. Essa estratégia busca minimizar a dependência no conjunto de dados.

Dessa forma, chegou-se a um conjunto de dados de 785 instâncias, 18 atributos e três classes, para o AIBL-18, o que representa 95,03% do total de pacientes, resultando num significativo aproveitamento dos dados originais, sem que tenha sido feita imputação de dados nem a utilização de dados de mais de um atendimento de qualquer paciente.

A mesma estratégia de pré-processamento foi aplicada para os dados ADNI-21, resultando em uma base de dados com 21 atributos, três classes e 1033 instâncias, o que representa 59,47% do total de pacientes do estudo.

O estudo OASIS possui originalmente quatro classes, sendo CDR 0, CDR 0.5, CDR 1 e CDR 2. Porém, tendo em vista que o número de instâncias para a classe CDR 2 é muito pequeno, optou-se por juntar as instâncias das classes CDR 1 e CDR 2 em uma única classe chamada CDR 1-2. O resultado do pré-processamento então resultou em uma base de dados com 216 instâncias, três classes e sete atributos.

3.2 ANÁLISE DOS DADOS

A base de dados AIBL-18 pós-processada resultou em 785 instâncias, 18 atributos e três classes, distribuídas conforme a Tabela 3.7.

Tabela 3.7. AIBL-18: instance distribution by class

Class	Number of instances
CN	547
MCI	125
AD	113

Source: Own authorship.

Da mesma forma, a base de dados ADNI-21 pós-processada resultou em 1033 instâncias, 21 atributos e três classes, distribuídas conforme a Tabela 3.8.

Tabela 3.8. ADNI-21: instance distribution by class

Class	Number of instances
CN	505
MCI	367
AD	161

Source: Own authorship.

A base de dados OASIS-7 pós-processada resultou em 217 instâncias, sete atributos e três classes, distribuídas conforme a Tabela 3.9.

Tabela 3.9. OASIS-7: instance distribution by class

Class	Number of instances
CDR 0	133
CDR 0.5	57
CDR 1-2	26

Source: Own authorship.

É natural esperar que todas as bases de dados demonstrem um desequilíbrio das instâncias entre as classes, tendo em vista que a presença da doença é menos frequente que o diagnóstico na população como um todo.

Porém, os modelos de classificação são afetados quando treinados com conjunto de dados que possuem classes desbalanceadas. O classificador treinado pode garantir acerto substantivo se ele simplesmente decidir sempre pela classe majoritária, para qualquer instância apresentada.

Para prevenir o mal condicionamento devido ao desbalanceamento das classes, pode-se ponderar a função de erro de classificação, de modo a dar mais peso aos erros que ocorrem na classe minoritária (CUI *et al.*, 2019).

Outra alternativa é a de sobreamostrar as classes minoritárias, sintetizando novas instâncias a partir das instâncias existentes. Essa técnica é conhecida na literatura como SMOTE (*Synthetic Minority Oversampling Technique*) (CHAWLA *et al.*, 2002). Variações da técnica SMOTE têm sido propostas e uma delas é a ADASYN (*Adaptive Synthetic Sampling Method*), que gera instâncias sintéticas na região do espaço de características onde a densidade de exemplos minoritários é baixa (HE *et al.*, 2008). Essas técnicas foram avaliadas nesta tese, mas não produziram resultados relevantes que justificassem ser incorporadas ao trabalho.

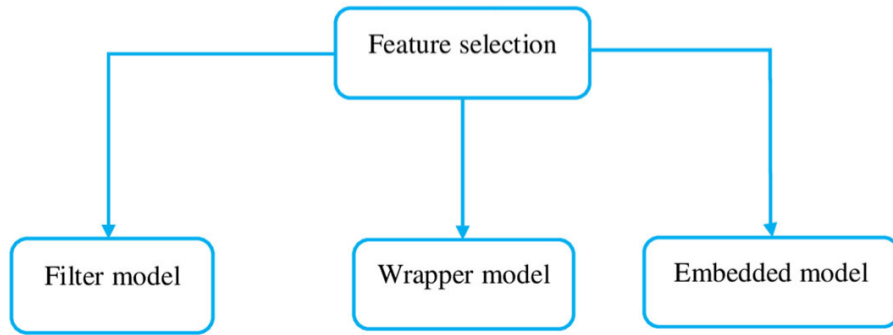
Fica o alerta de que a avaliação do modelo de classificação deve considerar o fato de todas as bases de dados serem desbalanceadas e as métricas de avaliação devem considerar isso.

3.3 SELEÇÃO DE ATRIBUTOS

A seleção de atributos é a abordagem pela qual você escolhe um subconjunto de atributos específicos para o projeto de modelos. Esse processo traz muitas vantagens, sendo a mais notável o aprimoramento de desempenho de um algoritmo de aprendizado de máquina.

Uma segunda vantagem inclui a diminuição da complexidade computacional. Se o número de atributos em um modelo for reduzido, mais simples se torna o cálculo dos parâmetros do modelo. Outra vantagem inclui uma melhor compreensão dos resultados. Com a redução dos atributos, adquire-se mais consciência de como os atributos se relacionam uns com os outros e contribuem para o resultado.

Não existe uma solução direta e simples para escolher quais atributos incluir em um modelo de ML. No entanto, existem diversas técnicas que podem ser usadas de maneira eficaz para selecionar os atributos (DHAL; AZAD, 2022), que podem ser agrupadas como métodos de filtro, envólucro e incorporados, como mostrado na Figura 3.1.

Figura 3.1. Feature selection methods.

Source: Adapted from (DHAL; AZAD, 2022).

3.3.1 Métodos de filtro

Os métodos de filtro são métodos de seleção de atributos que são executados antes mesmo da execução de um modelo, como um estágio de pré-processamento. Eles funcionam analisando aspectos de relacionamento entre as variáveis. Vários indicadores são usados para decidir quais elementos serão omitidos e quais permanecerão, com base no modelo que está sendo usado. Os métodos de filtro geralmente retornam uma classificação de atributos que informa como os recursos são organizados em comparação entre si e excluem as variáveis consideradas obsoletas.

Uma abordagem de filtro típica na regressão linear, por exemplo, é remover os atributos fortemente correlacionados entre si. Usar um limite de variação é outra abordagem de filtro. Essa abordagem estabelece um certo limite para a variação necessária entre os atributos para incorporá-los em um modelo. A lógica por trás disso é que as variáveis não se ajustarão significativamente se não tiverem uma alta variância e, portanto, não terão muito efeito sobre a variável dependente.

Alguns exemplos de métodos de filtro, utilizados nos resultados desse trabalho (Capítulo 5), são Chi-square (JIN *et al.*, 2006), Mutual Information (HOQUE *et al.*, 2014), Anova F-value (DING *et al.*, 2014), Fisher Score (HART *et al.*, 2000) e MultiSURF (URBANOWICZ *et al.*, 2018).

3.3.2 Métodos de invólucro

Fazem várias combinações de atributos para treinar modelos e, em seguida, medir a eficiência. Os métodos de invólucro avaliam o subconjunto ideal de atributos. Um determinado subconjunto é usado para treinar modelos e, em seguida, é avaliado em um conjunto de teste. Como os métodos de invólucro consomem muito tempo, usá-los para grandes conjuntos de atributos se torna difícil. Como seria de se esperar, os métodos de invólucro podem acabar sendo muito caros computacionalmente, mas ao decidir o subconjunto ideal, eles são altamente eficientes.

A eliminação de atributos recursivos, que começa com todos os atributos presentes em um modelo e os elimina um a um, é um exemplo de método de invólucro em regressão linear. A seleção direta é o inverso desse processo, começando com um único atributo e prosseguindo pela adição dos demais na busca pela melhora da eficiência do modelo.

Alguns exemplos de métodos de invólucro, utilizados nos resultados desse trabalho (Capítulo 5), são Recursive Feature Elimination ([SELJO-PARDO *et al.*, 2019](#)), Permutation Importance ([ALTMANN *et al.*, 2010](#)), Shapley additive explanations ([LUNDBERG; LEE, 2017](#)) e Boruta ([KURSA; RUDNICKI, 2010](#)).

3.3.3 Métodos incorporados

Os métodos incorporados são métodos de seleção de atributos usados no projeto do algoritmo de aprendizado de máquina. A regularização, particularmente a regularização Lasso, é a forma mais popular de método incorporado, pois tem o potencial de reduzir automaticamente a coleção de atributos. A regressão de Lasso é muitas vezes referida como regularização de norma L1.

Alguns exemplos de métodos incorporados, utilizados nos resultados desse trabalho (Capítulo 5), são Embedded Random Forest e Embedded LightGBM ([SAEYS *et al.*, 2008](#)).

3.3.4 Algoritmo proposto

Em (NIYAS; P, 2021b), os autores realizaram uma seleção de atributos baseada no *Fischer score* sobre as bases de dados da ADNI e AIBL, avaliaram as métricas da avaliação da classificação para as classes CN, MCI e AD e os resultados foram promissores, segundo os autores.

Considerando um conjunto de dados (\mathbf{x}, y) com n instâncias, d atributos e c classes, representado por $\{(\mathbf{x}_i, y_i)_{i=1}^n\}$, onde $\mathbf{x}_i \in \mathbb{R}^d$ corresponde aos valores dos atributos da instância i e $y_i \in \{1, 2, \dots, c\}$ corresponde ao rótulo recebido pela instância i , o problema da seleção de atributos é encontrar um subconjunto de tamanho $m < d$ que contém a informação máxima representativa do conjunto de dados.

Uma das possibilidades de resolver esse problema é utilizar o escore de Fisher para selecionar atributos do conjunto de dados, de tal forma que, no espaço de dados delimitado pelos atributos selecionados, as distâncias entre os pontos de dados em classes diferentes são tão grandes quanto possível, enquanto as distâncias entre pontos de dados da mesma classe são as menores possíveis. Desse modo, esse mecanismo de seleção de atributos pode ser classificado na categoria de método de filtro.

Tomando μ_k^j e σ_k^j , como a média e o desvio padrão da k -ésima classe, do j -ésimo atributo, pode-se calcular o escore de Fisher como:

$$F(\mathbf{x}^j) = \frac{\sum_{k=1}^c n_k (\mu_k^j - \mu^j)^2}{(\sigma^j)^2}, \quad (3.1)$$

onde $(\sigma^j)^2 = \sum_{k=1}^c n_k (\sigma_k^j)^2$, sendo que μ^j e σ^j correspondem à média e ao desvio padrão de todo o conjunto de dados referente ao j -ésimo atributo, e n_k corresponde ao tamanho da k -ésima classe no espaço reduzido de dados.

Após o cálculo do escore de Fisher para cada atributo, faz-se a seleção dos m atributos com a maior pontuação do escore de Fisher. Devido ao fato de que a pontuação de cada atributo é calculada independentemente, os atributos selecionados pelo critério do escore de Fisher são subótimos.

A revisão do *Fischer score* como algoritmo de seleção de atributo é para lembrar que duas

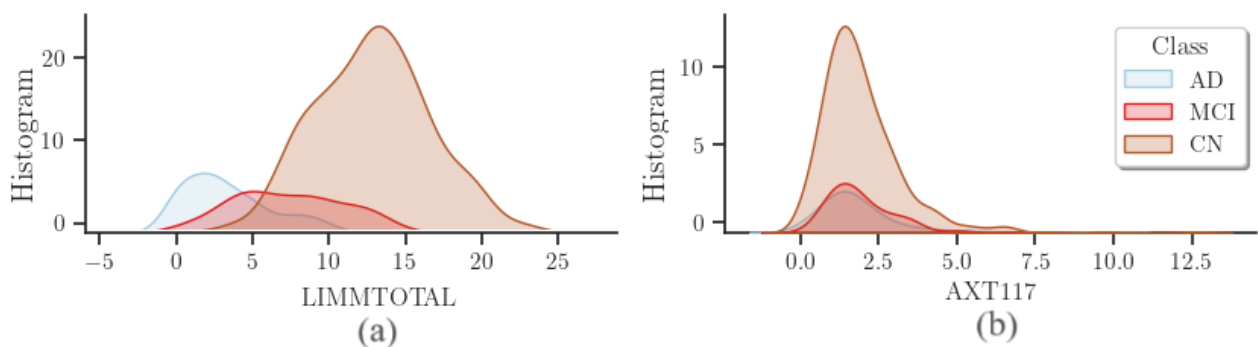
grandezas estatísticas (média e variância, ou desvio padrão) são consideradas para a construção do resultado.

3.3.4.1 O algoritmo FMFS

Esta tese propõe um novo método de avaliação de atributos, que foi batizado de FMFS (do inglês: *first moment feature selection*), baseado no cálculo do primeiro momento estatístico com relação à origem, dos valores dos atributos com relação às suas classes.

O FMFS surge da observação do histograma das classes em cada atributo. Na Figura 3.2, nota-se que as classes do atributo LIMMTOTAL estão melhor separadas que no atributo AXT117 da base AIBL. A hipótese então é que o atributo AXT117 possa ser excluído da base de dados por contribuir menos na distinção entre as classes, relativamente ao atributo LIMMTOTAL.

Figura 3.2. Class histograms for the LIMMTOTAL(a) and AXT117(b) features in the AIBL dataset.



Source: Own authorship.

O FMFS está descrito pelos seguintes passos:

1. Escalonamento de todas os atributos pela média global da base;
2. Cálculo das médias das classes em cada recurso e acúmulo da distância euclidiana entre as médias;
3. Varredura logarítmica de 100 limiares no intervalo entre as distâncias máxima e mínima;
4. Seleção dos atributos que superam o limiar, varridos em ordem decrescente;

5. Avaliação do modelo para a base selecionada em cada limiar;
6. Escolha da base que resultou na melhor acurácia.

Matematicamente, o FMFS pode ser construído a partir das equações a seguir, atendendo a nomenclatura definida na Figura 3.3, onde \mathbf{x} representa toda a base de dados com \mathbf{M} instâncias e \mathbf{N} atributos, $x_{i,j}$ corresponde à amostra da i -ésima instância e j -ésimo atributo. O valor y_{i,c_k} corresponde à classe c_k associada a i -ésima instância. As possíveis classes variam de c_1 até c_p .

Figura 3.3. Nomenclature definition.

$$\mathbf{x} = \mathbf{M} \text{ instances} \left\{ \begin{array}{l} \overbrace{\left[\begin{array}{cccc} \cdot & & & \\ \cdot & & & \\ \cdot & \cdot & x_{i,j} & \cdot \\ \cdot & & & \\ \cdot & & & \end{array} \right]}^{\mathbf{N} \text{ features}} \\ i \in [0 \dots \mathbf{M}-1] \\ j \in [0 \dots \mathbf{N}-1] \\ x \in R^{\mathbf{M} \times \mathbf{N}} \end{array} \right. \quad y = \begin{bmatrix} \cdot \\ \cdot \\ y_{i,c_k} \\ \cdot \\ \cdot \end{bmatrix} \begin{array}{l} c_k \in [c_1 \dots c_p] \\ i \in [0 \dots \mathbf{M}-1] \\ y \in R^{\mathbf{N} \times 1} \end{array}$$

Source: Own authorship.

Assim, pode-se calcular a média global do conjunto de dados - μ_g , pela Equação 3.2.

$$\mu_g = \frac{1}{\mathbf{N}} \frac{1}{\mathbf{M}} \sum_{i=0}^{\mathbf{M}-1} \sum_{j=0}^{\mathbf{N}-1} x_{i,j} \quad (3.2)$$

A média de cada atributo - μ_j pode-se ser calculada pela Equação 3.3.

$$\mu_j = \frac{1}{\mathbf{M}} \sum_{i=0}^{\mathbf{M}-1} x_{i,j} \quad (3.3)$$

Com as médias μ_g e μ_j calculadas, faz-se o escalonamento de todas as amostras, resultado em $z_{ij} = x_{ij}\mu_g/\mu_j$, de tal modo que todos os atributos passam a ter a mesma média μ_g .

Calcula-se a média de cada classe c_k para cada um dos atributos j - μ_{j,c_k} , pela Equação 3.4.

$$\mu_{j,c_k} = \frac{1}{\mathbf{M}_{j,c_k}} \sum_{i=0}^{\mathbf{M}_{j,c_k}-1} z_{i,j} \text{ onde } (y_j = c_k) \quad (3.4)$$

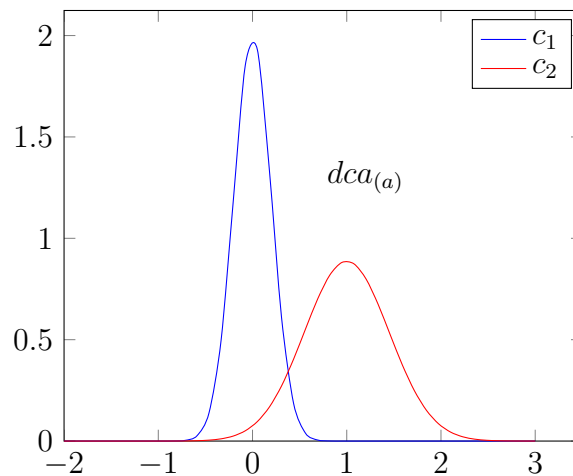
onde \mathbf{M}_{j,c_k} corresponde ao número de amostras que o atributo j possui da classe c_k .

Finalmente, a distância entre todas as médias das classes de cada atributo j - dca_j , pode ser calculada pela Equação 3.5.

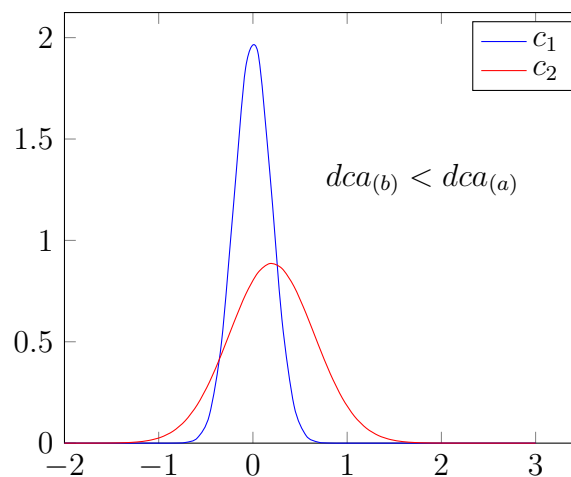
$$dca_j = \sqrt{\sum_{\substack{0 \leq k \leq p \\ 0 \leq q \leq p \\ k \neq q}} (\mu_{j,c_k} - \mu_{j,c_q})^2} \quad (3.5)$$

A Figura 3.4 mostra um exemplo hipotético do histograma das classes c_1 e c_2 em dois atributos 1 e 2, respectivamente em (a) e (b). No atributo 1, as médias das classes c_1 e c_2 estão mais distantes do que as médias das mesmas classes no atributo 2. Nessa condição, o valor de dca_1 certamente será maior que o valor de dca_2 , que é o resultado esperado calculado pelo método FMFS.

Figura 3.4. Features with different class means.



(a) Histogram of feature 1's classes c_1 and c_2 .



(b) Histogram of feature 2's classes c_1 and c_2 .

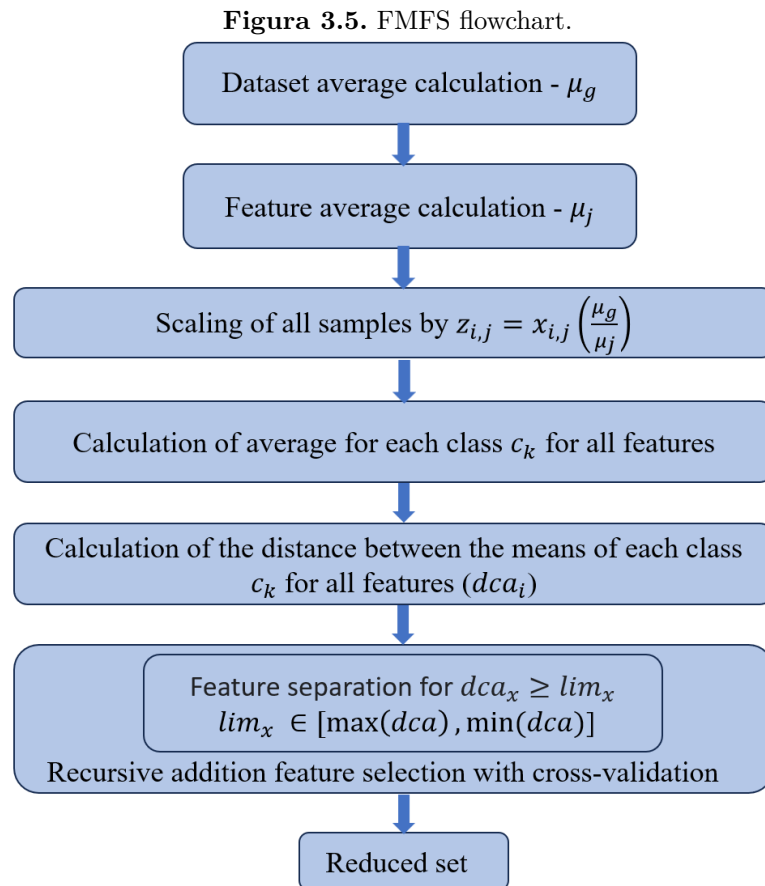
Source: Own authorship.

Os atributos que irão compor o conjunto de dados selecionados deverão ser os que possuem os maiores valores de dca , tendo em vista que o valor de dca reflete o grau de separação das classes em cada atributo. Assim, foi proposta uma varredura *top-down*, partindo de $\max(dca)$ até $\min(dca)$. O número de limiares a serem avaliados e também o tipo de escala (linear ou logarítmica) podem produzir resultados mais ou menos adequados. De modo geral, considerando uma quantidade de w limiares, os valores podem ser calculados como na Equação 3.6, para uma varredura linear e, como na Equação 3.7, para a varredura logarítmica.

$$lim_x = \max(dca) - x (\max(dca) - \min(dca)) / w \quad (3.6)$$

$$lim_x = \max(dca) - x \frac{(\max(dca) - \min(dca))}{w} \log \left(\frac{(x+1)}{w} - 1 \right) \quad (3.7)$$

A Figura 3.5 mostra um fluxograma do algoritmo FMFS, com destaque para os blocos principais.



Source: Own authorship.

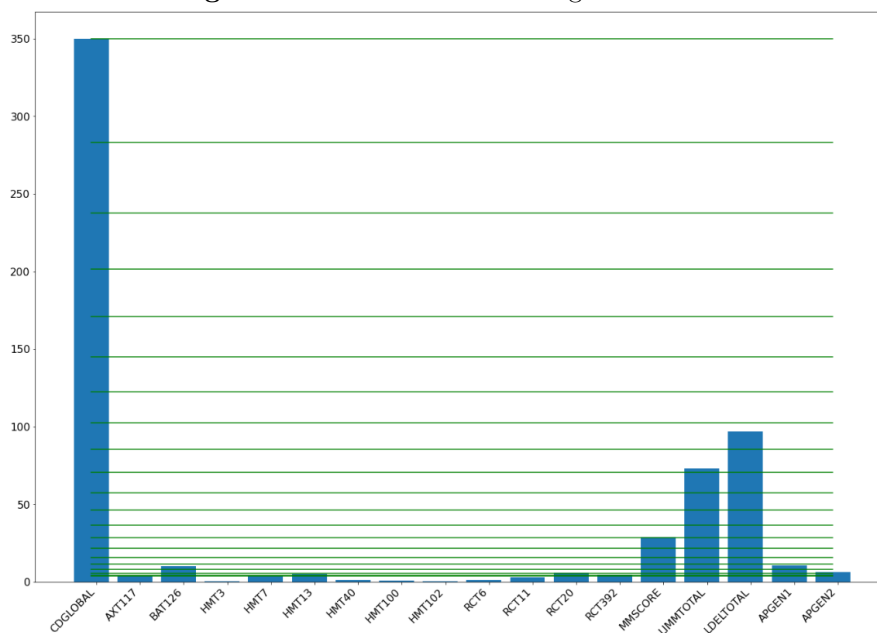
Para cada lim_x considerado, são selecionados todos os atributos em que $dca \geq lim_x$. A base de dados selecionada é então submetida aos classificadores para avaliação do desempenho e, ao final, faz-se a escolha da base de dados que deu o melhor desempenho, com o menor conjunto de atributos possível.

3.3.4.2 Aplicação do FMFS às bases de dados

O algoritmo FMFS foi aplicado às bases de dados ADNI, AIBL e OASIS, considerando uma varredura logarítmica com $w = 100$, experimentalmente determinado. Para cada uma das bases de dados, foram plotados os dados do ranqueamento da importância dos atributos e também a medida de acurácia da avaliação dos classificadores kNN e SVM para cada conjunto de atributos selecionados pela varredura logarítmica.

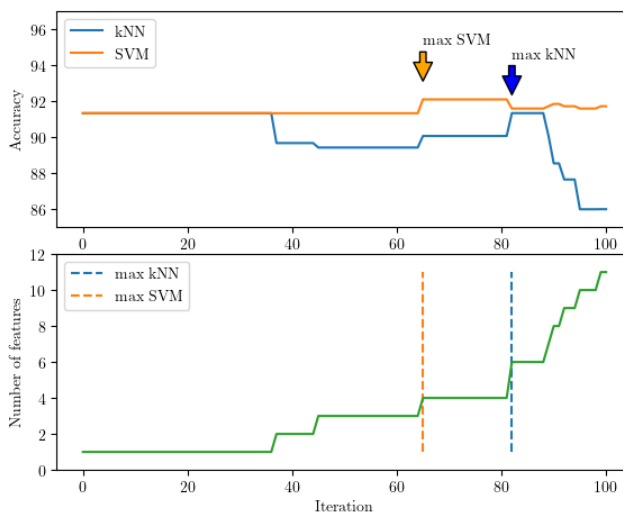
Na Figura 3.6, está mostrada graficamente a avaliação da métrica de cada atributo segundo o algoritmo proposto para a base de dados AIBL. Nota-se com clareza a grande relevância das métricas dos atributos CDGLOBAL, LDELTOTAL, LIMMTOTAL e MMSCORE em detrimento dos demais atributos. Os atributos que apresentam as menores métricas tendem a ter valores muito próximos, o que levou a uma varredura do maior para o menor valor da métrica e seguindo uma escala logarítmica, para montagem da base de dados a ser avaliada. Na Figura 3.7, pode-se acompanhar a evolução do desempenho de cada conjunto de dados selecionado pela varredura e avaliado pelos algoritmos kNN e SVM. As linhas verticais identificam os limiares do desempenho ótimo para cada um dos algoritmos.

Figura 3.6. AIBL feature ranking - *dca* values.



Source: Own authorship.

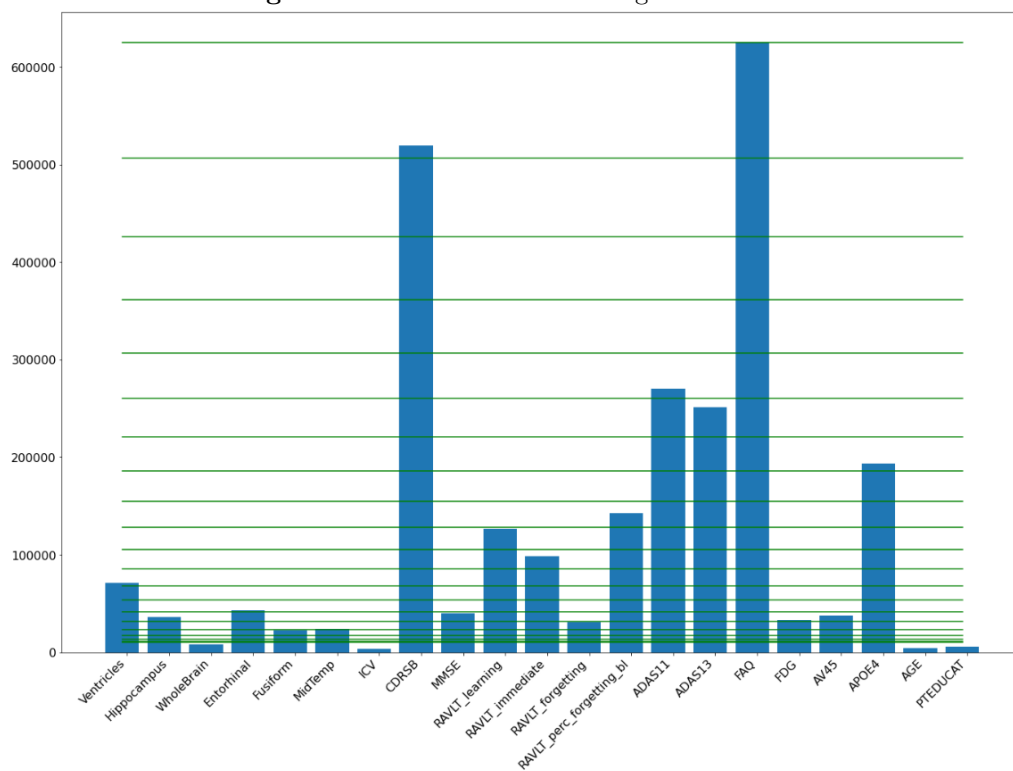
Figura 3.7. AIBL performance measurement. a) The orange and blue curves represent the evolution of SVM and kNN accuracy, respectively, with indicated maximums. b) The green curve represents the evolution of the number of selected features. The orange and blue vertical lines represent the point of maximum accuracy for SVM and kNN respectively.



Source: Own authorship.

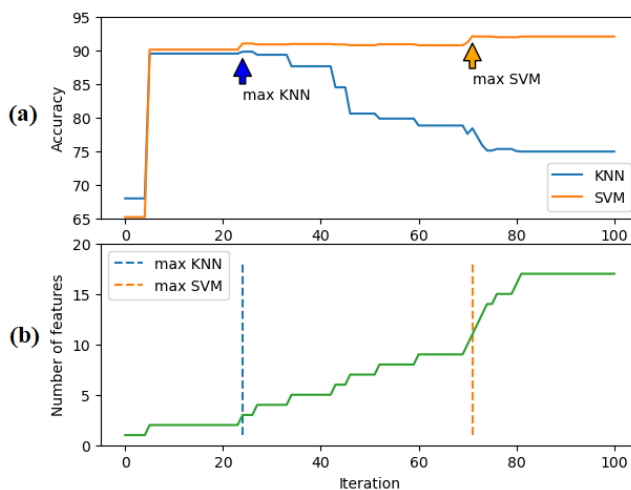
Da mesma forma, a Figura 3.8 mostra graficamente a avaliação da métrica de cada atributo segundo o algoritmo proposto para a base de dados ADNI e a Figura 3.9 a evolução do desempenho de cada conjunto de dados selecionados e os pontos de seleção do desempenho ótimo para cada algoritmo kNN e SVM.

Figura 3.8. ADNI feature ranking - *dca* values.



Source: Own authorship.

Figura 3.9. ADNI performance measurement. a) The orange and blue curves represent the evolution of SVM and kNN accuracy, respectively, with indicated maximums. b) The green curve represents the evolution of the number of selected features. The orange and blue vertical lines represent the point of maximum accuracy for SVM and kNN respectively.

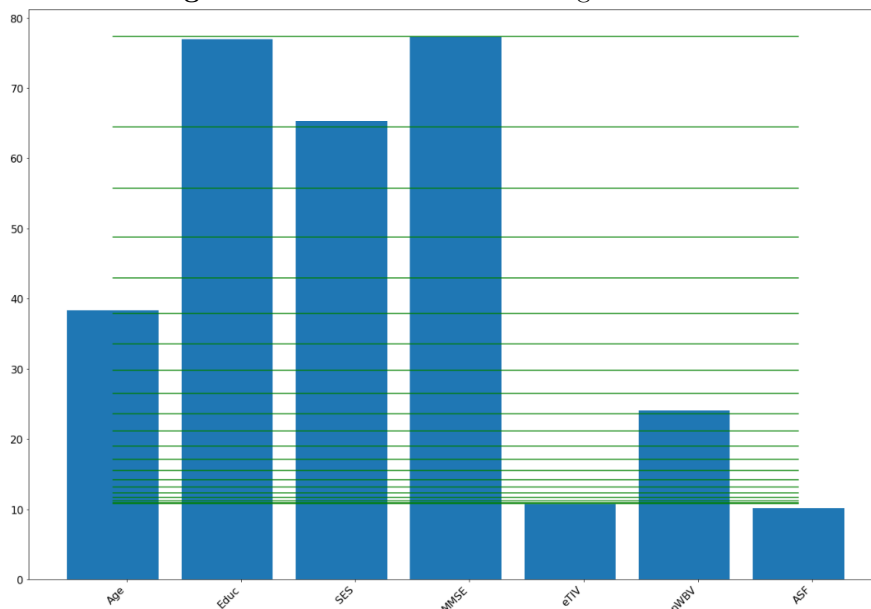


Source: Own authorship.

Finalmente, a Figura 3.10 mostra graficamente a avaliação da métrica de cada atributo segundo o algoritmo proposto para a base de dados OASIS e a Figura 3.11 a evolução do desempenho de cada conjunto de dados selecionados e os pontos de seleção do desempenho

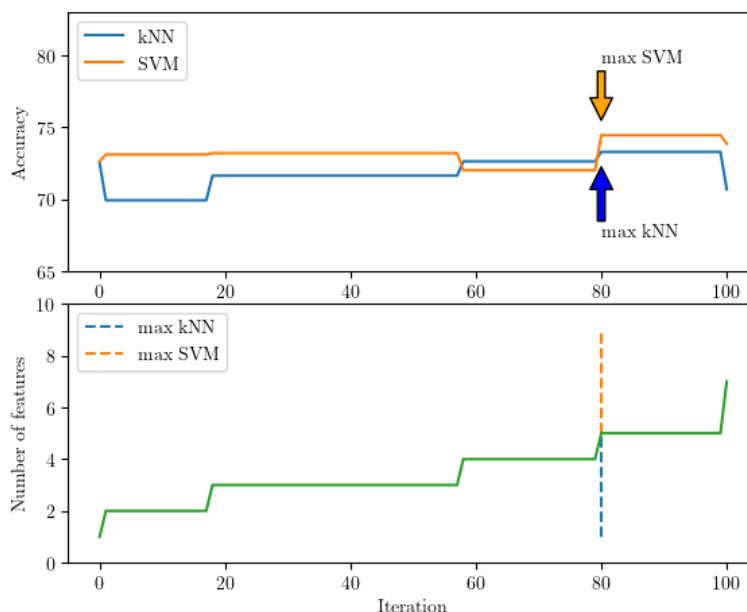
ótimo para cada algoritmo kNN e SVM.

Figura 3.10. OASIS feature ranking - *dca* values.



Source: Own authorship.

Figura 3.11. OASIS performance measurement. a) The orange and blue curves represent the evolution of SVM and kNN accuracy, respectively, with indicated maximums. b) The green curve represents the evolution of the number of selected features. The orange and blue vertical lines represent the point of maximum accuracy for SVM and kNN respectively.



Source: Own authorship.

3.4 MÉTRICAS DE AVALIAÇÃO DE DESEMPENHO

Como base no conceito da matriz de confusão (Figura 3.12), que dá os resultados numéricos de erros e acertos do modelo avaliado sobre o conjunto de testes, pode-se definir os seguintes possíveis resultados: TP - verdadeiros positivos, TN - verdadeiros negativos, FP - falsos positivos e FN - falsos negativos.

Figura 3.12. Confusion matrix.

		Predicted class	
		<i>P</i>	<i>N</i>
Actual Class	<i>P</i>	True Positives (TP)	False Negatives (FN)
	<i>N</i>	False Positives (FP)	True Negatives (TN)

Source: Own authorship.

Assim, é possível definir as métricas que são utilizadas neste trabalho. As métricas estão calculadas para cada classe e o valor final é a média macro desses valores.

3.4.0.1 Acurácia - A_{cy}

A métrica elementar de avaliação de qualquer sistema de classificação é a acurácia, como definido na Equação (3.8). A acurácia é calculada pela razão entre o número de resultados corretos e a soma de todos os resultados obtidos.

$$A_{cy} = \frac{TP + TN}{TP + FP + FP + TN} \quad (3.8)$$

3.4.0.2 Precisão - P_{rn}

A precisão é uma medida da exatidão do diagnóstico, ou seja, a proporção dos doentes diagnosticados por um sistema, que foram realmente atingidos pela doença. Matematicamente, pode ser definido como na Equação (3.9). É calculada como a razão entre o número de resultados positivos corretos e o número total de resultados positivos.

$$P_{rn} = \frac{TP}{TP + FP} \quad (3.9)$$

3.4.0.3 Sensibilidade - S_{ny}

Enquanto A_{cy} lida com ambos os resultados positivos e negativos, o desempenho de um modelo específico em termos de detecção de resultados positivos é avaliado usando S_{ny} (ou *recall*). S_{ny} é também chamada por taxa de verdadeiro positivo, e é dado pela proporção de resultados positivos que são corretamente considerados positivos em relação a todos os possíveis resultados positivos, calculado conforme a Equação (3.10).

$$S_{ny} = \frac{TP}{TP + FN} \quad (3.10)$$

3.4.0.4 Especificidade - S_{py}

A S_{py} é também chamada por taxa de verdadeiro negativo, e é dada pela proporção de resultados negativos que são corretamente considerados negativos em relação a todos os possíveis resultados negativos. A Equação (3.11) faz o cálculo da S_{py} e mede o desempenho de um modelo específico em termos de detecção de resultados negativos.

$$S_{py} = \frac{TN}{TN + FP} \quad (3.11)$$

3.4.0.5 Acurácia balanceada - B_{ca}

A B_{ca} é uma métrica utilizada para medir a acurácia de um conjunto de dados que possui um número desbalanceado de exemplos para cada classe. Numa situação de base de dados desbalanceada, o cálculo da A_{cy} leva a uma avaliação otimista do modelo. A B_{ca} pode ser calculada como a média aritmética entre a taxa de verdadeiro positivo e a taxa de verdadeiro negativo, conforme a Equação (3.12).

$$B_{ca} = \frac{S_{ny} + S_{py}}{2} \quad (3.12)$$

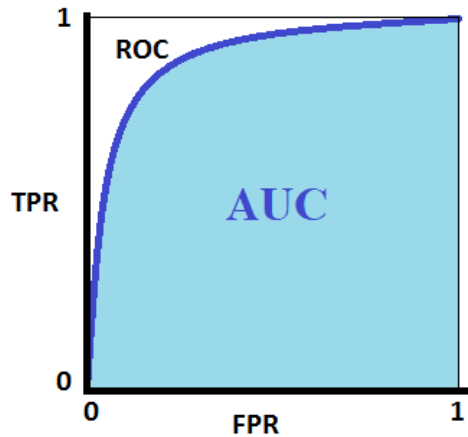
3.4.0.6 Pontuação f1 - $F1$

A $F1$ foi projetada para também funcionar adequadamente em base de dados desequilibradas. O objetivo da $F1$ é combinar as métricas P_{rn} e S_{ny} em uma única métrica, realizada a média harmônica entre essas duas métricas, conforme mostrado na Equação (3.13).

$$F1 = \frac{2}{\frac{1}{P_{rn}} + \frac{1}{S_{ny}}} \quad (3.13)$$

3.4.0.7 Área sobre a curva ROC - AUC

A *Receiver Operating Characteristic* (ROC) é uma curva de probabilidade avaliada em várias configurações de limite, entre a taxa de verdadeiros positivos (TPR) e a taxa de falsos positivos (FPR). A TPR é a S_{ny} , enquanto a FPR é dada por $1 - S_{py}$, formando o plano TPR x FPR, conforme pode-se ver na Figura 3.13, num exemplo hipotético. A área sobre a curva ROC (AUC) representa o grau ou medida de separabilidade entre as classes. Ela informa o quanto o modelo é capaz de distinguir entre as classes. Quanto maior a AUC, melhor o modelo é para prever a classe 0 como 0 e a classe 1 como 1. Por analogia, quanto maior a AUC, melhor o modelo distingue entre pacientes com e sem a doença.

Figura 3.13. Receiver Operating Characteristics (ROC) curve.

Source: Own authorship.

Em um modelo multiclasse, podemos plotar um número N de curvas ROC para cada uma das N classes usando a metodologia de um contra todos (*One vs ALL*). Conseqüentemente, a métrica passa a se chamar MAUC (*multi-class AUC*) e é a métrica utilizada neste trabalho, tendo em vista que as bases de dados possuem três classes de etapas da AD.

Diferentes métricas de desempenho implicam em diferentes conclusões para um modelo de classificação. Enquanto um modelo pode dar excelentes resultados em termos de P_{rn} , pode dar resultados muito pobres em termos de S_{py} . Baseado nessa lógica, resumimos os resultados em forma de tabela, incluindo as métricas de desempenho indicadas.

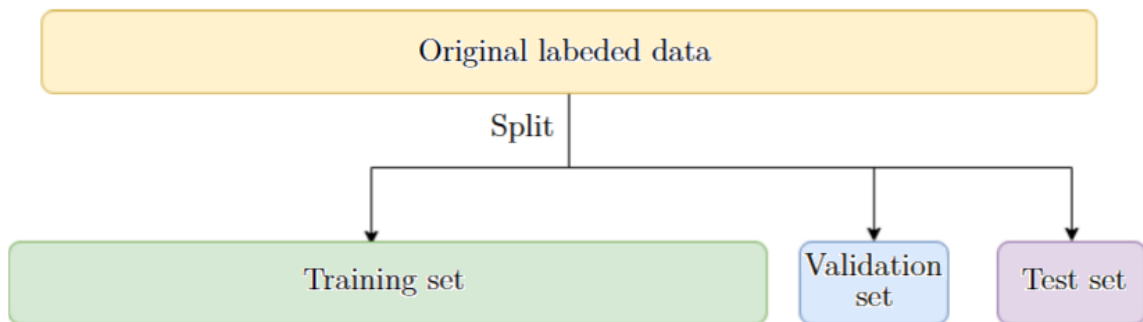
3.5 VALIDAÇÃO CRUZADA ANINHADA

Há sempre a necessidade de verificar a consistência do modelo de ML. Não se pode ajustar o modelo aos dados de treinamento e esperar que funcione com precisão para os dados reais que nunca foram vistos antes. É necessário alguma garantia de que o modelo tenha a maioria dos padrões dos dados aprendidos corretamente e que apresente baixo viés e variância. Assim, é recomendado que os dados originais sejam divididos em três partes: treino, validação e testes, conforme mostrado na Figura 3.14.

A divisão dos dados nas três partes naturalmente reduz a quantidade de dados que o modelo pode ser treinado e corre-se o risco de perder padrões/tendências importantes no conjunto de dados, o que, por sua vez, aumenta o erro induzido pelo viés.

Esse problema é resolvido pela técnica chamada *k-Fold*, onde k representa a quantidade de vezes em que o processo é repetido sobre os dados de treino e validação. Isso reduz significativamente o viés, pois está sendo usada a maioria dos dados para ajuste, e também reduz expressivamente a variância, pois a maioria dos dados também será usada no conjunto de validação. A troca dos conjuntos de treinamento e teste também aumenta a eficácia desse método. Essa técnica é chamada de validação cruzada e busca reduzir situações em que o modelo possa se tornar subajustando ou superajustando.

Figura 3.14. Train, validation and test data split.



Source: Own authorship.

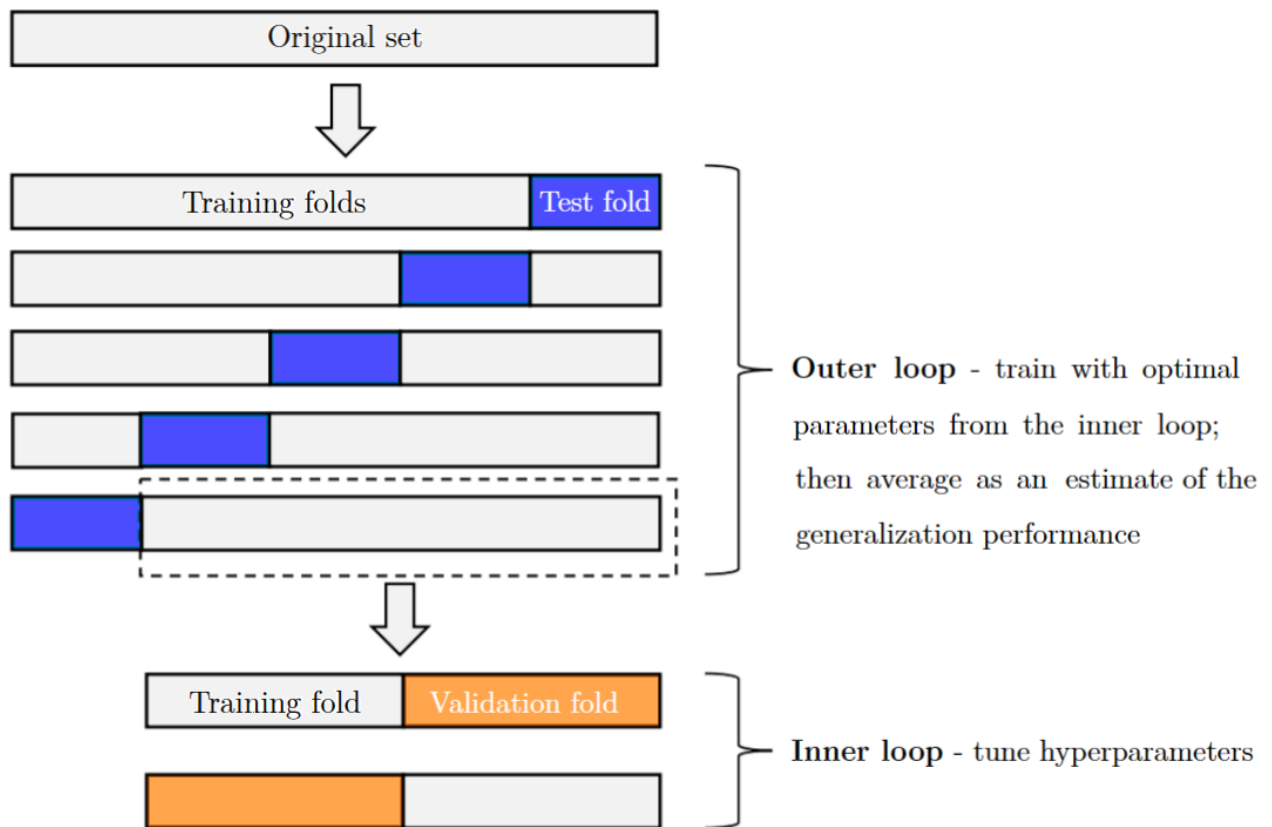
No entanto, um processo de seleção de modelo padrão geralmente inclui uma fase de otimização de hiper-parâmetros, na qual, através do uso de uma técnica de validação, como validação cruzada de *k-Fold*, um modelo ótimo será selecionado com base nos resultados de um teste de validação. No entanto, esse processo é vulnerável a uma forma de viés de seleção, o que o torna pouco confiável em muitas aplicações (CAWLEY; TALBOT, 2010).

Assim, nesta tese optou-se por utilizar um método baseado no *k-Fold*, chamado de validação cruzada aninhada (*nested cross-validation*), que realiza dois ciclos de *k-Fold* de forma aninhada (PARVANDEH *et al.*, 2020). A Figura 3.15 mostra como a validação cruzada aninhada é aplicada, em seis passos, como descritos a seguir:

- Passo 1: os dados são divididos em treino e teste pelo ciclo externo (cinco ciclos externos na ilustração). Para cada ciclo externo realizar o passo 2, iniciando no primeiro conjunto *Training folds*, na figura;
- Passo 2: dividir os dados em treino e validação no ciclo interno (dois ciclos internos na ilustração), para ajuste de hiper-parâmetros, usando busca em *grid*;

- Passo 3: utilizar o modelo de melhor desempenho para ser avaliado nos dados de teste do ciclo externo (*Test fold*);
- Passo 4: salvar os parâmetros do melhor modelo desse ciclo externo. Repetir os passos de 2 a 4 para os demais ciclos externos;
- Passo 5: escolher o melhor modelo do ciclo externo. Treinar sobre os dados completos para criar o modelo final;
- Passo 6: validar o modelo final sobre dados independentes.

Figura 3.15. Nested cross validation.

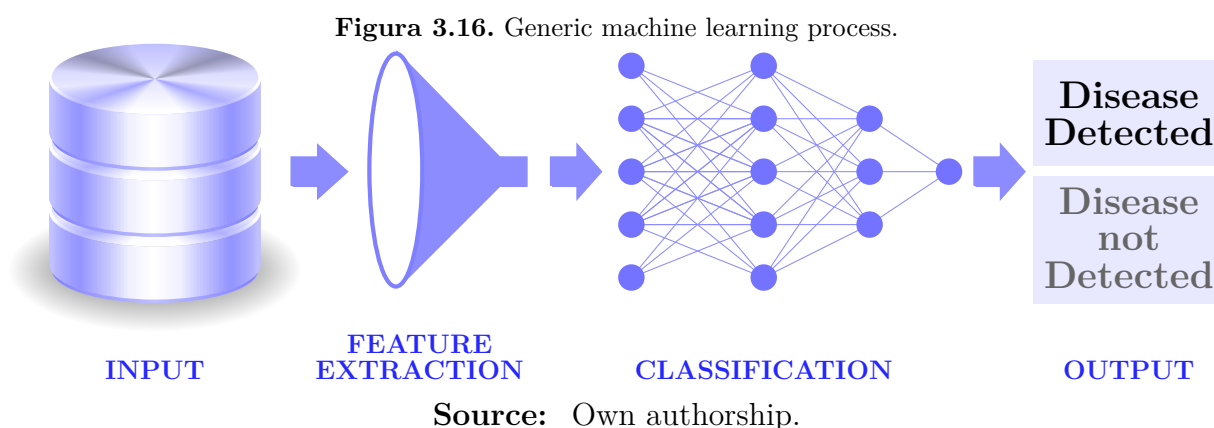


Source: Adapted from (PARVANDEH *et al.*, 2020).

Segundo (VARMA; SIMON, 2006), o uso adequado de validação cruzada para estimar o erro real de um classificador, requer que todas as etapas do algoritmo, incluindo o ajuste dos parâmetros do classificador, sejam repetidas em cada laço da validação cruzada. Nesse caso, a validação cruzada aninhada fornece uma estimativa quase imparcial do verdadeiro erro de classificação.

3.6 MODELOS DE ML UTILIZADOS

A tarefa de classificação em ML possui o esquema geral mostrado na Figura 3.16. Na primeira etapa, faz-se o pré-processamento e a análise dos dados na direção de se construir um conjunto de dados que seja representativo do objetivo que se almeja. A etapa seguinte é a de seleção dos atributos que possam facilitar o processo de classificação, removendo atributos de menor importância, reduzindo a dimensionalidade, simplificando o conjunto de dados na direção de melhor desempenho. Um dos objetivos desta tese foi apresentar um algoritmo de seleção de atributos - FMFS. Por fim, realiza-se a etapa de classificação na busca do melhor desempenho para o diagnóstico da doença, no caso desta tese. Faz-se a ressalva de que o esquema de classificador que aparece na Figura 3.16 mais parece com uma rede neural, mas ele apenas representa todos os outros modelos de classificação disponíveis em ML.



Para os testes iniciais, optou-se por trabalhar com os classificadores SVM e kNN, por motivos de facilidade de comparação com outros estudos ([NIYAS; P, 2021b](#)) ([NGUYEN *et al.*, 2018](#)), os quais serão brevemente revisados nos tópicos a seguir.

3.6.1 SVM : support vector machine

O algoritmo SVM ([CORTES; VAPNIK, 1995](#)) é um dos mais populares algoritmos usados por cientistas de dados, é poderoso e generaliza muito bem, em muitos casos.

Na Figura 3.17, estão mostrados os vetores de suporte, marcados com quadrados cinza, definindo a margem para a maior separação entre as classes binárias, neste exemplo, que é

chamada de margem ótima. O hiperplano ótimo, que separa as duas classes, divide a margem ótima ao meio. O hiperplano é descrito pela Equação (3.14), onde $\mathbf{x} = [x_1 \ x_2]^T$ e $\mathbf{w} = [a \ 1]^T$, para o caso binário.

$$\mathbf{w}^T \mathbf{x} + b = 0 \quad (3.14)$$

Portanto, dado um \mathbf{x}_i , o classificador SVM vai decidir por $h(\mathbf{x}_i) = 1$ ou $h(\mathbf{x}_i) = -1$, segundo a Equação (3.15).

$$h(\mathbf{x}_i) = \begin{cases} 1 & \mathbf{w}^T \mathbf{x} + b \geq 0 \\ -1 & \mathbf{w}^T \mathbf{x} + b < 0 \end{cases} \quad (3.15)$$

O problema de encontrar os valores de \mathbf{w} e b é chamado de problema de otimização com restrições, que corresponde a achar o hiperplano ótimo e é realizado durante a fase de treinamento do SVM.

Dado um conjunto de dados $D = (\mathbf{x}_i, y_i) | \mathbf{x}_i \in R^n, y_i \in \{-1, 1\}_{i=1}^m$, onde \mathbf{x}_i corresponde às instâncias com n atributos e y_i corresponde aos m rótulos, pode-se calcular os valores de \mathbf{w} e b , que correspondem ao hiperplano ótimo, pelas Equações (3.16) e (3.17).

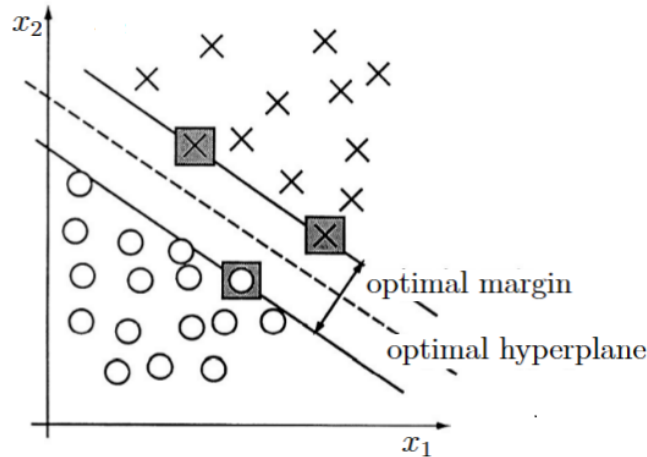
$$\mathbf{w} = \sum_{i=1}^m \alpha_i y_i \mathbf{x}_i, \quad (3.16)$$

onde α_i corresponde ao vetor que contém os multiplicadores de Lagrange para cada instância.

$$b = \frac{1}{S} \sum_{i=1}^S (y_i - \mathbf{w}^T \mathbf{x}), \quad (3.17)$$

onde S é o número de vetores de suporte.

Figura 3.17. SVM algorithm - the support vectors, marked with grey squares, define the margin of largest separation between the two classes.



Source: Adapted from (CORTESE; VAPNIK, 1995).

Uma estratégia utilizada no cálculo dos parâmetros ótimos para o SVM é inserir variáveis de folga às restrições do problema de otimização. Também é feito o acréscimo de um parâmetro de regularização C para determinar a importância da folga, o que significa o quanto se deseja evitar classificar incorretamente cada instância de treino. O problema de otimização regularizado passa a ser:

$$\max_{\alpha} \sum_{i=1}^m \alpha_i - \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m \alpha_i \alpha_j y_i y_j \mathbf{x}_i^T \mathbf{x}_j, \quad (3.18)$$

sujeito a

$$0 \leq \alpha_i \leq C, i = 1 \dots m, e \sum_{i=1}^m \alpha_i y_i = 0.$$

O parâmetro C : corresponde a um parâmetro de regularização incluído no processo de otimização do SVM, que flexibiliza o atendimento a todas as restrições para o cálculo do hiperplano ótimo. A presença de *outliers* ou de dados não linearmente separáveis pode tornar o cálculo do hiperplano ótimo quase impossível. Assim, com valores de C variando de zero até infinito, controla-se a rigidez no atendimento das restrições. Pequenos valores de C resultarão em uma margem mais ampla, ao custo de algumas classificações incorretas; valores grandes de C farão o classificador tolerar menos violações às restrições. É necessário encontrar um valor de C que não faça com que a solução seja impactada pelos dados ruidosos.

A função *kernel*: é utilizada para auxiliar na transformação dos dados que originalmente não são linearmente separáveis. A função *kernel* transforma os dados para outra dimensão que eventualmente possa facilitar encontrar um hiperplano que consiga separar os dados. Alguns exemplos de função *kernel* são: o *kernel* linear, definido por:

$$K(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{x}_i^T \mathbf{x}_j, \quad (3.19)$$

e o *kernel* polinomial, dado por:

$$K(\mathbf{x}_i, \mathbf{x}_j) = (\mathbf{x}_i^T \mathbf{x}_j + c)^d, \quad (3.20)$$

onde c é uma constante e d corresponde aos graus de liberdade da função.

O *kernel Radial Basis Function* (RBF), ou *Gaussian kernel*, é definido por:

$$K(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\gamma \|\mathbf{x}_i - \mathbf{x}_j\|^2), \quad (3.21)$$

onde um pequeno valor do parâmetro γ faz com que o modelo se comporte como um SVM linear e um grande valor fará com que o modelo seja fortemente impactado pelos exemplos dos vetores de suporte.

O algoritmo SVM faz parte de uma categoria chamada métodos de *kernel*, juntamente com os algoritmos *radial base functions* e *linear discriminate analysis*.

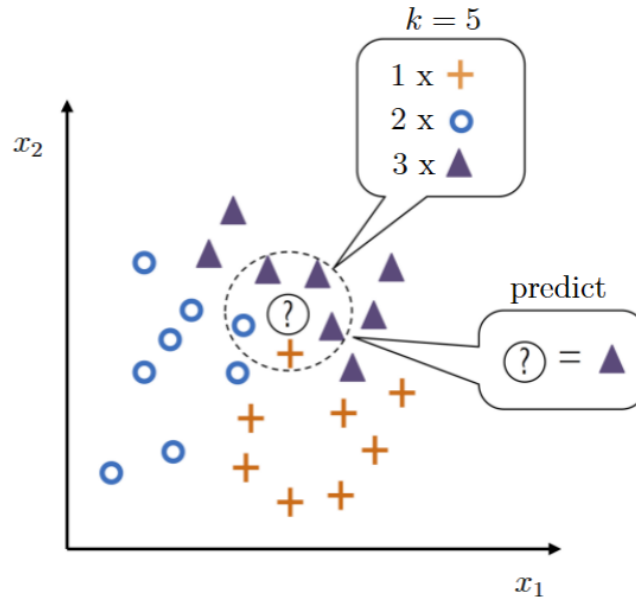
3.6.2 kNN : k-nearest neighbors

O algoritmo kNN ([COVER; HARTI, 1967](#)), que prima pela simplicidade, é um classificador de aprendizado supervisionado não paramétrico, que usa a proximidade para fazer classificações ou previsões sobre o agrupamento de um ponto de dados individual. O kNN trabalha com a suposição de que pontos semelhantes podem ser encontrados próximos uns dos outros.

Na Figura [3.18](#) é mostrado um exemplo com três classes (triângulo, círculo, cruz) e o ponto a ser classificado é representado por uma interrogação (?). O algoritmo avalia o número k de vizinhos mais próximos ($k = 5$, neste exemplo), contando o número de pontos pertencentes a

cada classe e decide pela classe que possui a maioria dos pontos (triângulo, no caso do exemplo).

Figura 3.18. kNN algorithm - example with three classes and $k = 5$.



Source: Adapted from ([RASCCHKA, 2018](#)).

Para determinar quais pontos de dados estão mais próximos de um determinado ponto de consulta, a distância entre o ponto de consulta e os outros pontos de dados precisará ser calculada. Vários tipos de medidas de distância podem ser utilizadas para realizar o kNN, a exemplo da distância euclidiana, distância de Manhattan, a distância de Hamming e distância de Minkowski, definida a seguir:

$$dist(\mathbf{x}, \mathbf{z}) = \left(\sum_{i=1}^d |x_i - z_i|^p \right)^{1/p}. \quad (3.22)$$

Dado um ponto de teste x e definindo o conjunto S_x dos k vizinhos mais próximos de x , selecionados pelo cálculo das distâncias entre x e o conjunto de dados, o classificador kNN vai decidir segundo:

$$h(x) = mode(y : (x, y) \in S_x), \quad (3.23)$$

onde a função $mode(\cdot)$ seleciona o rótulo de maior ocorrência em S_x .

O algoritmo kNN faz parte de uma categoria chamada métodos *instance-based*, juntamente com os algoritmos *learning vector quantization* e *self-organizing map*.

3.6.3 Outros modelos de ML

No sentido de explorar novas possibilidades e ampliar os resultados obtidos, também foram utilizados os classificadores *Random Forest* (BREIMAN, 2001), *Logistic Regression* (LAWALLEY, 2008), *Naive Bayes* (RISHI, 2001), *Decision Tree* (FREUND; MASON, 1999), *Ada Boosting* (SCHAPIRE, 2013), *Linear Discriminant Analysis* (BALAKRISHNAMA; GANAPATHIRAJU, 1998), *Gradiente Boosting* (FRIEDMAN, 2001), *Histogram Gradiente Boosting* (SHAFFER *et al.*, 1996) e *Extreme Gradiente Boosting* (CHEN; GUESTRIN, 2016)

Este capítulo fez a apresentação das bases de dados clínicas primárias utilizadas, o tratamento das bases para a montagem de banco de dados adequados para a avaliação do diagnóstico da AD, e as propostas tecnológicas desenvolvidas, com destaque para o algoritmo de seleção de atributos proposto. No próximo capítulo, são também feitas propostas tecnológicas para utilização de imagens de ressonância magnética no auxílio ao diagnóstico da AD.

ANÁLISE BASEADA EM IMAGENS

Este capítulo faz o detalhamento dos dados de imagens utilizados e as propostas tecnológicas desenvolvidas.

Técnicas de imagem em cortes transversais, como tomografia computadorizada e ressonância magnética, fornecem um método não invasivo valioso para a detecção da atrofia cortical que é normalmente observada na AD. A espectroscopia por ressonância magnética (MRI) e as técnicas de imagem metabólica, como a tomografia por emissão de pósitrons (PET), são ainda mais específicas e fornecem evidências de alterações metabólicas únicas nas vias da memória, fornecendo assim uma riqueza de novas informações para os investigadores (NORFRAY; PROVENZALE, 2004). A ressonância magnética possui a importante vantagem de não emitir radiação ionizante.

4.1 DIAGNÓSTICO DE AD BASEADO EM IMAGENS DE MRI

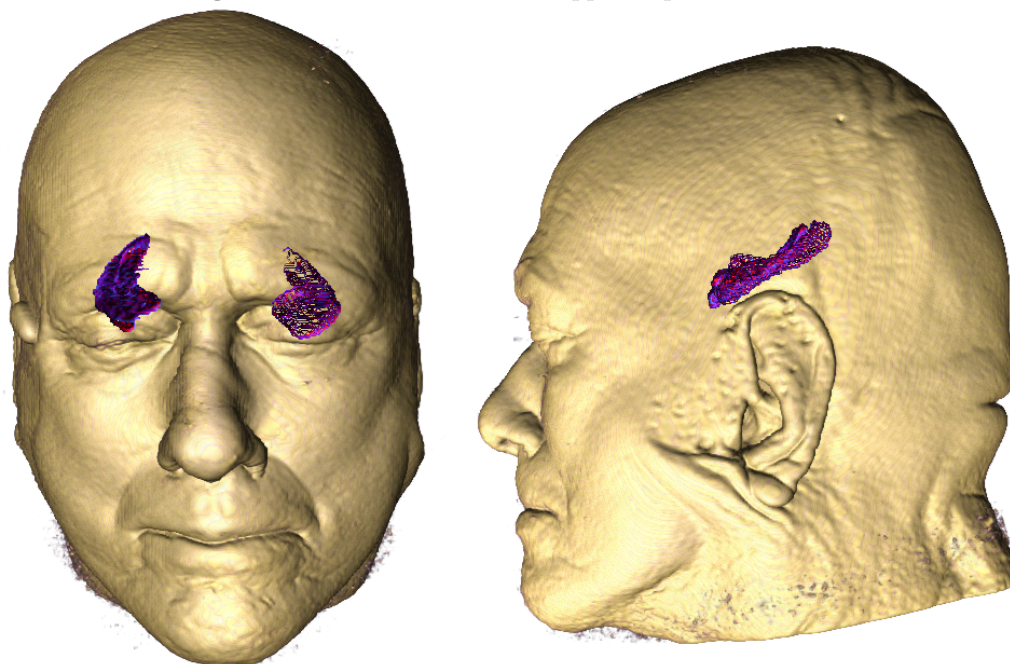
O principal papel da ressonância magnética (e da tomografia computadorizada) no diagnóstico da doença de Alzheimer é a avaliação da alteração de volume em locais característicos que podem render uma precisão diagnóstica de até 87% (NORFRAY; PROVENZALE, 2004).

A atrofia do lobo temporal mesial, particularmente do hipocampo, córtex entorrinal e córtex perirrinal, pode ser utilizada para realizar o diagnóstico de AD (PATEL *et al.*, 2020), de modo não invasivo.

Medições de volume cerebral, avaliadas com segmentação, demonstram que pacientes com doença de Alzheimer têm taxas aceleradas de perda de volume cerebral, tipicamente em torno de duas vezes o normal (1% vs 0,5% ao ano). Isso é ainda mais acentuado no hipocampo, com indivíduos afetados apresentando três vezes a perda de volume ao ano (4,5% vs 1,5% ao ano) (MILLER-THOMAS *et al.*, 2016).

A Figura 4.1 ilustra a posição e a estrutura do hipocampo na cavidade cerebral (imagem sob licença CC BY-NC-ND¹). O hipocampo é estrutura complexa do lobo temporal mesial e faz parte do sistema límbico e do circuito da memória, que mede cerca de 4,5 cm no seu diâmetro longitudinal, possui três partes: cabeça, também denominada pé do hipocampo, corpo e cauda (ROGACHESKI *et. al.*, 1998).

Figura 4.1. Illustration of the hippocampus in the brain.

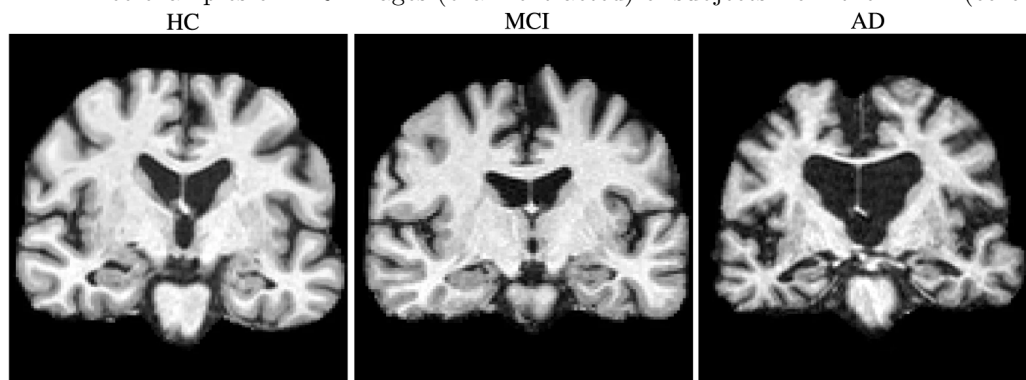


Source: Own authorship.

Em (LEDIG *et. al.*, 2018) são estudados volumes da estrutura cerebral a partir de uma base de dados de 5.074 MRI. Particularmente, é apresentada uma amostra de três casos de pacientes em diferentes estágios da AD, sendo um indivíduo de controle saudável HC (healthy control) - homem, 84,8 anos; indivíduo MCI - feminino, 71,8 anos; e um paciente com AD - homem, 77,5 anos. A Figura 4.2 mostra um corte coronal em cada um dos três pacientes, cada um em diferente estágio da AD.

¹<https://circulodeestudios-centrohistorico.blogspot.com/2015/09/sigue-comiendo-chatarra-y-se-te-va.html>

Figura 4.2. Three examples of MRI images (brain-extracted) of subjects from the ADNI (coronal slices).

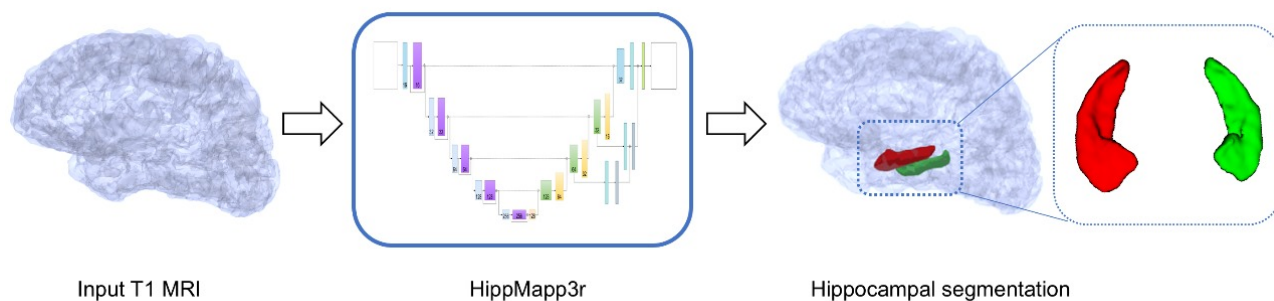


Source: Adapted from (LEDIG *et al.*, 2018).

A percepção de alterações e perdas de estruturas cerebrais fica evidenciada nas imagens de MRI e será explorada neste trabalho.

Em (GOUBRAN *et al.*, 2020) foi proposta a segmentação hipocampal total a partir da MRI. O algoritmo proposto, baseado em uma rede neural convolucional, foi chamado HippMapp3r. A Figura 4.3 mostra o esquema geral do processamento realizado.

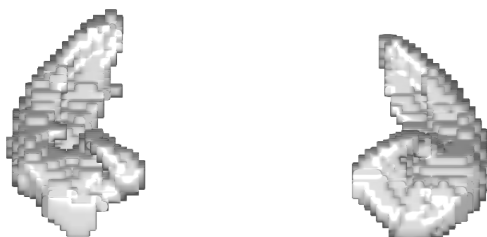
Figura 4.3. Proposed hippocampus extraction scheme.



Source: (GOUBRAN *et al.*, 2020).

Um exemplo de resultado gerado pelo HippMapp3r está mostrado na Figura 4.4.

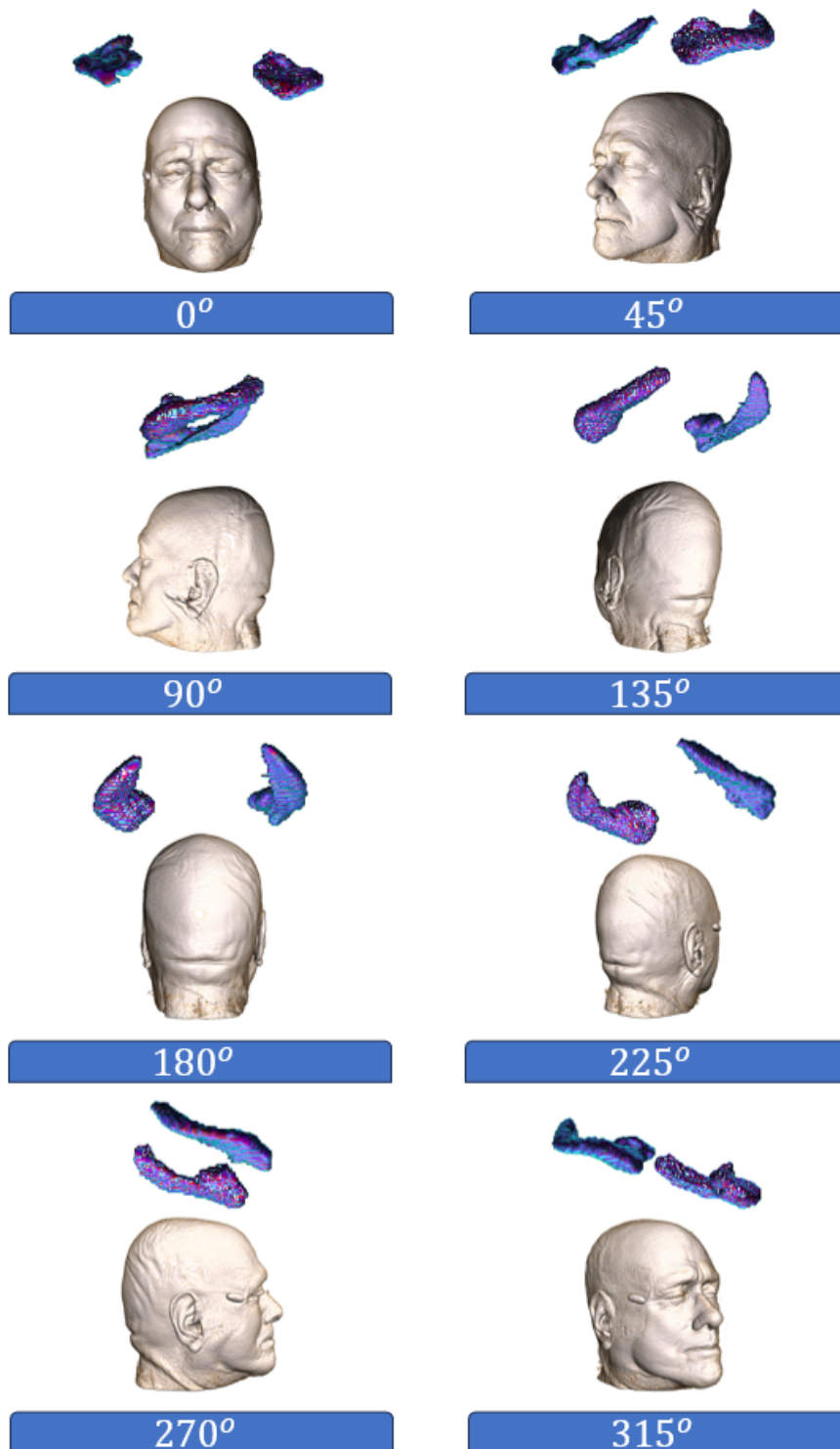
Figura 4.4. Extracted hippocampus.



Source: Own authorship.

A Figura 4.5 mostra o hipocampo destacado sobre a cabeça do indivíduo em vários ângulos de visualização. A cabeça do indivíduo e o hipocampo estão em escalas diferentes, mas no mesmo ângulo.

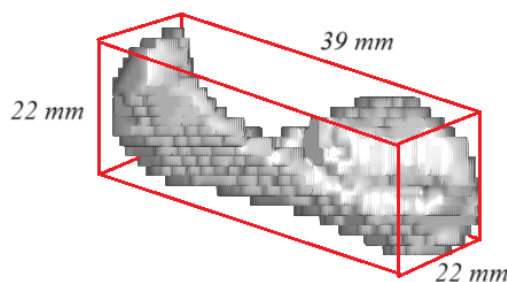
Figura 4.5. Visualization of the hippocampus from various angles..



Source: Own authorship.

Utilizando este mesmo exemplo, foi possível realizar a medição das dimensões máximas do hipocampo nos 3 planos (Figura 4.6), resultando em de 22 mm quando medido da esquerda para a direita, 22 mm quando medido de cima para baixo e 39 mm quando medido de frente para trás da cabeça do indivíduo.

Figura 4.6. Maximum dimensions of the hippocampus in the 3 planes. Right hippocampus rotated 300 degrees.



Source: Own authorship.

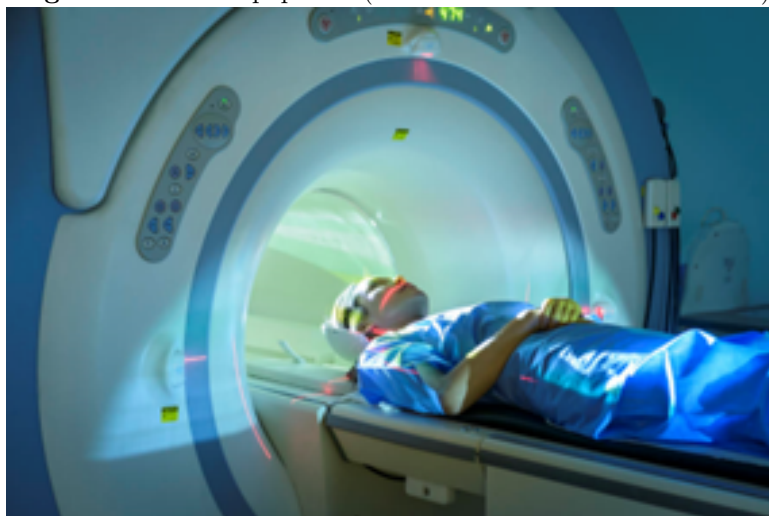
Essas medidas revelam que o plano coronal possui a maior quantidade de *slices* da MRI que conseguem cruzar alguma região do hipocampo.

4.2 IMAGENS DE RESSONÂNCIA MAGNÉTICA

A MRI é uma tecnologia de imagem não invasiva que produz imagens anatômicas tridimensionais detalhadas. É frequentemente usada para detecção de doenças, diagnóstico e monitoramento de tratamento. É baseada em tecnologia sofisticada que excita e detecta a mudança na direção do eixo de rotação dos prótons encontrados na água que compõe os tecidos vivos.

Para obter uma imagem de ressonância magnética, um paciente é colocado sob o efeito de um grande campo magnético e deve permanecer imóvel durante o processo de geração da imagem. Na Figura 4.7, é mostrado um ambiente com um equipamento de MRI numa visão característica (imagem sob licença CC BY-ND)([MELBOURNE, 2018](#)). Como a ressonância magnética não usa raios-x, e portanto não emite radiação ionizante, é a modalidade de imagem de preferida quando imagens frequentes são necessárias para diagnóstico ou terapia, especialmente no cérebro.

Figura 4.7. MRI equipment (creative commons license BY-ND).



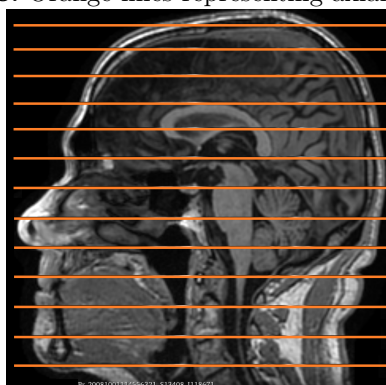
Source: (MELBOURNE, 2018).

4.2.1 Formato digital das imagens de ressonância magnética

Ao final do procedimento de realização do exame de ressonância magnética, tem-se um arquivo digital com os dados coletados, que resultam numa imagem das estruturas internas sob avaliação. O arquivo é composto por um cabeçalho e os dados dos *slices* nas três dimensões X, Y e Z.

A Figura 4.8 mostra a representação dos *slices* de um dos planos, obtidos pela MRI da cabeça de um paciente.

Figura 4.8. Orange lines representing axial MRI slices.



Source: Own authorship.

O formato do arquivo de imagem é muitas vezes um aspecto confuso para alguém que de-seja processar imagens médicas. Os formatos predominantes em radiologia são: *Digital Imaging*

and Communications in Medicine (DICOM), *Neuroimaging Informatics Technology Initiative* (NIFTI), *Philips MRI scanner formats* (PAR/REC), *Analyze 7.5* (ANALYZE), *Nearly Raw Raster Data* (NRRD) e *Montreal Neurological Institute* (MINC). Segundo (LAROUBIN; MURPHY, 2013), os principais formatos usados em imagens médicas são: ANALYZE, NIFTI, MINC e o DICOM.

4.2.1.1 Formato NIFTI

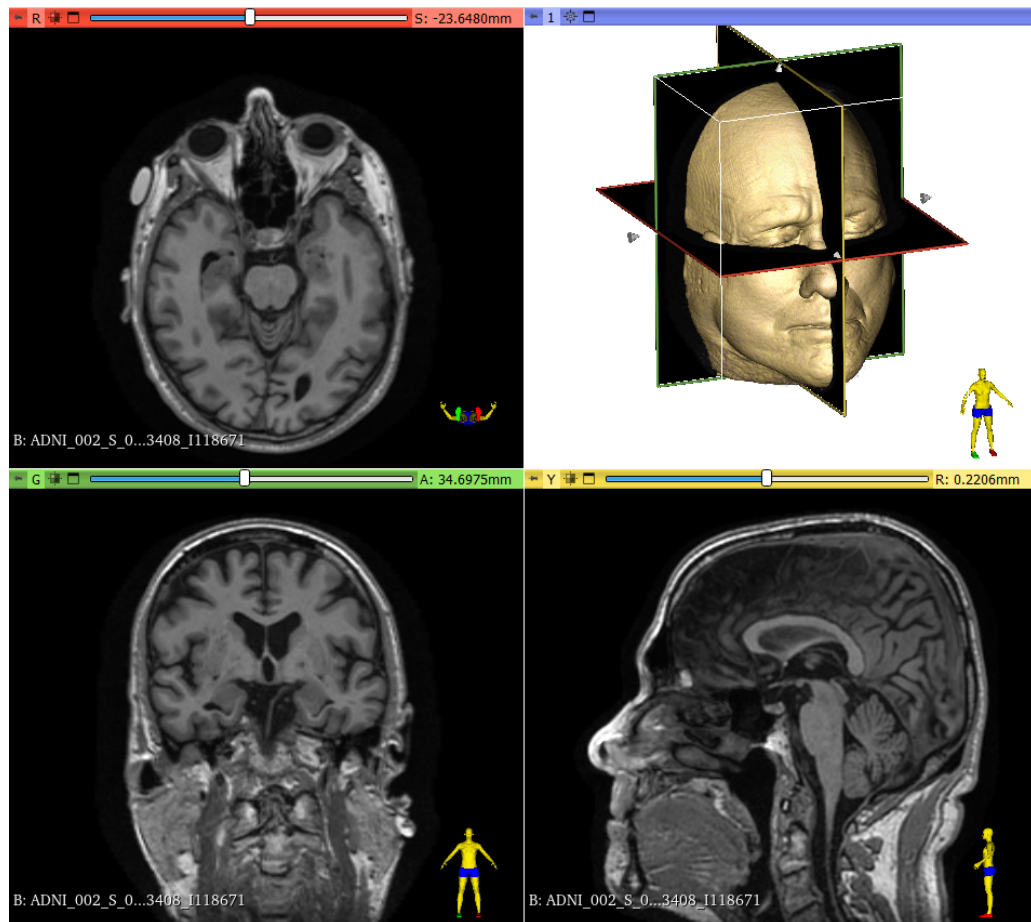
O formato NIFTI é muito usado em informática de imagem para neurociência e até mesmo nas pesquisas em neurorradiologia e foi utilizado neste trabalho. No atendimento clínico, o formato DICOM é padrão, no entanto, algumas ferramentas de informática de imagem podem converter arquivos DICOM para o formato NIFTI automaticamente (LI *et al.*, 2016).

A sigla NIFTI é um acrônimo que vem da *Neuroimaging Informatics Technology Initiative*, que foi um grupo de trabalho no NIH (*National Institutes of Health* nos EUA) no início dos anos 2000. A *Neuroimaging Informatics Technology Initiative* queria criar um tipo de arquivo que fosse semelhante, mas aprimorado a partir do formato de arquivo ANALYZE. Naquela época, o formato de arquivo MINC e o ANALYZE eram os tipos de arquivos comuns usados na pesquisa de neuroimagem.

De posse do arquivo NIFTI, é possível visualizar e processar a MRI para o objetivo específico. A Figura 4.9 mostra a visão do corte axial (barra vermelha), coronal (barra verde) e sagital (barra amarela) de uma MRI no formato NIFTI, de um paciente masculino com 85 anos, com diagnóstico neurológico normal. Também é possível visualizar um modelo 3D do crânio do paciente (barra azul), construído a partir das imagens dos três planos. A figura foi capturada a partir do software *3D Slicer image computing platform*², software gratuito e de código aberto para visualização, processamento, segmentação, registro e análise de imagens e malhas médicas, biomédicas e outras 3D; e planejamento e navegação de procedimentos guiados por imagem. O software *3D Slicer* está referenciado em (FEDOROV *et al.*, 2012).

²<https://www.slicer.org/>

Figura 4.9. Axial (red bar), coronal (green bar) and sagittal (yellow bar) slice views of an MRI in NIFTI format.



Source: Own authorship.

4.3 DETALHAMENTO DA BASE DE DADOS UTILIZADA

No site Kaggle³, por exemplo, existem algumas bases de dados de MRI utilizados na determinação do diagnóstico de AD⁴. Esta tese fez a opção por trabalhar com dados obtidos diretamente nos estudos referenciados sobre AD, como é o caso da *Alzheimers Disease Neuroimaging Initiative* (ADNI).

4.3.1 Estudo ADNI

O *Alzheimers Disease Neuroimaging Initiative* (ADNI)([MARINESCU et al., 2018](#)) é um estudo norte-americano multicêntrico e longitudinal projetado para desenvolver biomarcadores

³<https://www.kaggle.com/>

⁴<https://www.kaggle.com/datasets/tourist55/alzheimers-dataset-4-class-of-images>

clínicos, de imagem, genéticos e bioquímicos para a detecção precoce e rastreamento da AD. Desde o seu lançamento, há mais de uma década, a parceria público-privada histórica trouxe grandes contribuições para a pesquisa de AD, permitindo o compartilhamento de dados entre pesquisadores de todo o mundo. Informações complementares podem ser encontradas no site⁵ do estudo. Este estudo também classifica o estágio da progressão da doença em três categorias: CN, MCI e AD.

Os dados do estudo ADNI podem ser baixados na plataforma *Image and Data Archive* (IDA)⁶, que é gerenciada pelo Laboratory of Neuro Imaging (LONI) do Mark and Mary Stevens Neuroimaging and Informatics Institute associado à University of Southern California.

4.3.1.1 MRI no ADNI

A coleta de imagens se mostrou fundamental para que a ADNI pudesse atingir o objetivo de desenvolver biomarcadores para rastrear a progressão da AD. A aquisição das imagens foi realizada em várias fases, ADNI 1, ADNI GO, ADNI 2 e ADNI 3, com outras diversas variações das condições tecnológicas e de aquisição.

Neste trabalho, foram selecionados os dados de MRI da fase ADNI 1, particularmente do momento da triagem dos pacientes. Essa base de dados está padronizada pela ADNI e chama-se *ADNI1:Screening 1.5T*, que tem como características gerais:

- possui dados de 1075 pacientes, com idade entre 55 e 91 anos;
- coletados em uma única e primeira visita ao estudo;
- 619 pacientes do sexo masculino e 456 do sexo feminino;

Os dados da *ADNI1:Screening 1.5T* foram coletados no período de 03 de janeiro a 29 de setembro de 2006. Nessa oportunidade, os 1075 pacientes foram diagnosticados nos estágios CN, MCI e AD, de acordo com o estágio da AD, resultando no número de pacientes por classe conforme mostrado na Tabela 4.1.

⁵<https://adni.loni.usc.edu/>

⁶<https://ida.loni.usc.edu/login.jsp>

Tabela 4.1. ADNI1:Screening 1.5T - Number of patients per class

Class	Quantity
CN	307 patients
MCI	525 patients
AD	243 patients

Source: Own authorship.

Segundo a ADNI, o protocolo de ressonância magnética do *ADNI1:Screening 1.5T* objetivou adquirir imagens estruturais consistentes em *scanners* de 1,5 Tesla usando sequências ponderadas em T1 e T2. Na maioria dos ambientes clínicos, um *scanner* de ressonância magnética de 1,5 T é mais do que adequado para gerar imagens claras e fornecer evidências para diagnóstico e prognóstico. O *scanner* 3T também foi utilizado pela ADNI em fases de aquisição de dados, como no ADNI 3, para atingir outros objetivos.

Através do controle das radiofrequências de pulso e das ondas de gradientes, é possível controlar o destaque da imagem obtida. As imagens ponderadas em T1 mostram a anatomia de tecidos moles e gordura. Imagens ponderadas em T2 mostram com destaque os líquidos e patologias.

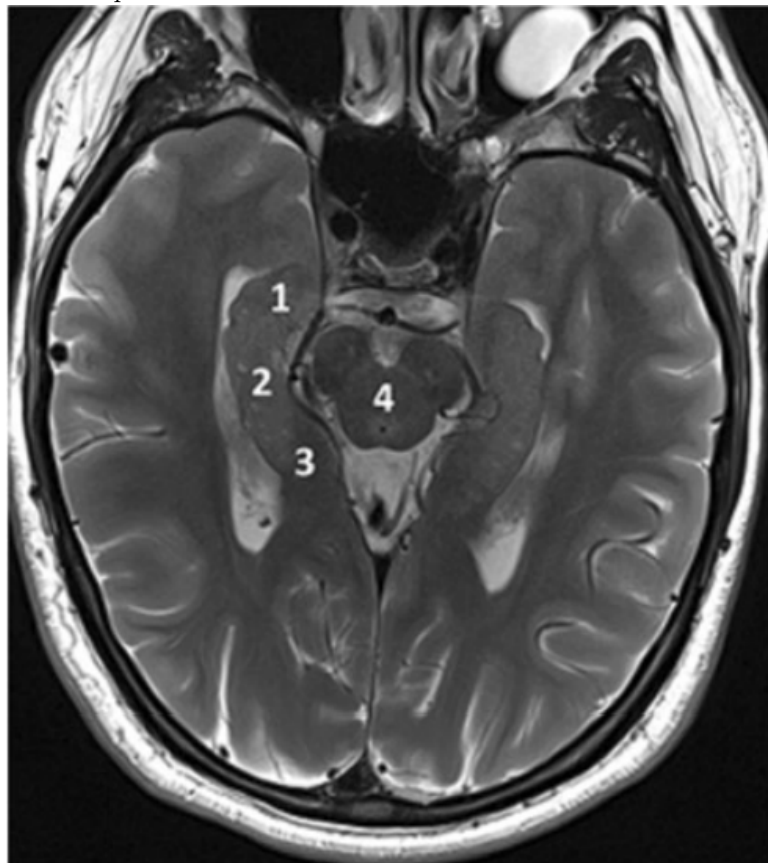
As imagens no *ADNI1:Screening 1.5T* estão disponíveis no formato NIFTI e podem ser baixadas na plataforma IDA. A Figura 4.9, antes já referenciada, mostra a MRI do paciente anonimizado 002_S_0295, que faz parte da base de dados *ADNI1:Screening 1.5T*, com destaque para um particular *slice* no plano axial, coronal e sagital.

4.3.1.2 Escolha dos *slices* para análise

Dado que o acesso e os dados primários foram definidos, o próximo desafio para esta tese foi o de fazer as escolhas adequadas das imagens a serem analisadas. Nota-se que nem todos os *slices* da MRI são significativos para o estudo da AD. Os *slices* de interesse certamente são aqueles que mostram as estruturas que sofrem alteração pela evolução da AD, como é o caso do hipocampo ([DEKEYZER; BOSSCHE, 2017](#)), que em indivíduos afetados apresenta uma perda de volume três vezes superior ao indivíduo normal ao ano ([MILLER-THOMAS *et al.*, 2016](#)).

Para exemplificação, a Figura 4.10 mostra um *slice* no plano axial de uma MRI 3T ponderada em T2, mostrando a identificação do hipocampo e de suas partes, apresentada em (DEKEYZER; BOSSCHE, 2017).

Figura 4.10. Anatomy of the hippocampal formation: 1 = hippocampal head, 2 = hippocampal body, 3 = hippocampal tail, 4 = mesencephalon.

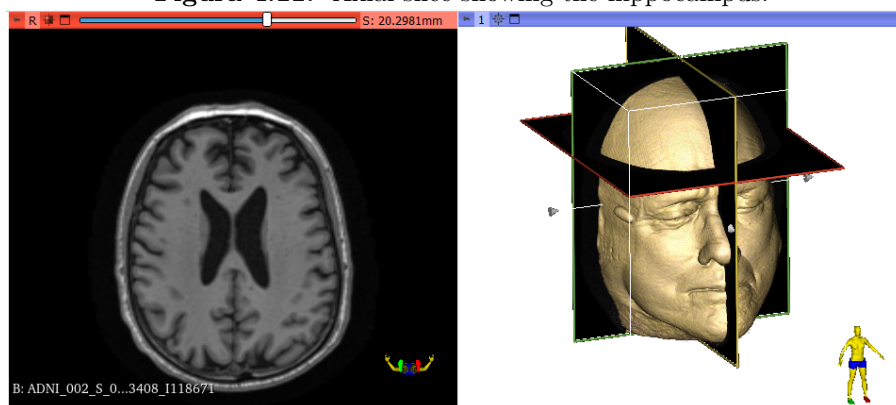


Source: Adapted from (DEKEYZER; BOSSCHE, 2017).

Certamente os *slices* em qualquer dos três planos axial, coronal ou sagital, vão revelar aspectos do hipocampo, tendo em vista que a aquisição de dados foi feita em toda extensão da cabeça do paciente, numa resolução espacial bastante fina. Porém, nesta tese optou-se por escolher *slices* no plano axial e coronal, para a geração das imagens que vão compor a base de dados de análise. Naturalmente, os *slices* no plano sagital não mostram a parte direita e esquerda do hipocampo simultaneamente, e por isso não foram avaliados neste trabalho.

4.4 AVALIAÇÃO NO PLANO AXIAL

A Figura 4.11 mostra um *slice* no plano axial selecionado pelo software *3D Slicer*.

Figura 4.11. Axial slice showing the hippocampus.

Source: Own authorship.

Uma das possibilidades de continuidade deste trabalho seria contar com a contribuição de um médico neurorradiologista com a função de inspecionar cada uma das MRI e escolher o *slice* que melhor represente a região do hipocampo. Um trabalho a ser feito em todos os 1075 pacientes da base *ADNI1:Screening 1.5T*.

No entanto, sendo a premissa desta tese o desenvolvimento autônomo de ferramentas de inteligência computacional para o auxílio ao diagnóstico de AD, optou-se por desenvolver um algoritmo de seleção de *slices* que faça uma varredura em todas as MRI e calcule o *slice* de interesse.

4.4.1 Algoritmo de seleção de *slices* proposto

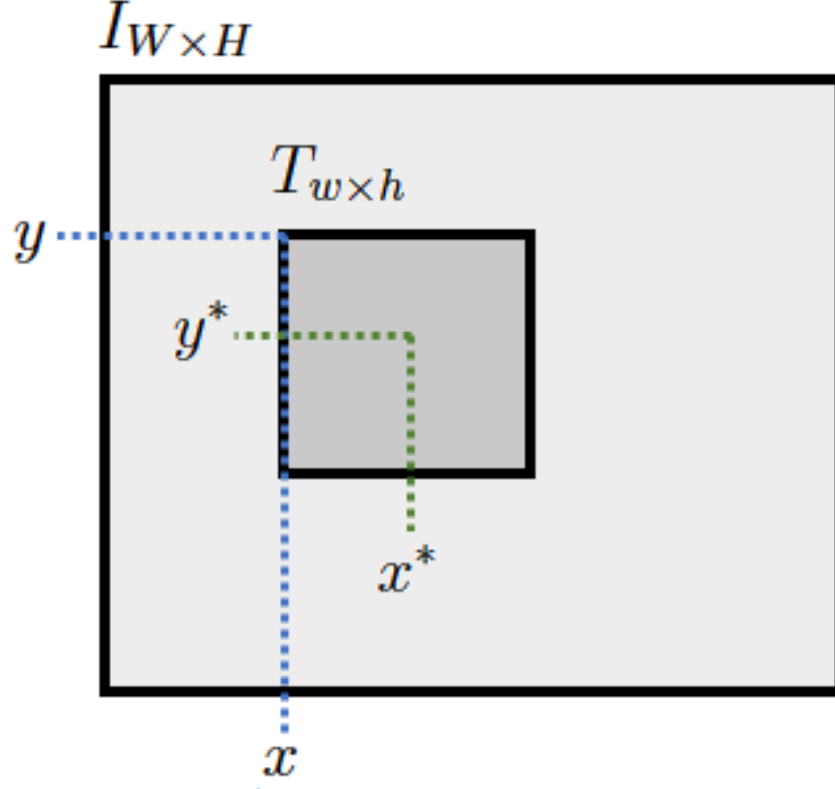
Dado uma imagem I de dimensões $W \times H$, correspondente a um específico *slice* no plano axial, o objetivo é verificar se essa imagem possui semelhanças com uma imagem gabarito T , representativa da anatomia do hipocampo, de dimensões $w \times h$, sendo $w \leq W$ e $h \leq H$, e escolher o *slice* que resultar na melhor similaridade.

4.4.1.1 O algoritmo SMR

Esta tese propõe um algoritmo de seleção de *slices* por similaridade, que foi batizado de **SMR** (do inglês: *similarity matching rule*), baseado na combinação de três funções de casamento de gabaritos. As três funções fazem o deslizamento do gabarito T sobre a imagem I ,

conforme indicado na Figura 4.12, para todo $x \leq W - w + 1$ e $y \leq H - h + 1$, resultando em uma imagem $R(x,y)$, de resultados de similaridade, de dimensão $[W - w + 1 \times H - h + 1]$.

Figura 4.12. Image matching.



Source: Own authorship.

A primeira função é a $R_{diff}(x,y)$, dada pela Equação 4.1, que realiza o somatório normalizado da diferença quadrática de cada ponto do gabarito com a imagem, a partir do ponto (x,y) . O valor mínimo de $R_{diff}(x,y)$ determina o ponto (x,y) de maior similaridade entre a imagem e o gabarito.

$$R_{diff}(x,y) = \frac{\sum_{x^*,y^*} (T(x^*,y^*) - I(x + x^*,y + y^*))^2}{\sqrt{\sum_{x^*,y^*} T(x^*,y^*)^2 \sum_{x^*,y^*} I(x + x^*,y + y^*)^2}} \quad (4.1)$$

A segunda função é a $R_{dot}(x,y)$, dada pela Equação 4.2, que realiza o somatório normalizado do produto de cada ponto do gabarito com a imagem, a partir do ponto (x,y) . O valor máximo de $R_{dot}(x,y)$ determina o ponto (x,y) de maior similaridade entre a imagem e o gabarito.

$$R_{dot}(x,y) = \frac{\sum_{x^*,y^*} (T(x^*,y^*) I(x + x^*,y + y^*))}{\sum_{x^*,y^*} T(x^*,y^*)^2 \sum_{x^*,y^*} I(x + x^*,y + y^*)^2} \quad (4.2)$$

A terceira função é a $R_{cdot}(x,y)$, dada pela Equação 4.3, que realiza o somatório normalizado do produto de cada ponto do gabarito, subtraído da média, com a imagem, subtraída da média, a partir do ponto (x,y) . O valor máximo de $R_{cdot}(x,y)$ determina o ponto (x,y) de maior similaridade entre a imagem e o gabarito.

$$R_{cdot}(x,y) = \frac{\sum_{x^*,y^*} (T^*(x^*,y^*) I^*(x+x^*,y+y^*))}{\sum_{x^*,y^*} T(x^*,y^*)^2 \sum_{x^*,y^*} I(x+x^*,y+y^*)^2} \quad (4.3)$$

Importante notar que $R_{cdot}(x,y)$ realiza o produto dos valores dos *pixels* subtraídos da média, dados pelas Equações 4.4 e 4.5.

$$T^*(x^*,y^*) = T(x^*,y^*) - 1/(w h) \sum_{x^{**},y^{**}} T(x^{**},y^{**}) \quad (4.4)$$

$$I^*(x+x^*,y+y^*) = I(x+x^*,y+y^*) - 1/(w h) \sum_{x^{**},y^{**}} I(x+x^{**},y+y^{**}) \quad (4.5)$$

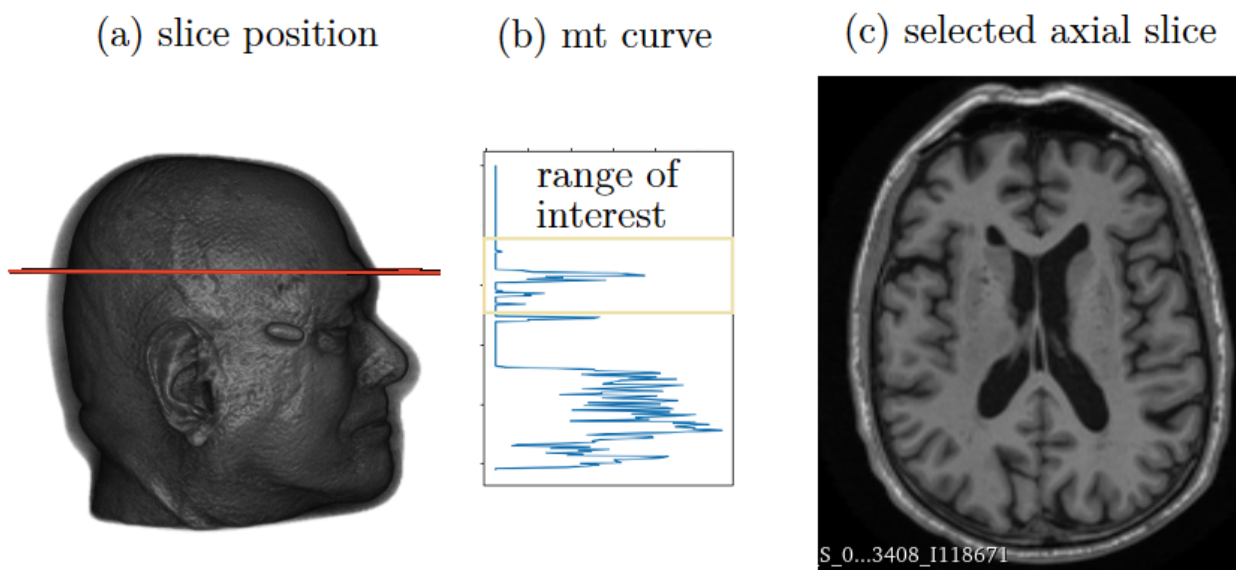
Portanto, a função $R_{cdot}(x,y)$ calcula a covariância normalizada entre a imagem I e o gabarito T .

Por fim, o algoritmo SMR faz a combinação das três funções para cada *slice* analisado, conforme a Equação 4.6, onde th é um limiar de 0.01.

$$mt = \max(R_{dot}) \max(R_{cdot}) (1 - \min(R_{diff}) + th) \quad (4.6)$$

Na Figura 4.13, está evidenciado um exemplo de utilização do algoritmo SMR sobre a MRI do paciente anonimizado 002_S_0295, que faz parte da base de dados *ADNI1:Screening 1.5T*. A linha vermelha em (a) mostra a posição do *slice* selecionado com relação à cabeça do paciente. A faixa de interesse da busca do *slice* está destacada em amarelo sobre a curva mt mostrada em (b). Naturalmente, a posição do *slice* selecionado coincide com o pico da curva mt , dentro da faixa de interesse. Finalmente, em (c) está mostrada a imagem do *slice* selecionado.

Figura 4.13. Example SMR result. The red line in (a) shows the position of the selected slice with respect to the patients head. The range of interest on the mt curve (b) for the slice search is outlined in yellow. The image of the selected slice is shown in (c).



Source: Own authorship.

A partir do algoritmo SMR é possível selecionar um ou mais **slices** da MRI com a intenção de realizar aumento de dados, se necessário. Tendo em vista que a base *ADNI1:Screening 1.5T* possui 1075 MRI e as redes convolucionais necessitam de grandes volumes de dados para treinamento, os *slices* laterais ao *slice* selecionado pelo SMR podem ser também utilizados para o treinamento da rede convolucional.

4.4.1.2 Aplicação do SMR na base de dados

A base *ADNI1:Screening 1.5T* possui 1075 MRI de pacientes diagnosticados nas três classes, CN, MCI e AD. As MRI estão formatadas em NIFTI, com *slices* nos planos axial, coronal e sagital.

Assim, o algoritmo SMR foi utilizado para varrer todos os *slices* do plano axial das 1075 MRI, gerando os resultados *mt*, sendo selecionado o slice correspondente ao maior *mt*, dentro da faixa de interesse.

O gabarito T utilizado está mostrado na Figura 4.14, que possui 70 pixels de largura e 103 pixels de altura. A imagem I tem dimensões 166×256 pixels, correspondente a um específico

slice no plano axial.

Figura 4.14. Template.

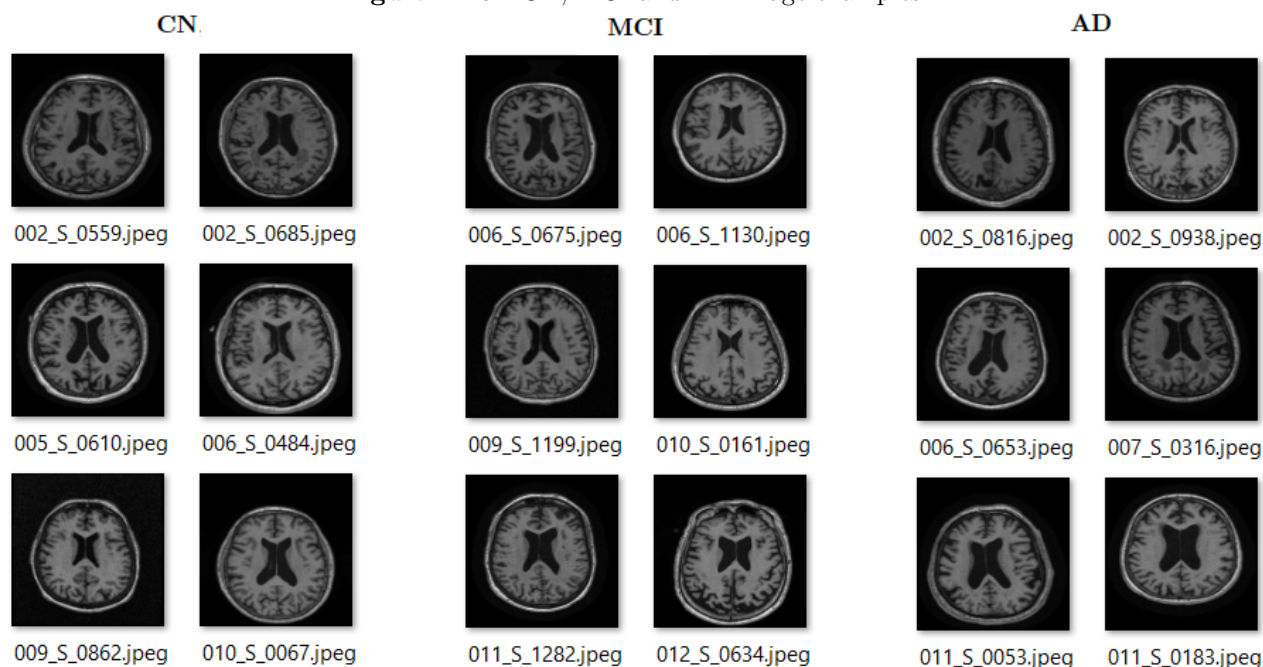


Source: Own authorship.

4.4.2 Banco de imagens resultante

Ao final do processamento de toda a base de dados *ADNI1:Screening 1.5T*, foi possível criar diretórios CN, MCI e AD para separar as imagens de cada classe. A Figura 4.15 mostra um resumo de como ficaram as imagens selecionadas para serem processadas no próximo estágio.

Figura 4.15. CN, MCI and AD image examples.



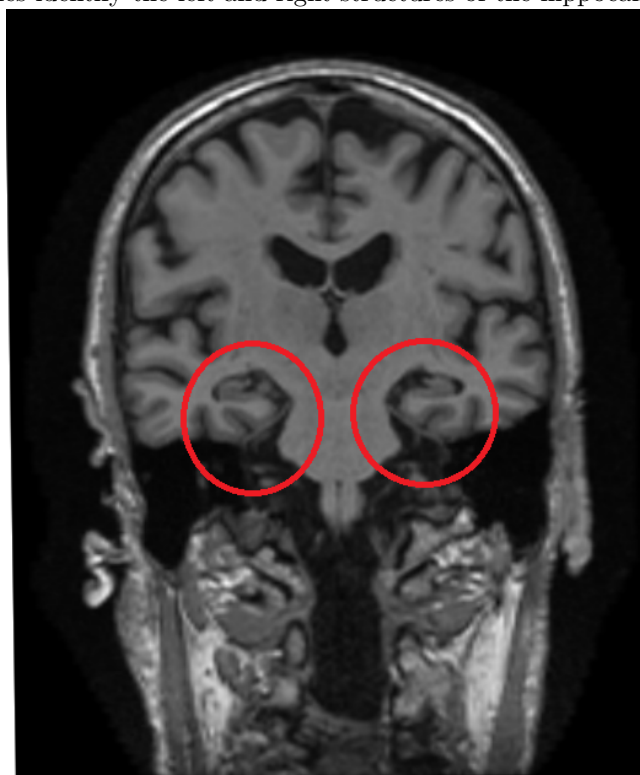
Source: Own authorship.

Todas as imagens foram salvas em formato JPEG com dimensões 176 pixels por 176 pixels.

4.5 AVALIAÇÃO NO PLANO CORONAL

No plano coronal é possível visualizar simultaneamente as duas partes, esquerda e direita, que compõem as estruturas do hipocampo, conforme mostrado na Figura 4.16.

Figura 4.16. Red circles identify the left and right structures of the hippocampus in the coronal plane.

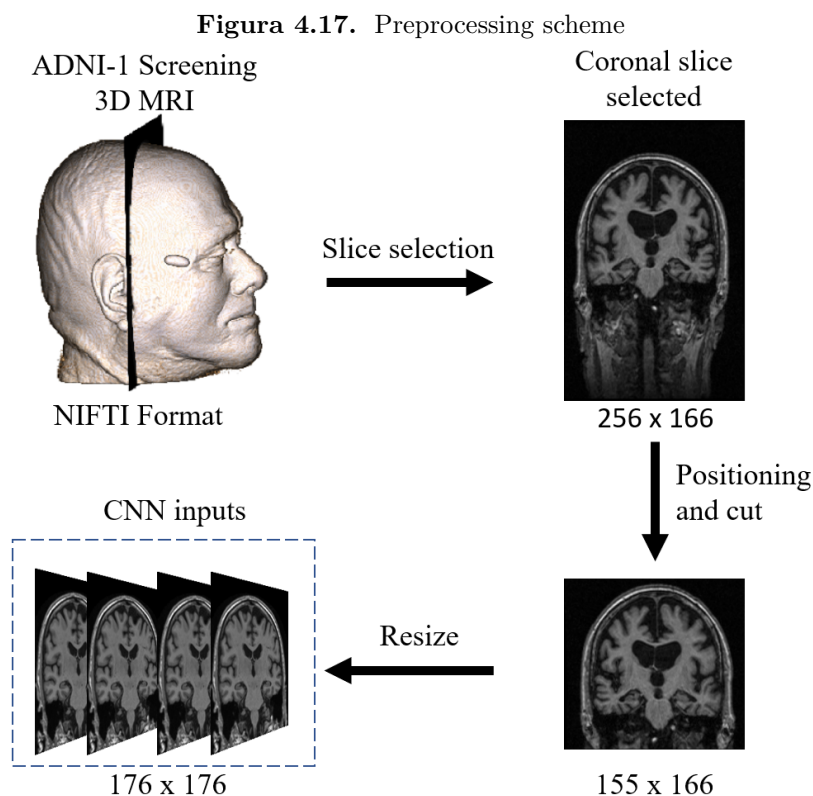


Source: Own authorship.

Nota-se na Figura 4.16 que cada parte esquerda e direita, circundadas na cor vermelho, são estruturas bastante semelhantes e espelhadas em relação ao eixo central do plano coronal. Essa característica foi utilizada para a realização da seleção de *slices*.

4.5.1 Estratégia para construção dos bancos de imagens

No plano coronal, decidiu-se ter uma experiência diferente da do plano axial e foi realizada uma varredura manual sobre as MRIs de toda base de dados *ADNI1:Screening 1.5T* das classes CN e AD. A Figura 4.17 mostra o fluxo de processamento realizado até se chegar às imagens de interesse para a entrada nas redes convolucionais.



Source: Own authorship.

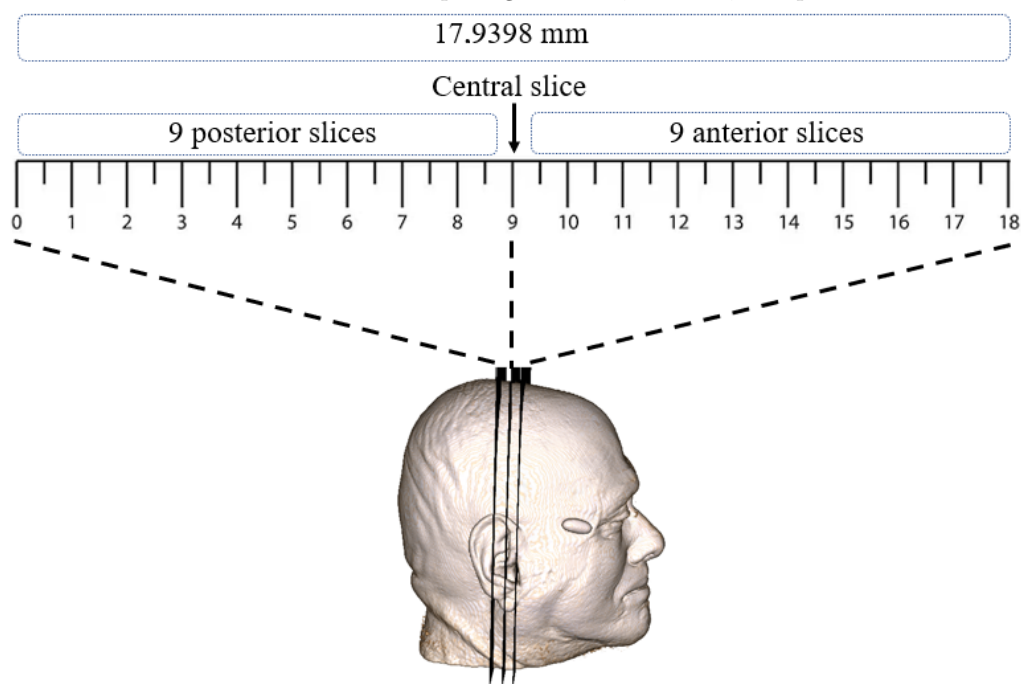
A primeira parte do processamento consistiu em fazer a leitura da MRI 3D da base de dados *ADNI1:Screening 1.5T* e fazer a varredura dos *slices* no plano coronal, na busca de encontrar as estruturas que evidenciam a presença de hipocampo. Nesse ponto, é feito um redimensionamento da imagem para 256 pixels x 166 pixels, tendo em vista que nem todas as MRIs foram geradas nas mesmas dimensões.

O passo seguinte foi o enquadramento da região de interesse e a eliminação de partes da imagem, sobretudo na parte inferior da cabeça, privilegiando a região da presença do cérebro, onde se localizam as estruturas do hipocampo. Assim, a imagem selecionada ficou com as dimensões 155 pixels x 166 pixels.

Finalmente, as imagens foram convertidas para tom de cinza em 8 bits, redimensionadas para adequação para o processamento pelas redes convolucionais em 176 pixels x 176 pixels e salvas em disco. A conversão para 8 bits foi feita por arredondamento, sem perda substancial de qualidade, tendo em vista que a imagem original tem baixa faixa dinâmica. Assim, não foi avaliada a hipótese de usar técnicas de preservação da qualidade da imagem, como *Tone Mapping* (SALIH *et al*, 2012).

Uma opção para resultar em aumento de dados foi gerada, tendo em vista a necessidade das redes convolucionais treinarem com grande volume de dados. A Figura 4.18 mostra a estratégia que foi utilizada. Escolhido o *slice* central por inspeção visual, fica a opção de selecionar até mais outros nove *slices* na posição anterior e na posterior, totalizando 19 *slices* por paciente, numa única MRI, cobrindo uma área total de até 17,9398 mm, tendo em vista que os *slices* são espaçados 0,9442mm um do outro.

Figura 4.18. Coronal slices with 0.9442 mm spacing. Central, anterior, and posterior slices are highlighted.



Source: Own authorship.

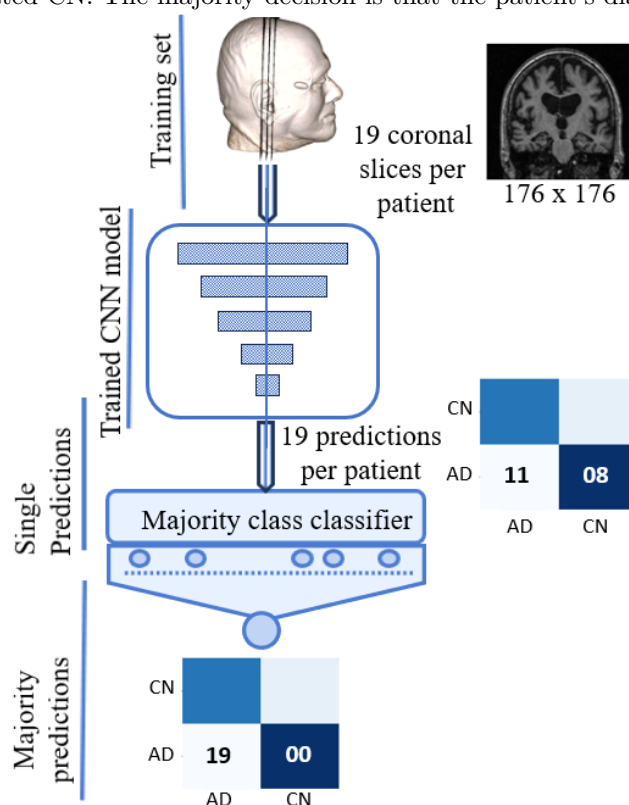
4.5.2 Avaliação por múltiplas instâncias

O uso de múltiplas instâncias por paciente é justificado pelo aumento necessário de dados para a etapa de treinamento da rede e também para viabilizar a decisão do diagnóstico por maioria. Conforme mostrado na Figura 4.6, o hipocampo possui a maior extensão na direção medida de frente para trás da cabeça do indivíduo, podendo chegar a 39 mm, no exemplo medido. Portanto, para as MRIs do ADNI1: Screening 1.5T, cerca de até 40 *slices* coronais podem ser selecionados para cobrir toda a extensão do hipocampo. Nesse trabalho, foi escolhido um *slice* central, que manualmente foi escolhido na região central do hipocampo, e outros nove *slices* na direção anterior e na posterior, totalizando 19 *slices* por paciente. A estratégia

utilizada neste trabalho corresponde a utilização de cerca de 50% da extensão do hipocampo. Os outros 50% foram deixados como tolerância a estimativa inicial do *slice* central e as variações anatômicas inerentes à diversidade de pacientes.

A Figura 4.19 mostra um exemplo de um paciente hipotético onde foram utilizadas 19 instâncias. Neste exemplo, 11 instâncias foram avaliadas individualmente como AD e 08 instâncias indicaram CN. A decisão por maioria dá o resultado final de AD para este paciente.

Figura 4.19. Multiple instance decision scheme. In this case, a patient was evaluated and 11 instances indicated AD and 08 instances indicated CN. The majority decision is that the patient's diagnosis is AD.



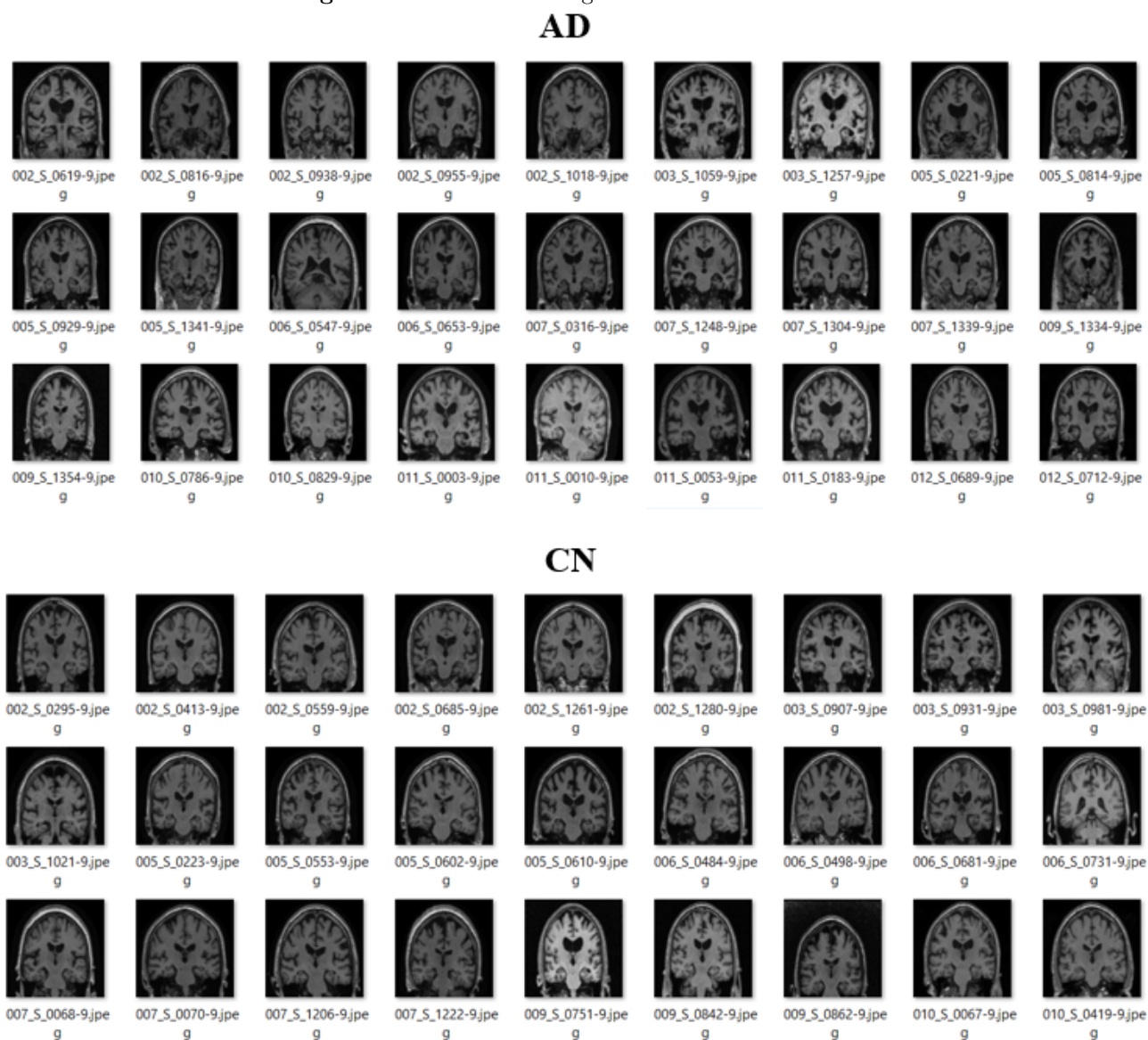
Source: Own authorship.

Além disso, é importante ressaltar que a decisão por maioria ajuda a tornar a escolha do *slice* central mais robusta, considerando que os *slices* anteriores e posteriores expandem a região espacial de análise. Outro fator é a maior tolerância a uma possível falha na inferência da rede, que pode variar o diagnóstico em *slices* individuais de um mesmo paciente.

4.5.3 Bancos de imagens resultantes

Foi gerado um banco de imagens com *slices* centrais para pacientes com CN e com AD. Uma amostra do banco de imagens está apresentada na A Figura 4.20.

Figura 4.20. Coronal image dataset of central slices.

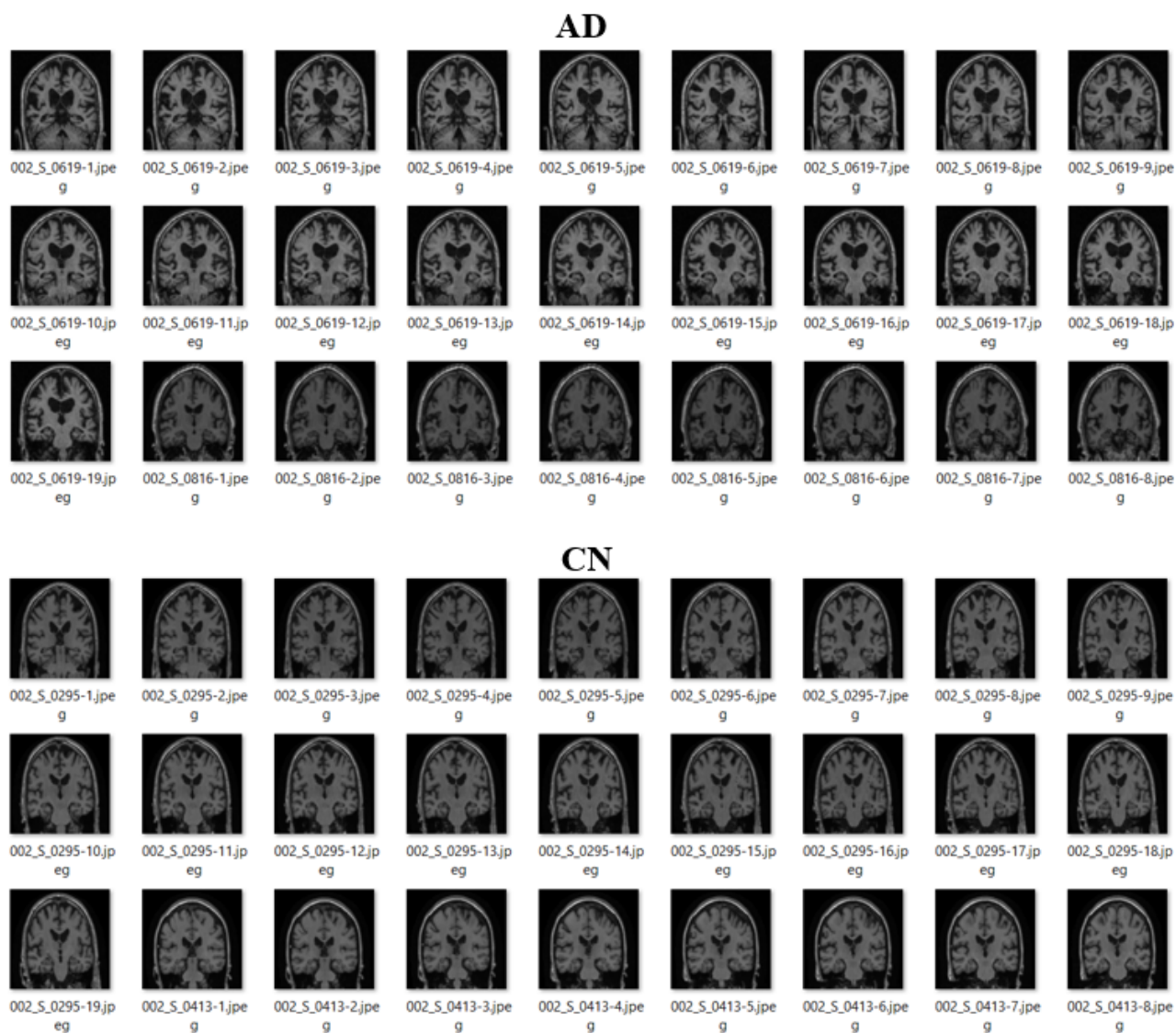


Source: Own authorship.

Nem todas as MRIs avaliadas puderam ser aproveitadas por dados faltantes, por erro na leitura do arquivo ou por inconsistência dos dados. Assim, da base de dados original *ADNI1:Screening 1.5T*, que possui 307 pacientes em CN, 525 pacientes em MCI e 243 pacientes em AD, foram geradas imagens de 226 pacientes em CN e 173 pacientes em AD. Naturalmente,

uma qualidade insuficiente para treinamento de redes convolucionais, o que reforça a necessidade de se fazer aumento de dados. Nesse sentido, foi pensada a opção de aumentar os dados com uso de *slices* anteriores e posteriores. A Figura 4.21 apresenta uma amostra do banco de imagens com o *slice* central, nove *slices* anteriores e nove *slices* posteriores.

Figura 4.21. Coronal dataset with a center slice, 9 anterior slices, and 9 posterior slices. The slice ending in -10 is the central slice chosen.

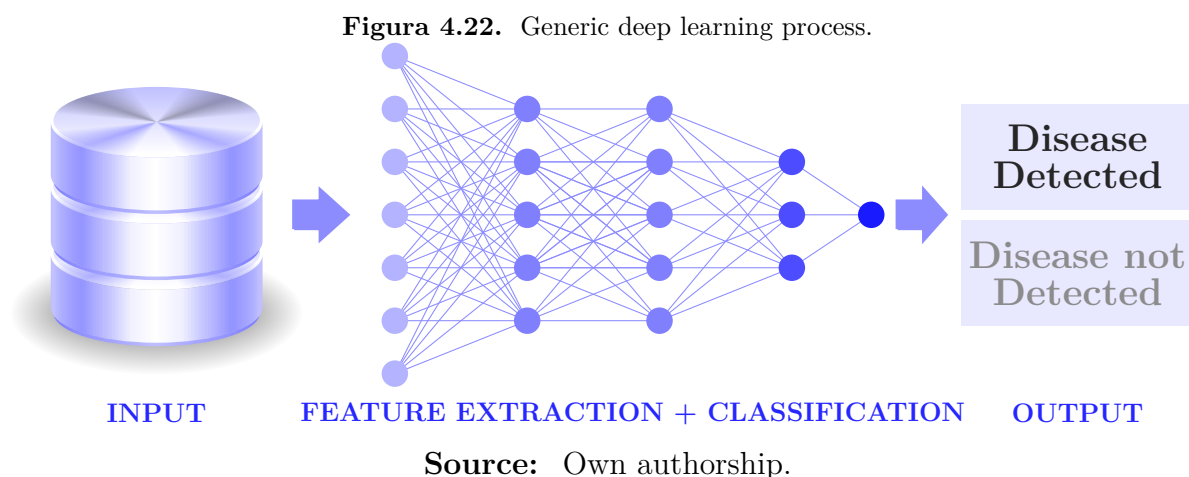


Source: Own authorship.

Com a estratégia aqui proposta para aumento do banco de imagens, pode-se chegar até 4.294 imagens de CN e 3.287 imagens de AD, o que já viabiliza o treinamento de redes convolucionais.

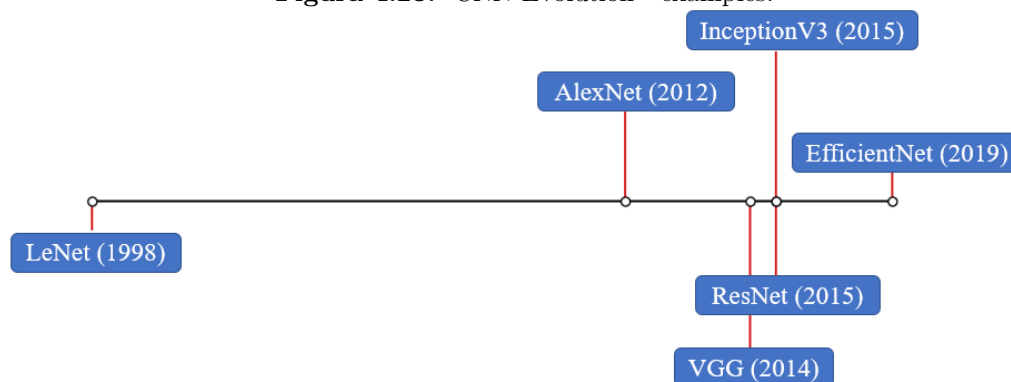
4.6 MODELOS DE DL UTILIZADOS

Os modelos de DL são compostos por múltiplas camadas de processamento para aprender representações de dados com múltiplos níveis de abstração. As redes DL do tipo *convolutional neural networks* (CNN) trouxeram avanços no processamento de imagens, vídeo, fala e áudio, ao passo que as redes recorrentes brilharam em dados sequenciais como o texto e a fala (LECUN *et al*, 2015). A Figura 4.22 mostra um esquema geral do processamento dos modelos de DL. Há de se observar que a extração de atributos foi incorporada diretamente ao modelo e é realizada nas primeiras camadas de processamento.



As arquiteturas CNN típicas empilham algumas camadas convolucionais, depois uma camada de *pooling*, depois outras poucas camadas convolucionais, depois outra camada de *pooling* e assim por diante. Ao longo dos anos, variantes dessa arquitetura fundamental foram desenvolvidas, levando a avanços surpreendentes no campo (YAMASHITA *et al*, 2018). A Figura 4.23 mostra alguns exemplos de CNNs que marcaram a evolução da área.

Figura 4.23. CNN Evolution - examples.



Source: Own authorship.

4.6.1 Modelos clássicos de CNNs

A LeNet ([LECUN *et al.*, 1998](#)) foi originalmente desenvolvida para reconhecimento de dígitos manuscritos e é formada por duas camadas convolucionais, seguidas por *pooling*, e duas camadas totalmente conectadas.

A AlexNet ([KRIZHEVSKY *et al.*, 2012](#)) foi a primeira vencedora do desafio "ImageNet". Possui oito camadas no total (cinco camadas convolucionais e outras três camadas totalmente conectadas). Introduziu a função de ativação *rectified linear unit* (ReLU).

A VGG ([SIMONYAN; ZISSERMAN, 2014](#)) foi proposta pela equipe do *Visual Geometry Group* (VGG), da Universidade de Oxford. A variante VGG-16 possui 13 camadas convolucionais e três totalmente conectadas. Também foi proposta a VGG-19, que é uma variante ainda mais profunda, com 16 camadas convolucionais e três totalmente conectadas.

A ResNet ([HE *et al.*, 2015](#)) incorporou a ideia de "blocos residuais", que são conectados uns aos outros por meio de conexões salto, e conseguiu ir ainda mais profundo, sem comprometer o poder de generalização do modelo. A ResNet-50 é um exemplo dessa estratégia.

A Inception, em vez de empilhar camadas convolucionais, empilha módulos ou blocos, dentro dos quais estão as camadas convolucionais. A Inception V2 e V3 ([SZEGEDY *et al.*, 2015](#)) passa a utilizar *batch normalisation*.

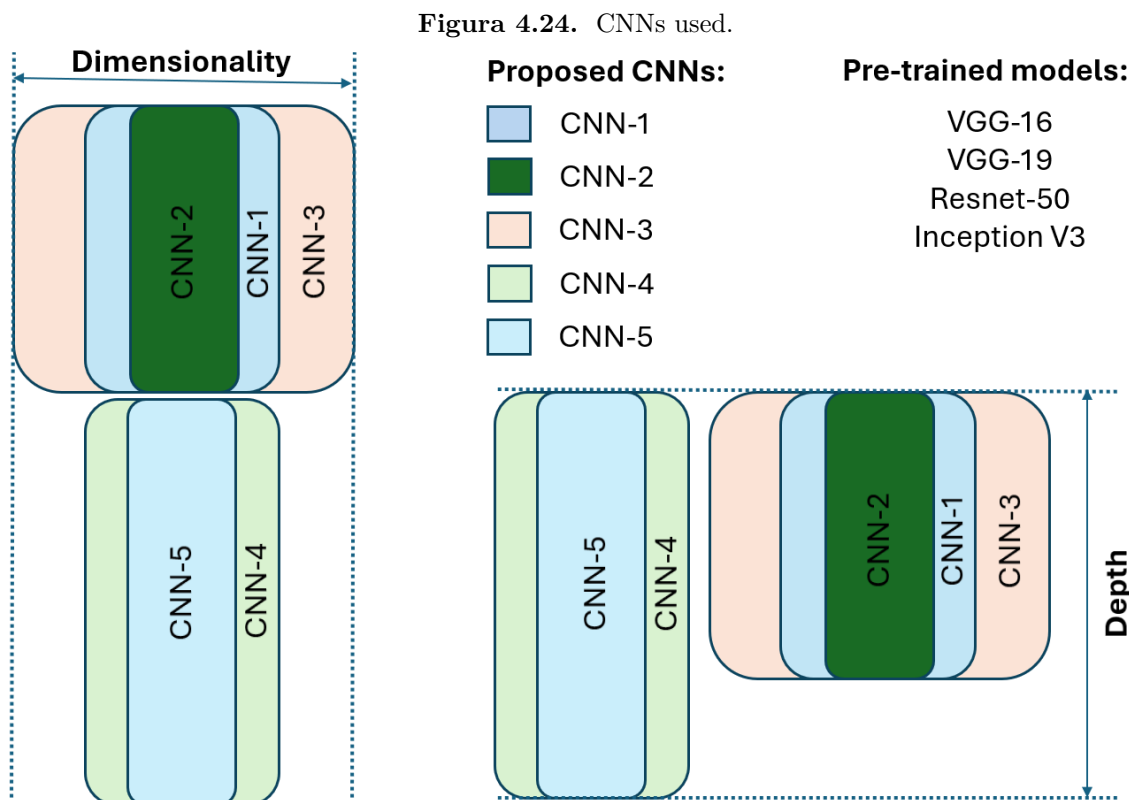
A EfficientNet ([LIAN; LE, 2019](#)) usa uma estratégia para dimensionar uniformemente a largura, a profundidade e a resolução da rede.

Neste trabalho, optou-se por utilizar as redes VGG-16, VGG-19, ResNet-50 e InceptionV3 para avaliação dos resultados a partir dos bancos de imagem montados, por serem redes frequentemente utilizadas no problema da classificação de imagens.

4.6.2 Modelos propostos de CNNs

Também foi feita a opção por propor cinco modelos de CNNs construídas a partir das camadas estruturais: *convolutional*, *batch normalization*, *max pooling*, *dropout*, *flatten*, *dense*, que foram chamados de CNN-1, CNN-2, CNN-3, CNN-4 e CNN-5.

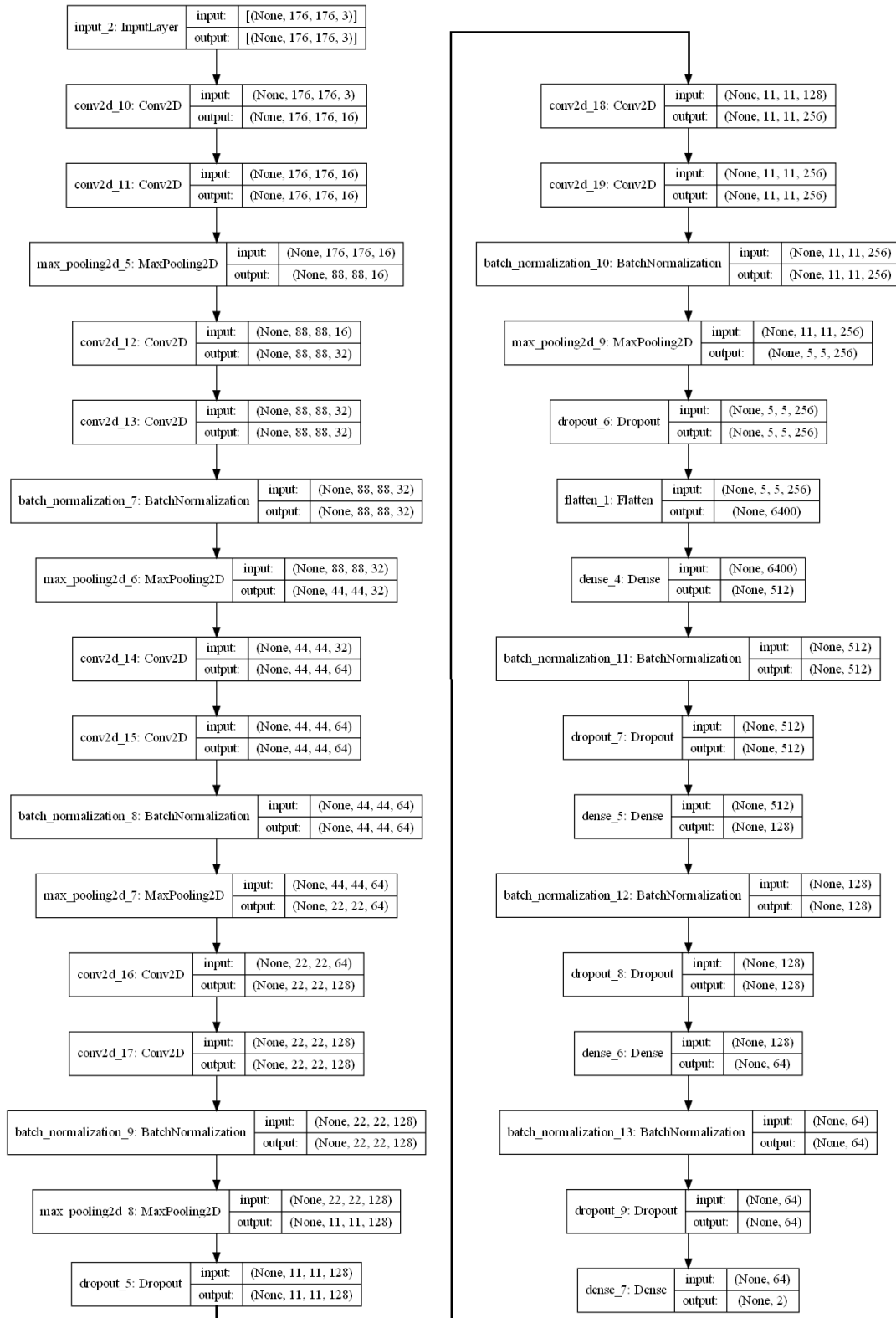
A Figura 4.24 mostra uma visão geral das CNNs utilizadas neste trabalho. Os modelos são detalhados a seguir.



Source: Own authorship.

O modelo CNN-1 está mostrado na Figura 4.25.

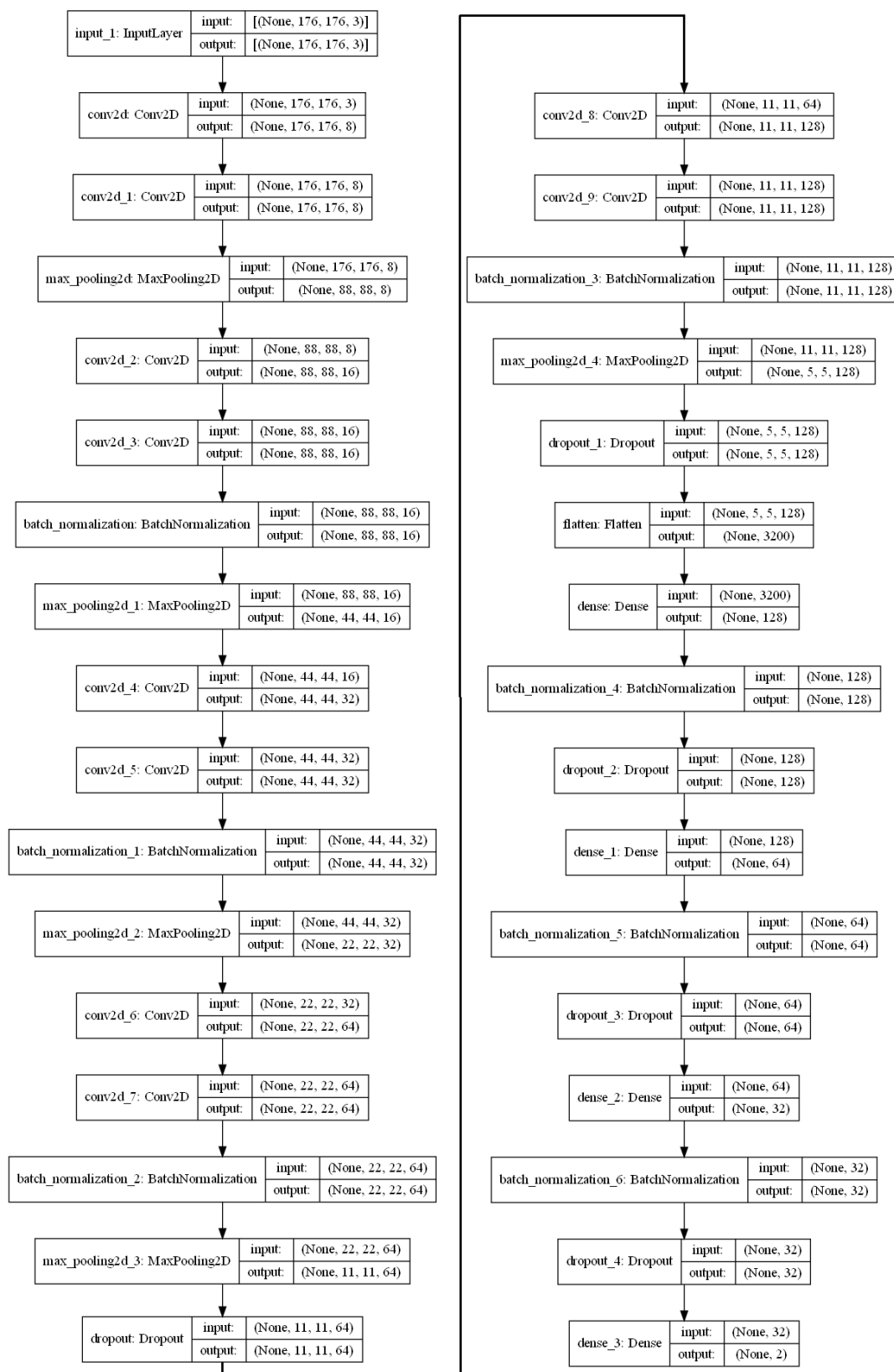
Figura 4.25. CNN 1 model - total parameters: 4,534,866. Initial model.



Source: Own authorship.

O modelo CNN-2 está mostrado na Figura 4.26.

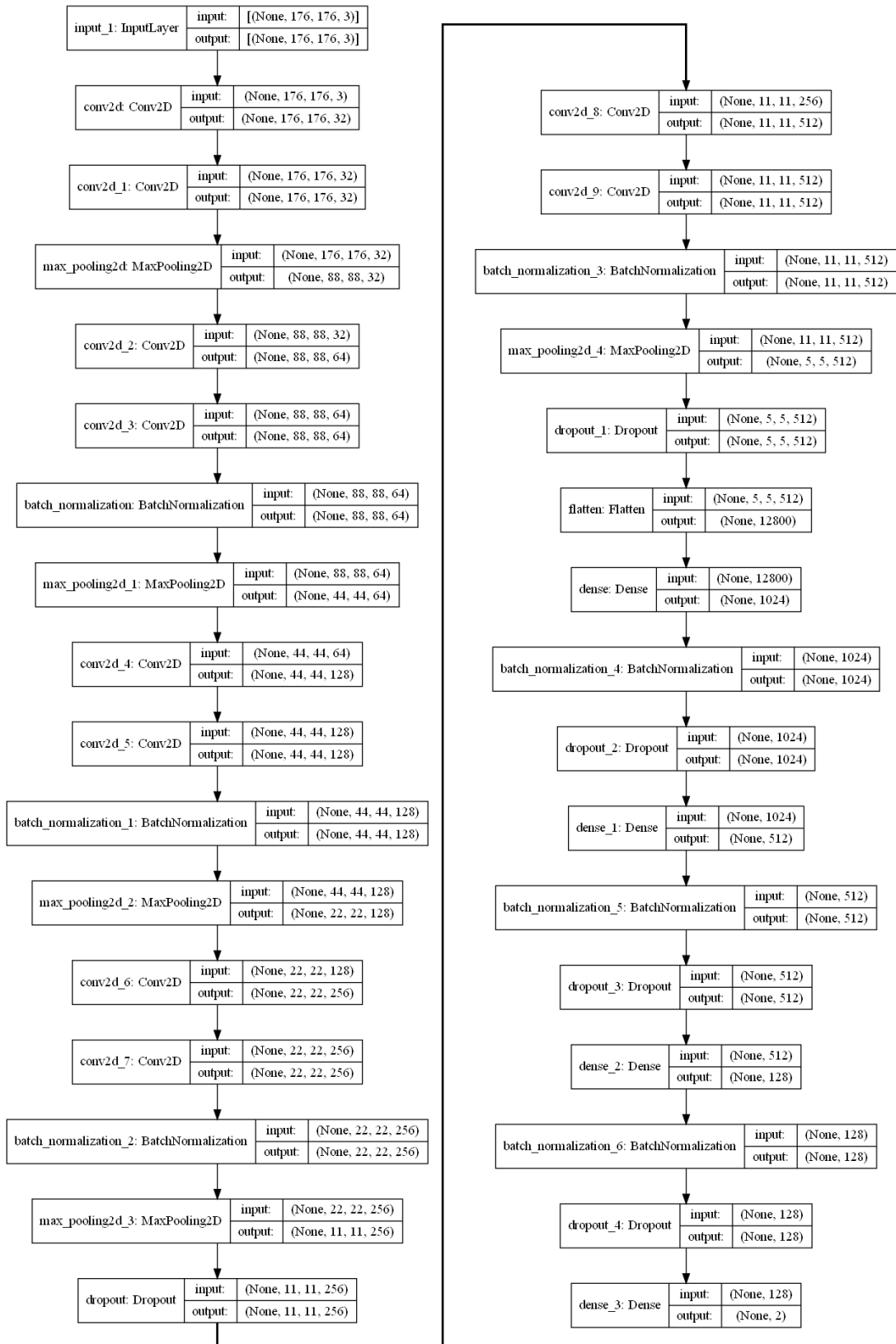
Figura 4.26. CNN 2 model - total parameters: 717,034. Reduced dimensionality with respect to CNN-1.



Source: Own authorship.

O modelo CNN-3 está mostrado na Figura 4.27.

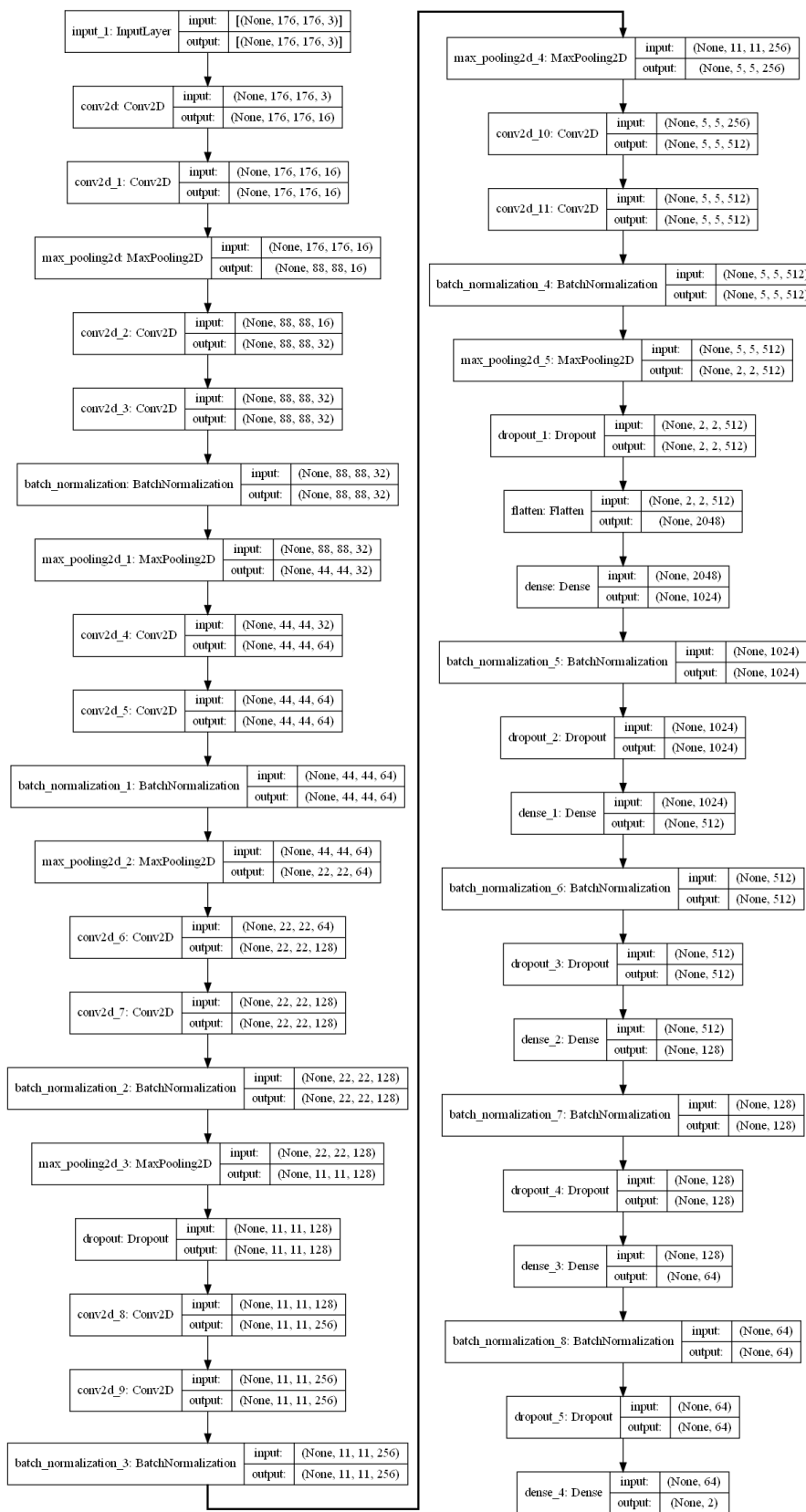
Figura 4.27. CNN 3 model - total parameters: 18,421,666. Increased dimensionality with respect to CNN-1.



Source: Own authorship.

O modelo CNN-4 está mostrado na Figura 4.28.

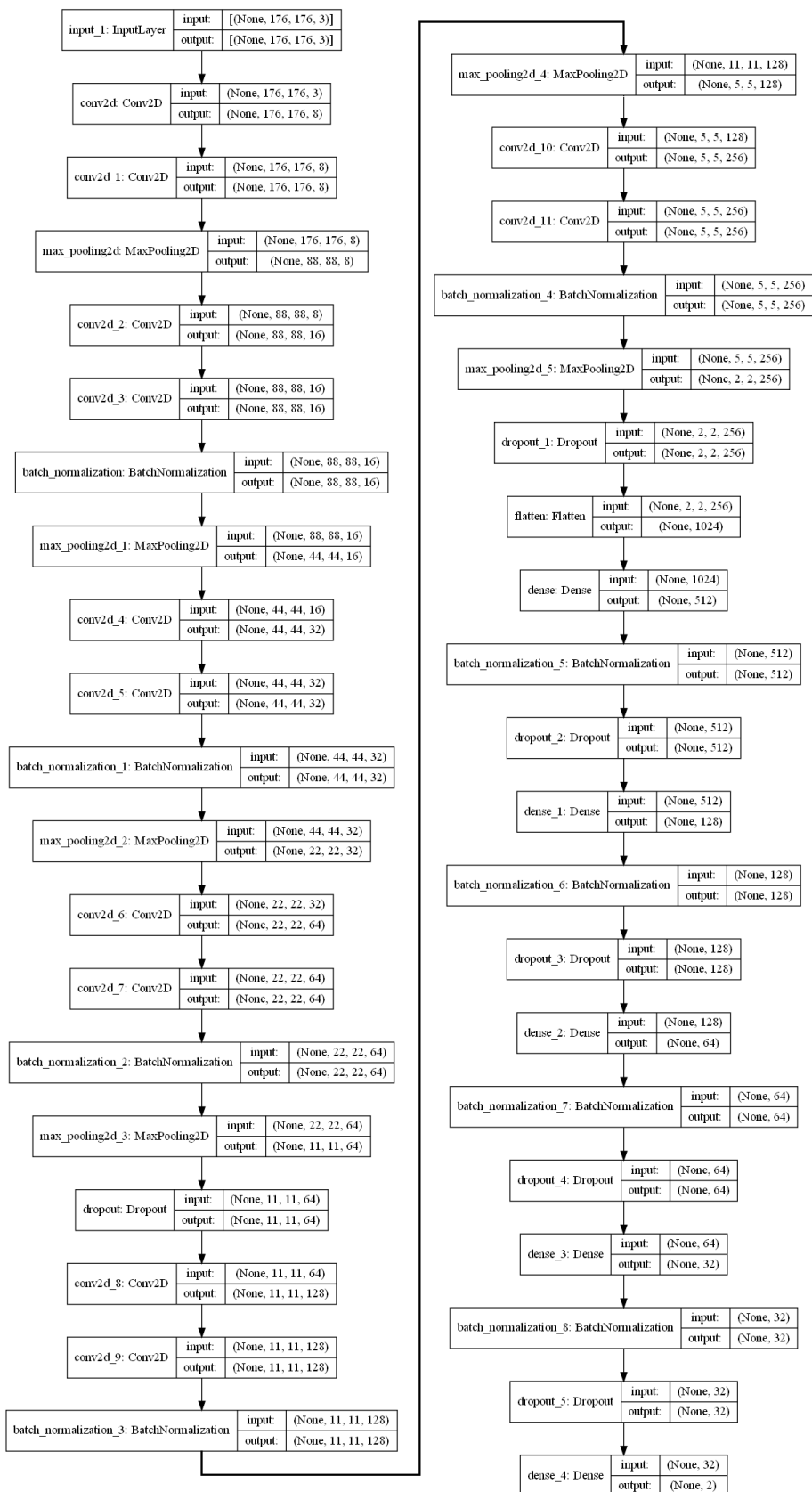
Figura 4.28. CNN 4 model - total parameters: 7,426,642. Increased depth compared to CNN-1.



Source: Own authorship.

O modelo CNN-5 está mostrado na Figura 4.29.

Figura 4.29. CNN 5 model - total parameters: 1,786,090. Reduced dimensionality with respect to CNN-4.



Source: Own authorship.

Como visto na Figura 4.25, a CNN-1 possui 10 camadas *convolutional*, cinco camadas *max pooling*, sete camadas *batch normalization*, cinco camadas *dropout*, uma camada *flatten*, e quatro camadas *dense*. A rede possui um total de 4.534.866 parâmetros, sendo 4.532.498 de parâmetros treináveis e 2.368 de parâmetros não treináveis.

A CNN-2 possui o mesmo número e as mesmas camadas da CNN-1. A diferença está na dimensionalidade de cada camada que foi reduzida, como pode ser visto na Figura 4.26. Assim, a rede ficou com um total de 717.034 parâmetros. Desse total, 716.106 são parâmetros treináveis e 928 são parâmetros não treináveis.

Já a CNN-3, que também possui o mesmo número e as mesmas camadas da CNN-1, teve a dimensionalidade das camadas aumentada, como pode ser visto na Figura 4.27. A rede resultou em um total de 18.421.666 parâmetros. Desse total, 18.416.418 são parâmetros treináveis e 5.248 são parâmetros não treináveis.

Na CNN-4 houve o aumento do número de camadas com relação a CNN-1, tendo sido acrescentadas duas camadas *convolutional*, uma camada *max pooling*, duas camadas *batch normalization*, duas camadas *dropout* e uma camada *dense*, como pode ser visto na Figura 4.28. A dimensionalidade das camadas é equivalente à da CNN-1, tendo resultado em uma rede com um total de 7.426.642 parâmetros. Desse total, 7.421.202 são parâmetros treináveis e 5.440 são parâmetros não treináveis.

Finalmente, a CNN-5 manteve o mesmo número e as mesmas camadas da CNN-4, porém com a dimensionalidade reduzida, como pode ser visto na Figura 4.29. A rede resultou em um total de 1.786.090 parâmetros. Desse total, 1.783.626 são parâmetros treináveis e 2.464 são parâmetros não treináveis.

Este capítulo fez a apresentação da base de ressonância magnética primária utilizada, do tratamento para a montagem dos bancos de dados adequados para a avaliação do diagnóstico da AD e das propostas tecnológicas desenvolvidas, com destaque para a proposta de aumento de dados realizada. No próximo capítulo, serão apresentados os resultados deste trabalho, a comparação com resultados obtidos por outros pesquisadores e a discussão dos resultados, tanto para dados clínicos como para imagem.

RESULTADOS E DISCUSSÃO

Este capítulo apresenta os resultados obtidos pelas análises clínicas e os resultados obtidos pela avaliação de imagens de ressonância magnética. Ao final, é feita uma discussão sobre os resultados e uma comparação com resultados atingidos por outros pesquisadores.

5.1 RESULTADOS DE ANÁLISE POR DADOS CLÍNICOS

Os resultados com dados clínicos foram produzidos para viabilizar a comparação do desempenho dos classificadores nas condições das bases de dados sem a seleção de atributos e com a seleção de atributos.

Os procedimentos práticos foram realizados em linguagem python versão 3.7.3, com as bibliotecas numpy versão 1.19.5, scikit-learn versão 0.24.1, e utilizado um processador Intel(R) Core(TM) i7-7500U CPU @ 2.70 GHz 2.90 GHz, com 16 GB de RAM.

O algoritmo de seleção de atributos proposto (FMFS) foi aplicado a cada uma das bases de dados, resultando em conjuntos reduzidos de atributos que foram submetidos a dois algoritmos de classificação: *k-nearest neighbors* (kNN) e *support vector machines* (SVM), e avaliados em termos das métricas: acurácia balanceada, precisão, sensibilidade, especificidade, score f1 e a *multi-class area under the curve* (MAUC). Ao final, foi possível comparar o desempenho da classificação entre o conjunto de dados completo e o conjunto de dados reduzido, evidenciando o efeito do algoritmo de seleção de atributos, assim como foi possível compará-lo com os resultados obtidos em outros estudos.

5.1.1 Seleção de atributos para a base AIBL

A base de dados clínicos AIBL, descrita no Capítulo 3, que possui originalmente 18 atributos, 785 instâncias, três classes, e é referenciada neste trabalho como AIBL-18, foi submetida ao algoritmo de seleção de atributos proposto (FMFS), com otimização para os modelos de classificação kNN e SVM. O resultado está explicitado na Tabela 5.1.

Tabela 5.1. Reduced AIBL dataset: selected features

Model	Dataset	Features
KNN	AIBL-6	CDGLOBAL BAT126 MMSCORE LIMMTOTAL LDELTOTAL APGEN1
SVM	AIBL-4	CDGLOBAL MMSCORE LIMMTOTAL LDELTOTAL

Source: Own authorship.

Assim, obteve-se duas novas bases de dados, AIBL-6 e AIBL-4, com seis e quatro atributos, respectivamente, que serão avaliadas nas seções seguintes.

5.1.2 Seleção de atributos para a base ADNI

A base de dados clínicos ADNI, descrita no Capítulo 3, que possui originalmente 21 atributos, 1033 instâncias, três classes, e é referenciada neste trabalho como ADNI-21, foi submetida ao algoritmo de seleção de atributos proposto (FMFS), com otimização para os modelos de classificação kNN e SVM. O resultado está explicitado na Tabela 5.2.

Tabela 5.2. Reduced ADNI dataset: selected features

Model	Dataset	Features
KNN	ADNI-3	CDRSB ADAS11 FAQ
SVM	ADNI-11	Ventricles Entorhinal CDRSB MMSE RAVLT_learning RAVLT_immediate RAVLT_perc_forgetting_bl ADAS11 ADAS13 FAQ APOE4

Source: Own authorship.

Assim, obteve-se duas novas bases de dados, ADNI-3 e ADNI-11, com três e 11 atributos, respectivamente, que serão avaliadas nas seções seguintes.

5.1.3 Seleção de atributos para a base OASIS

A base de dados clínicos OASIS, descrita no Capítulo 3, que possui originalmente sete atributos, 217 instâncias, três classes, e é referenciada neste trabalho como OASIS-7, foi submetida ao algoritmo de seleção de atributos proposto (FMFS), com otimização para os modelos de classificação kNN e SVM. O resultado está explicitado na Tabela 5.3.

Tabela 5.3. Reduced OASIS dataset: selected features

Model	Dataset	Features
KNN	OASIS-5	Age Educ SES MMSE nWBV
SVM	OASIS-5	Age Educ SES MMSE nWBV

Source: Own authorship.

O procedimento resultou em uma única nova base de dados OASIS-5, com cinco atributos iguais para os dois modelos de classificação kNN e SVM, que será avaliada nas seções seguintes.

5.1.4 Avaliação do desempenho para três classes

Foi avaliado o desempenho da classificação para todas as bases de dados e obtidos os resultados para as bases de dados completas e reduzidas. Nessa avaliação, a classificação foi feita para as classes CN, MCI e AD, no caso das bases AIBL e ADNI, e para as classes CDR 0, CDR 0.5 e CDR 12, para a base de dados OASIS.

5.1.4.1 Avaliação ADNI

A Tabela 5.4 resume os resultados de desempenho obtidos pela avaliação das três classes, em cada conjunto de dados ADNI, para cada modelo de classificação. Para a métrica B_{ca} , foi avaliado o intervalo de confiança para o nível de confiança de 95%. Também foi avaliada a significância estatística dos resultados do B_{ca} entre conjuntos de dados completos e reduzidos.

Nessas condições, os tempos de treinamento sobre as bases ADNI-3 e ADNI-11 ficaram reduzidos em 55% e 7%, respectivamente, comparados com as bases completas.

Tabela 5.4. ADNI dataset results - three classes

Model	Dataset	$B_{ca} \pm IC$	A_{cy}	P_{rn}	S_{ny}	S_{py}	F1	MAUC
KNN	ADNI-21	81,66 \pm 3,39	76,29	79,34	76,89	87,23	77,41	0,878
	ADNI-3	90,06* \pm 2,26	89,45	90,32	86,52	93,60	87,97	0,960
SVM	ADNI-21	92,60 \pm 2,06	91,97	92,52	90,09	95,11	91,07	0,969
	ADNI-11	92,61 \pm 2,29	91,77	92,34	90,20	95,02	90,90	0,969

IC - confidence interval for 95% confidence level

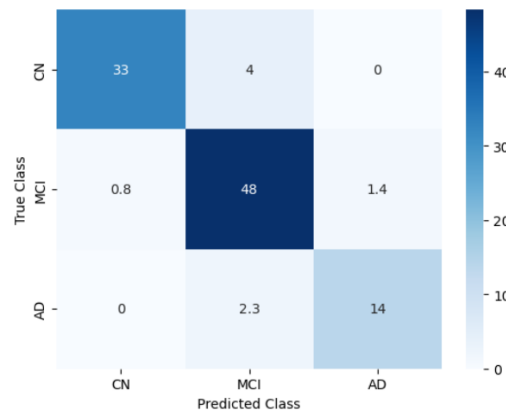
(*) Statistically significant result ($p < 0,05$)

Except for MAUC, every numerical result is expressed as a percentage.

Source: Own authorship.

Conforme verificado na Tabela 5.4, a base de dados ADNI-3 apresentou 10,3% de aumento da métrica B_{ca} com relação ao desempenho avaliado com a base ADNI-21, utilizando o classificador kNN. Esse resultado se mostrou estatisticamente significativo, correspondente a um valor de $p < 0,05$.

Já a base de dados ADNI-11, avaliada pelo classificador SVM, não resultou em desempenho estatisticamente superior quando comparado com o desempenho para a base ADNI-21, mas apresentou o maior desempenho global, tendo atingido a métrica B_{ca} de 92,61%. A Figura 5.1 mostra a matriz de confusão para esse caso.

Figura 5.1. ADNI-11 avaliada with the SVM classifier - average confusion matrix.

Source: Own authorship.

5.1.4.2 Avaliação AIBL

A Tabela 5.5 resume os resultados de desempenho obtidos pela avaliação das três classes, em cada conjunto de dados AIBL, para cada modelo de classificação. Para a métrica B_{ca} , foi avaliado o intervalo de confiança para o nível de confiança de 95%. Também foi avaliada a

significância estatística dos resultados do B_{ca} entre conjuntos de dados completos e reduzidos.

Nessas condições, os tempos de treinamento sobre as bases AIBL-6 e AIBL-4 ficaram reduzidos em 18% e 90%, respectivamente, comparados com as bases completas.

Tabela 5.5. AIBL dataset results - three classes

Model	Dataset	$B_{ca} \pm IC$	A_{cy}	P_{rn}	S_{ny}	S_{py}	F1	MAUC
KNN	AIBL-18	74,15 \pm 2,08	82,28	73,85	63,61	84,69	66,50	0,912
	AIBL-6	90,44* \pm 3,89	91,32	87,37	86,08	94,81	85,81	0,950
SVM	AIBL-18	91,96 \pm 2,25	91,72	86,97	88,32	95,61	87,09	0,967
	AIBL-4	92,52 \pm 2,50	92,22	88,25	89,06	95,98	87,74	0,959

IC - confidence interval for 95% confidence level

(*) Statistically significant result ($p < 0,05$)

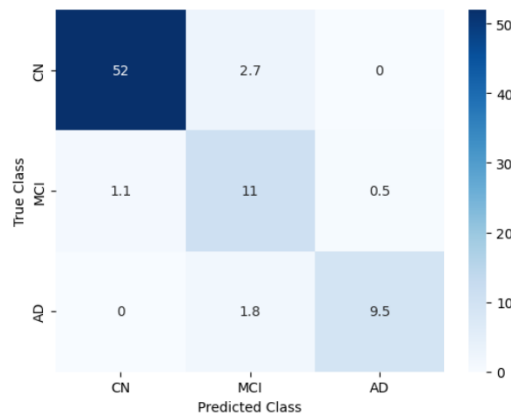
Except for MAUC, every numerical result is expressed as a percentage.

Source: Own authorship.

Conforme verificado na Tabela 5.5, a base de dados AIBL-6 apresentou 22% de aumento da métrica B_{ca} em relação ao desempenho avaliado com a base AIBL-18, utilizando o classificador kNN. Esse resultado se mostrou estatisticamente significativo, correspondente a um valor de $p < 0,05$.

A base de dados AIBL-4, avaliada pelo classificador SVM, apresentou o maior desempenho global e a Figura 5.2 mostra a matriz de confusão para esse caso.

Figura 5.2. AIBL-4 avaliada with the SVM classifier - average confusion matrix.



Source: Own authorship.

Também a base de dados AIBL-4, avaliada pelo classificador SVM, não resultou em desempenho estatisticamente superior quando comparado com o desempenho para a base AIBL-18, tendo atingido a métrica B_{ca} de 92,52%, cerca de 0,9% acima do desempenho avaliado com a base AIBL-18.

5.1.4.3 Avaliação OASIS

A Tabela 5.6 resume os resultados de desempenho obtidos pela avaliação das três classes, em cada conjunto de dados OASIS, para cada modelo de classificação. Para a métrica B_{ca} , foi avaliado o intervalo de confiança para o nível de confiança de 95%.

Tabela 5.6. OASIS dataset results - three classes

Model	Dataset	$B_{ca} \pm IC$	A_{cy}	P_{rn}	S_{ny}	S_{py}	F1	MAUC
SVM	OASIS-7	72.13 \pm 4.90	72.29	66.13	61.95	82.31	59.09	0.858
	OASIS-5	72.66 \pm 4.46	73.64	70.85	61.75	83.57	58.33	0.871
KNN	OASIS-7	67.71 \pm 3.39	69.94	62.79	55.10	80.32	53.51	0.800
	OASIS-5	68.17 \pm 6.29	71.36	64.80	54.53	81.82	51.40	0.804

IC - confidence interval for 95% confidence level

Except for MAUC, every numerical result is expressed as a percentage.

Source: Own authorship.

A base de dados OASIS-5, avaliada pelo classificador SVM, apresentou o maior desempenho global e a Figura 5.3 mostra a matriz de confusão para esse caso.

Figura 5.3. OASIS-5 avaliada com o classificador SVM - matriz de confusão média.



Source: Own authorship.

5.1.5 Avaliação do desempenho para duas classes

Neste tópico, as bases de dados ADNI-3, ADNI-11, AIBL-4 e AIBL-6, foram avaliadas pelos classificadores SVM e kNN, porém num arranjo de duas classes resultantes das combinações: NC v/s AD, MCI v/s AD e NC v/s MCI. Da mesma forma, a base de dados OASIS-5 foi avaliada pelos classificadores SVM e kNN com a seguinte combinação de classes: C00 v/s C12, C05 v/s C12 e C00 v/s C05. Sendo que C00 corresponde à classe CDR 0, C05 corresponde à

classe CRD 0.5 e C12 corresponde à junção das instâncias das classes CDR 1 e CDR 2.

5.1.5.1 Avaliação ADNI

A Tabela 5.7 resume os resultados de desempenho obtidos pela avaliação das combinações em duas classes, em cada conjunto de dados ADNI, para cada modelo de classificação.

Tabela 5.7. ADNI dataset results - two classes

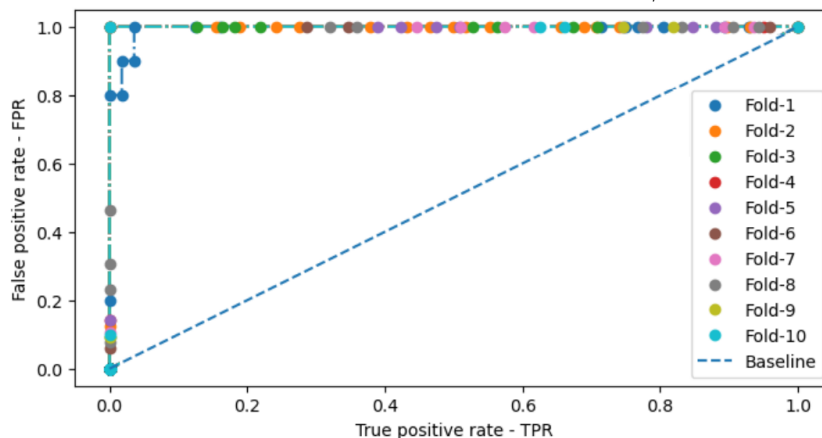
Model / Dataset	Type	B_{ca}	A_{cy}	P_{rn}	S_{ny}	S_{py}	F1	AUC
SVM / ADNI-11	CN v/s AD	99.88%	99.81%	99.55%	99.88%	99.88%	99.70%	1.000
	MCI v/s AD	91.12%	94.45%	92.93%	91.12%	91.12%	91.63%	0.986
	CN v/s MCI	93.98%	94.72%	95.33%	93.98%	93.98%	94.49%	0.975
KNN / ADNI-3	CN v/s AD	99.67%	99.81%	99.87%	99.67%	99.67%	99.76%	0.997
	MCI v/s AD	86.95%	91.89%	89.65%	86.95%	86.95%	87.82%	0.975
	CN v/s MCI	92.79%	93.58%	94.25%	92.79%	92.79%	93.28%	0.967

Source: Own authorship.

Os melhores resultados de avaliação do desempenho foram atingidos para a combinação de classes NC v/s AD em ambas as bases de dados e nas duas combinações de classificadores. A métrica B_{ca} atingiu o valor de 99,88% e a AUC média chegou a 1.000, na precisão de três casas decimais.

As curvas ROC para a base de dados ADNI-11 com as classes NC v/s AD, e avaliadas pelo classificador SVM, são mostradas na Figura 5.4. As curvas ROC, em diferentes cores, foram calculadas para cada uma das dobras (*folds*) do processo de validação cruzada aninhada. As curvas ROC calculadas têm comportamentos muito próximos do comportamento ideal.

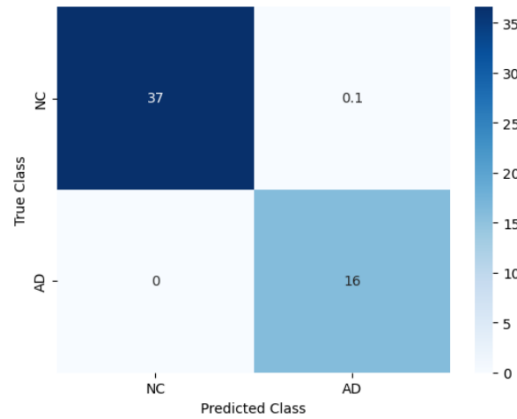
Figura 5.4. ADNI-11 avaliada com o classificador SVM sobre as classes CN v/s AD - curvas ROC.



Source: Own authorship.

A Figura 5.5 mostra a matriz de confusão para a avaliação da base de dados ADNI-11 pelo classificador SVM na condição das classes NC v/s AD.

Figura 5.5. ADNI-11 avaliated with the SVM classifier - CN v/s AD - average confusion matrix.



Source: Own authorship.

5.1.5.2 Avaliação AIBL

A Tabela 5.8 resume os resultados de desempenho obtidos pela avaliação das combinações em duas classes, em cada conjunto de dados AIBL, para cada modelo de classificação.

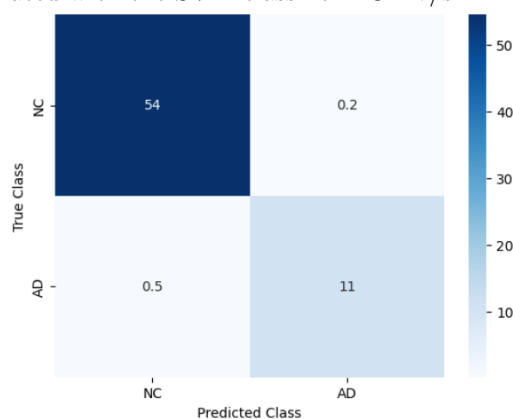
Tabela 5.8. AIBL dataset results - two classes

Model/ Dataset	Type	B_{ca}	A_{cy}	P_{rn}	S_{ny}	S_{py}	F1	AUC
SVM / AIBL-4	CN v/s AD	96.76%	98.94%	98.66%	96.76%	96.76%	97.55%	0.999
	MCI v/s AD	90.90%	91.16%	91.48%	90.90%	90.90%	90.82%	0.953
	CN v/s MCI	93.99%	94.51%	89.80%	93.99%	93.99%	91.55%	0.936
KNN / AIBL-6	CN v/s AD	95.80%	98.48%	97.67%	95.80%	95.80%	96.67%	0.986
	MCI v/s AD	87.69%	88.70%	89.83%	87.69%	87.69%	87.84%	0.910
	CN v/s MCI	93.72%	94.65%	90.19%	93.72%	93.72%	91.69%	0.941

Source: Own authorship.

Os melhores resultados de avaliação do desempenho foram atingidos para a combinação de classes NC v/s AD em ambas as bases de dados e nas duas combinações de classificadores. A métrica B_{ca} atingiu o valor de 96,76%.

A Figura 5.6 mostra a matriz de confusão para a avaliação da base de dados AIBL-4 pelo classificador SVM na condição das classes NC v/s AD.

Figura 5.6. AIBL-4 avaliated with the SVM classifier - CN v/s AD - average confusion matrix.

Source: Own authorship.

5.1.5.3 Avaliação OASIS

A Tabela 5.9 resume os resultados de desempenho obtidos pela avaliação das combinações em duas classes, em cada conjunto de dados OASIS, para cada modelo de classificação.

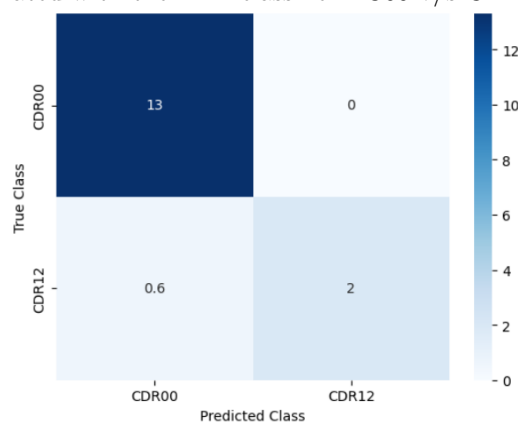
Tabela 5.9. OASIS dataset results - two classes

Model / Dataset	Type	B_{ca}	A_{cy}	P_{rn}	S_{ny}	S_{py}	F1	AUC
SVM / OASIS-5	C00 v/s C12	88.04%	94.29%	92.85%	88.04%	88.04%	88.66%	0.978
	C05 v/s C12	65.82%	72.50%	71.70%	65.82%	65.82%	63.54%	0.783
	C00 v/s C05	72.77%	81.58%	81.59%	72.77%	72.77%	72.71%	0.846
KNN / OASIS-5	C00 v/s C12	89.17%	96.17%	97.93%	89.17%	89.17%	91.23%	0.919
	C05 v/s C12	63.82%	71.25%	63.50%	63.82%	63.82%	62.18%	0.733
	C00 v/s C05	65.85%	78.42%	75.77%	65.85%	65.85%	64.92%	0.778

Source: Own authorship.

Os melhores resultados de avaliação do desempenho foram atingidos para a combinação de classes C00 v/s C12 em ambas as bases de dados e nas duas combinações de classificadores. A métrica B_{ca} atingiu o valor de 89,17%.

A Figura 5.7 mostra a matriz de confusão para a avaliação da base de dados OASIS-5 pelo classificador kNN na condição das classes C00 v/s C12.

Figura 5.7. OASIS-5 avaliada com o kNN classificador - C00 v/s C12 - average confusion matrix.

Source: Own authorship.

5.1.6 Avaliação do desempenho com outros classificadores

Neste tópico, as bases de dados ADNI-11, AIBL-4 e OASIS-5, que apresentaram o melhor desempenho nas avaliações anteriores, foram avaliadas por outros classificadores, em termos das métricas: acurácia balanceada, precisão, sensibilidade, especificidade, escore f1 e a multi-class area under the curve (MAUC). Nessa avaliação, a classificação foi feita para as classes CN, MCI e AD, no caso das bases AIBL-4 e ADNI-11, e para as classes CDR 0, CDR 0.5 e CDR 12, para a base de dados OASIS-5.

Além dos classificadores SVM e kNN, a avaliação foi feita com os classificadores *Random Forest*, *Logistic Regression*, *Naive Bayes*, *Decision Tree*, *Ada Boost*, *Linear Discriminant Analysis* (LDA), *Gradient Boosting*, *Extreme Gradient Boosting* e *Histogram Gradient Boosting*. Da mesma forma que nos outros casos, todas as avaliações foram feitas utilizando validação cruzada aninhada.

A Tabela 5.10 apresenta os resultados da avaliação do desempenho de vários classificadores para a base de dados AIBL-4. Esses resultados apontaram para o melhor desempenho do classificador *Gradient Boosting*, tendo a métrica B_{ca} atingido o valor de 92,88%.

Tabela 5.10. AIBL-4 - performance evaluation

Model	B_{ca}	A_{cy}	P_{rn}	S_{yn}	S_{py}	F1	MAUC
SVM	92,52%	92,22%	88,25%	89,06%	95,98%	87,74%	0,959
KNN	91.21%	91.46%	87.02%	87.12%	95.30%	86.23%	0.960
Random Forest	87.79%	89.29%	84.26%	82.77%	92.82%	82.83%	0.958
Logistic Regression	91.60%	91.71%	88.01%	88.03%	95.17%	87.28%	0.975
Naive Bayes	91.80%	91.72%	87.72%	88.08%	95.52%	87.10%	0.974
Decision Tree	91.96%	91.46%	87.73%	88.36%	95.57%	87.03%	0.959
Ada Boost	84.15%	87.77%	80.85%	77.22%	91.08%	78.27%	0.949
LDA	88.15%	90.18%	86.47%	82.93%	93.37%	83.86%	0.974
Gradient Boosting	92.88%	92.61%	88.94%	89.63%	96.13%	88.60%	0.974
Extreme GB	91.62%	91.96%	88.38%	87.87%	95.36%	87.38%	0.967
Histogram GB	87.97%	89.54%	84.38%	82.71%	93.24%	82.69%	0.961

Source: Own authorship.

A Tabela 5.10 apresenta os resultados da avaliação do desempenho de vários classificadores para a base de dados ADNI-11. Esses resultados apontaram o melhor desempenho para o classificador SVM, tendo a métrica B_{ca} atingido o valor de 92,61%.

Tabela 5.11. ADNI-11 - performance evaluation

Model	B_{ca}	A_{cy}	P_{rn}	S_{yn}	S_{py}	F1	MAUC
SVM	92,61%	91,77%	92,34%	90,20%	95,02%	90,90%	0,969
KNN	83.56%	78.13%	80.61%	79.54%	87.58%	79.57%	0.891
Random Forest	92.36%	91.39%	91.26%	89.83%	94.90%	90.34%	0.968
Logistic Regression	92.46%	91.10%	91.21%	90.14%	94.78%	90.54%	0.973
Naive Bayes	88.42%	83.17%	82.39%	85.71%	91.13%	83.35%	0.933
Decision Tree	91.85%	90.90%	91.12%	89.11%	94.59%	89.86%	0.952
Ada Boost	90.10%	88.10%	87.76%	87.04%	93.16%	86.84%	0.939
LDA	85.20%	80.94%	82.93%	81.33%	89.07%	81.84%	0.918
Gradient Boosting	92.05%	91.49%	92.33%	89.28%	94.81%	90.44%	0.967
Extreme GB	91.29%	90.23%	90.71%	88.42%	94.17%	89.33%	0.966
Histogram GB	92.00%	90.71%	90.65%	89.41%	94.60%	89.79%	0.963

Source: Own authorship.

A Tabela 5.11 apresenta os resultados da avaliação do desempenho de vários classificadores para a base de dados OASIS-5. Esses resultados evidenciaram o melhor desempenho do classificador *Decision Tree*, tendo a métrica B_{ca} atingido o valor de 76,25%.

Tabela 5.12. OASIS-5 - performance evaluation

Model	B _{ca}	A _{cy}	P _{rn}	S _{yn}	S _{py}	F1	MAUC
SVM	72.66%	73.64%	70.85%	61.75%	83.57%	58.33%	0.872
KNN	68.17%	71.36%	64.80%	54.53%	81.82%	51.40%	0.804
Random Forest	75.30%	76.41%	72.49%	65.05%	85.56%	62.50%	0.881
Logistic Regression	73.74%	74.55%	72.58%	63.10%	84.39%	60.34%	0.876
Naive Bayes	73.86%	75.00%	70.04%	63.05%	84.66%	62.38%	0.876
Decision Tree	76.25%	74.57%	67.98%	68.10%	84.40%	64.92%	0.842
Ada Boost	65.58%	64.20%	53.02%	50.71%	80.46%	49.48%	0.811
LDA	73.87%	75.00%	70.88%	64.34%	83.41%	63.00%	0.868
Gradiente Boosting	73.54%	74.98%	71.42%	62.95%	84.13%	61.66%	0.890
Extreme GB	73.49%	75.02%	68.72%	62.67%	84.31%	61.57%	0.863
Histogram GB	70.36%	70.30%	63.24%	57.39%	83.34%	55.13%	0.881

Source: Own authorship.

5.1.7 Avaliação com outros métodos de seleção de atributos

As bases de dados ADNI, AIBL e OASIS também foram submetidas a outros métodos de seleção de atributos, dos três grupos de técnicas referenciadas no Capítulo 3. Da mesma forma que feito para o algoritmo FMFS, cada uma das técnicas foi avaliada para os classificadores kNN e SVM, e encontrado o número de atributos que resultou na melhor acurácia.

A Tabela 5.13 mostra o valor encontrado para o número de atributos e acurácia ótimos, avaliados sobre o banco de dados ADNI, para cada um dos classificadores.

Tabela 5.13. ADNI - feature selection method evaluation

Feature selection method	kNN		SVM	
	Features	B _{ca}	Features	B _{ca}
FMFS	3	90.06%	11	92.61%
Chi-square	2	89.68%	17	92.81%
Mutual Information	1	90.31%	5	92.93%
Anova F-value	1	90.31%	4	92.93%
Variance Threshold	6	61.50%	11	74.66%
Fisher Score	13	83.82%	17	92.62%
MultiSURF	6	86.70%	6	92.07%
RFE (SVM)	1	90.31%	3	92.84%
Permutation Importance (LR)	2	91.72%	4	92.93%
SHAP (XGboost)	2	91.72%	3	92.97%
Boruta (RF)	3	90.54%	5	92.93%
Embedded Random Forest	1	90.31%	4	92.93%
Embedded LightGBM	1	90.31%	5	92.83%

Source: Own authorship.

A Tabela 5.14 mostra o valor encontrado para o número de atributos e acurácia ótimos, avaliados sobre o banco de dados AIBL, para cada um dos classificadores.

Tabela 5.14. AIBL - feature selection method evaluation

Feature selection method	kNN		SVM	
	Features	B _{ca}	Features	B _{ca}
FMFS	6	90.44%	4	92.52%
Chi-square	4	91.21%	4	92.52%
Mutual Information	1	88.30%	5	92.43%
Anova F-value	1	88.30%	7	92.27%
Variance Threshold	1	50.25%	10	80.90%
Fisher Score	1	88.30%	12	92.11%
MultiSURF	4	87.66%	11	92.93%
RFE (SVM)	3	91.86%	6	92.54%
Permutation Importance (LR)	1	88.30%	12	92.29%
SHAP (XGboost)	1	50.25%	9	92.49%
Boruta (RF)	2	90.82%	4	92.52%
Embedded Random Forest	1	88.30%	3	92.79%
Embedded LightGBM	1	88.30%	7	92.76%

Source: Own authorship.

A Tabela 5.15 mostra o valor encontrado para o número de atributos e acurácia ótimos, avaliados sobre o banco de dados OASIS, para cada um dos classificadores.

Tabela 5.15. OASIS - feature selection method evaluation

Feature selection method	kNN		SVM	
	Features	B _{ca}	Features	B _{ca}
FMFS	5	68.17%	5	72.66%
Chi-square	2	77.59%	6	72.14%
Mutual Information	2	77.59%	4	72.74%
Anova F-value	2	77.59%	3	75.30%
Variance Threshold	1	50.72%	4	70.03%
Fisher Score	5	67.63%	7	72.13%
MultiSURF	1	72.91%	6	71.97%
RFE (SVM)	1	72.91%	6	71.97%
Permutation Importance (LR)	1	72.91%	4	72.57%
SHAP (XGboost)	3	74.60%	7	72.13%
Boruta (RF)	2	77.59%	3	75.30%
Embedded Random Forest	2	77.59%	3	75.30%
Embedded LightGBM	4	68.14%	5	70.22%

Source: Own authorship.

5.2 RESULTADOS DA ANÁLISE POR IMAGEM

Os procedimentos práticos das análises por imagem foram realizados em linguagem python versão 3.9.0, com as bibliotecas numpy versão 1.19.2, scikit-learn versão 1.1.1, tensorflow versão 2.6.0, cuda toolkit versão 11.3.1 e utilizado um processador 12th Gen Intel(R) Core(TM) i7-12700H 2.30 GHz, com 16 GB de RAM. Também foi utilizada uma placa *Graphics Processing Unit* (GPU) da NVIDIA GeForce RTX 3060, com 6 GB de RAM.

Particularmente com relação às operações de DL, a utilização da GPU teve o mérito de reduzir bastante o tempo de processamento, aumentando a produtividade na geração dos resultados.

A avaliação de desempenho foi feita em termos das métricas: acurácia balanceada, acurácia, precisão, sensibilidade, especificidade, score f1 e a *area under the curve* (AUC). Ao final, foi possível comparar o desempenho da classificação entre os bancos de imagem gerados, assim como foi possível comparar com os resultados obtidos em outros estudos.

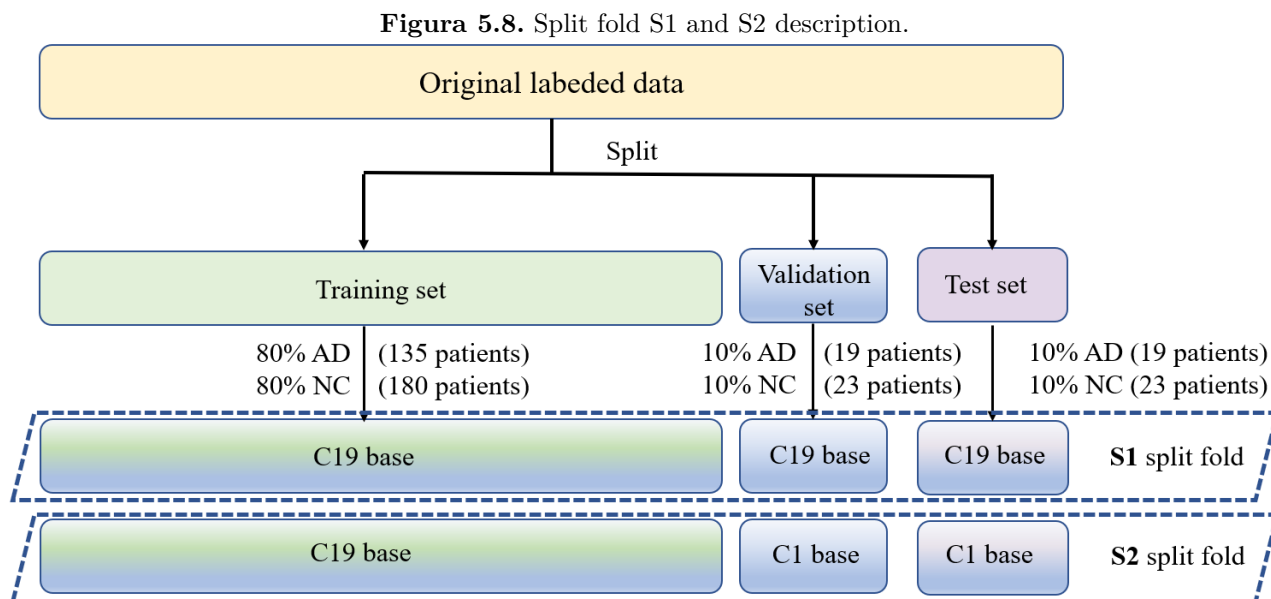
5.2.1 Análise do banco de imagens coronal

Conforme descrito no Capítulo 4, foram geradas duas versões de bancos de imagem coronal, sendo um com a presença somente do *slice* central (Figura 4.20) - banco C1, e outro com o *slice* central, adicionado dos nove *slices* anteriores e nove *slices* posteriores (Figura 4.21) - banco C19. O banco C1 é muito reduzido, contendo imagens geradas de 226 pacientes em CN e 173 pacientes em AD, o que inviabiliza o treinamento de redes convolucionais. Por outro lado, o banco C19 possui 4.294 imagens de CN e 3.287 imagens de AD, dado a estratégia de uso de *slices* anteriores e *slices* posteriores, e que pode viabilizar o treinamento de redes convolucionais.

5.2.1.1 Alternativas de montagem das partições

Conforme visto no Capítulo 3, é necessário construir um conjunto de três partições, ver Figura 3.14, que viabilize o treinamento das CNNs, sem correr qualquer risco de vazamento de dados entre as partições de treino e teste. Assim, foi proposta a utilização dos dois bancos de imagens C1 e C19 para a montagem das partições S1 e S2. A Figura 5.8 ilustra os elementos

de construção das partições.



Source: Own authorship.

A partição S1 é formada por imagens de 315 pacientes no banco C19 para compor a partição de treinamento, imagens de outros 42 pacientes do banco C19 para compor a partição de validação e imagens de também outros 42 pacientes do banco C19 para compor a partição de teste.

A partição S2 é formada por imagens de 315 pacientes no banco C19 para compor a partição de treinamento, imagens de outros 42 pacientes do banco C1 para compor a partição de validação e imagens de também outros 42 pacientes do banco C1 para compor a partição de teste.

O resultado final da composição das partições é mostrado na Tabela [5.16](#).

Tabela 5.16. Split fold composition S1 and S2

Split fold	Train	Validation	Test
S1	2.565 AD images	361 AD images	361 AD images
	3.420 CN images	437 CN images	437 CN images
S2	2.565 AD images	19 AD images	19 AD images
	3.420 CN images	23 CN images	23 CN images

Source: Own authorship.

5.2.1.2 Avaliação da partição S1

A avaliação de desempenho da partição S1 foi feita utilizando os modelos de DL propostos, CNN-1, CNN-2, CNN-3, CNN-4 e CNN-5, e os modelos pré-treinados, VGG-19, VGG-16, ResNet-50 e Inception V3. Os resultados estão explicitados na Tabela 5.17.

Tabela 5.17. S1 split fold - performance evaluation

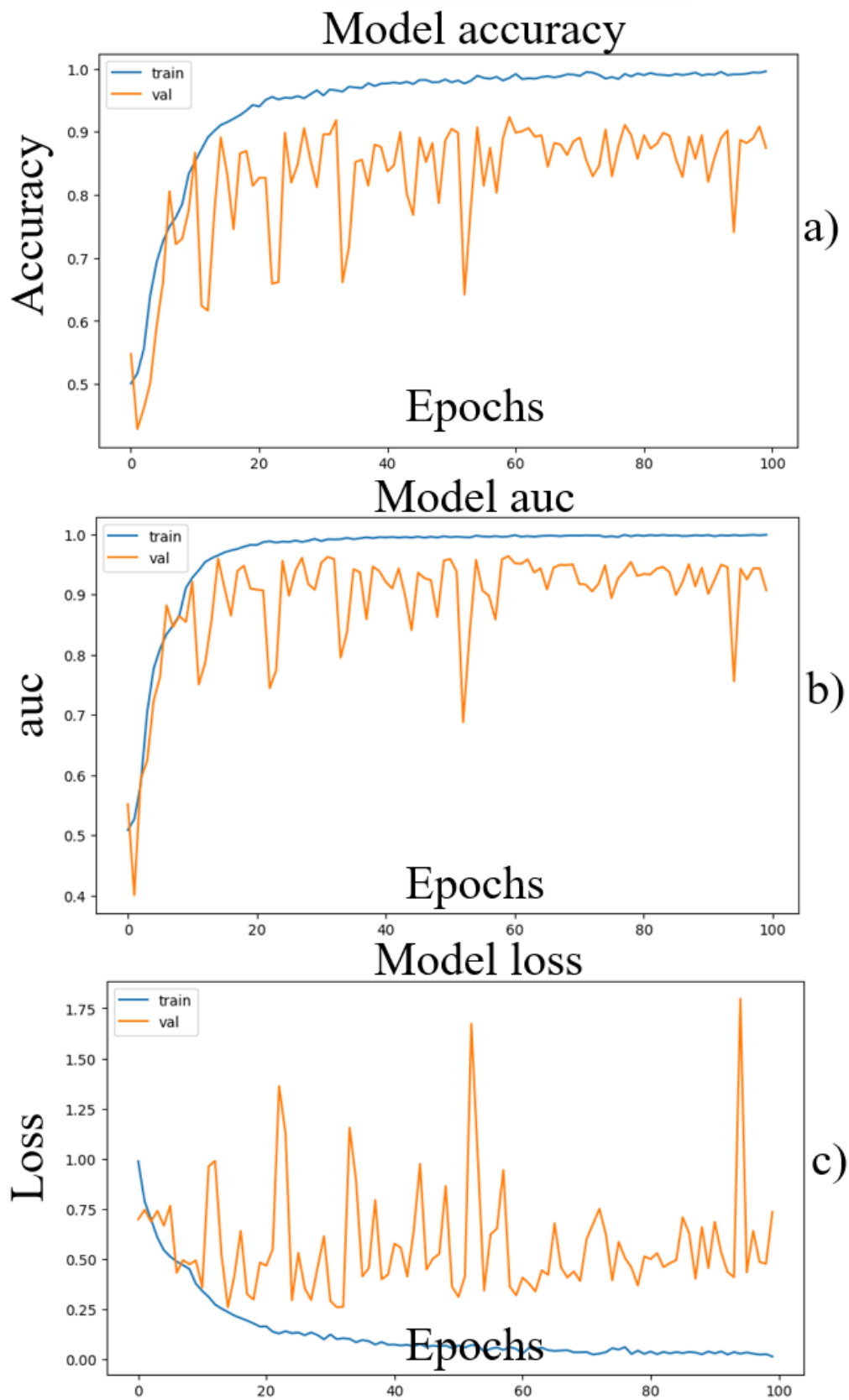
Model	B_{ca}	A_{cy}	P_{rn}	S_{yn}	S_{py}	F1	AUC
CNN-1	94.25%	94.36%	94.34%	95.42%	93.07%	94.88%	0.942
CNN-2	90.50%	90.23%	94.10%	87.64%	93.35%	90.76%	0.905
CNN-3	83.20%	84.71%	78.58%	99.08%	67.31%	87.65%	0.832
CNN-4	93.62%	93.36%	96.83%	90.85%	96.40%	93.74%	0.936
CNN-5	94.79%	94.74%	96.04%	94.28%	95.29%	95.15%	0.948
VGG-19	86.22%	86.59%	86.03%	90.16%	82.27%	88.04%	0.862
VGG-16	84.09%	84.71%	83.02%	90.62%	77.56%	86.65%	0.841
Resnet-50	83.13%	84.21%	80.19%	94.51%	71.75%	86.76%	0.831
Inception V3	81.33%	81.95%	80.84%	87.87%	74.79%	84.21%	0.813

Source: Own authorship.

Na fase de treino para este e demais casos, todos os modelos foram treinados utilizando o otimizador *Adam* com taxa de aprendizagem de 0.001, 100 épocas e tamanho do lote definido para 16 com base na memória da GPU (8 GB).

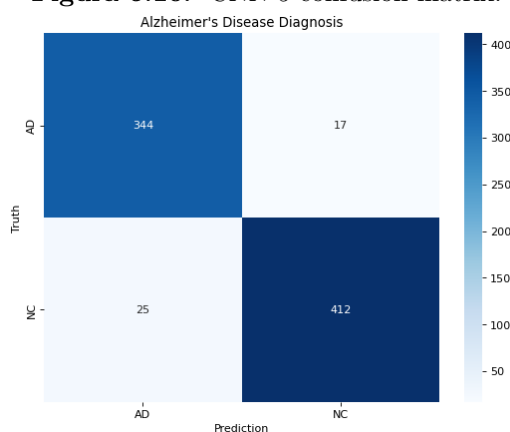
Os modelos pré-treinados, VGG-19, VGG-16, ResNet-50 e Inception V3, foram inicializadas com pesos do treinamento realizado no *dataset* ImageNet e tiveram as primeiras camadas congeladas para treinamento, até a penúltima camada convolucional. Os modelos propostos, CNN-1, CNN-2, CNN-3, CNN-4 e CNN-5, foram treinados por inteiro.

Para a CNN-5 foi calculada a curva de aprendizado, avaliada para 100 épocas, e o resultado pode ser visto na Figura 5.9.

Figura 5.9. CNN-5 model learning curves. a) Accuracy model. b) AUC model. c) Loss model.

Source: Own authorship.

Também para a CNN-5 a matriz de confusão foi avaliada e mostrada na Figura [5.10](#).

Figura 5.10. CNN-5 confusion matrix.

Source: Own authorship.

5.2.1.3 Avaliação da partição S2

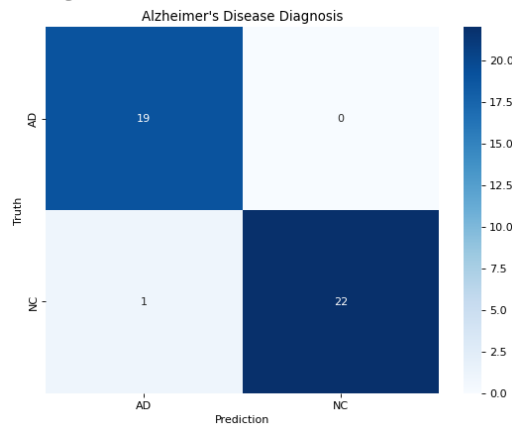
Da mesma forma que na partição S1, a avaliação de desempenho da partição S2 foi feita utilizando os modelos de DL propostos, CNN-1, CNN-2, CNN-3, CNN-4 e CNN-5. Os resultados estão explicitados na Tabela 5.18.

Tabela 5.18. S2 split fold - performance evaluation

Model	B_{ca}	A_{cy}	P_{rn}	S_{yn}	S_{py}	F1	AUC
CNN-1	95.65%	95.24%	100.0%	91.30%	100.0%	95.45%	0.957
CNN-2	91.30%	90.48%	100.0%	82.61%	100.0%	90.48%	0.913
CNN-3	92.11%	92.86%	88.46%	100.0%	84.21%	93.88%	0.921
CNN-4	97.83%	97.62%	100.0%	95.65%	100.0%	97.78%	0.978
CNN-5	89.93%	90.48%	88.00%	95.65%	84.21%	91.67%	0.899

Source: Own authorship.

Também para a CNN-4 a matriz de confusão foi avaliada e mostrada na Figura 5.11.

Figura 5.11. CNN-4 confusion matrix.

Source: Own authorship.

5.2.1.4 Reavaliação da partição S2 com aumento de dados

Adicionalmente ao uso da base C19 para compor a partição de treinamento em S2, onde já há 19 *slices* selecionados para cada paciente, foi feito um aumento de dados adicional em 2x sobre as imagens usando uma faixa de variação aleatória de [0.99, 1.01] no zoom, faixa de variação aleatória de [0.8, 1.2] no brilho e inversão aleatória horizontal. Essa estratégia corresponde a um aumento de dados clássico aplicados sobre as imagens. Os resultados de desempenho dessa estratégia estão apresentados na Tabela 5.19.

Tabela 5.19. Data augmentation in S2 training dataset - performance evaluation

Model	B_{ca}	A_{cy}	P_{rn}	S_{yn}	S_{py}	F1	AUC
CNN-1	89.93%	90.48%	88.00%	95.65%	84.21%	91.67%	0.899
CNN-2	92.11%	92.86%	88.46%	100.0%	84.21%	93.88%	0.921
CNN-3	89.47%	90.48%	85.19%	100.0%	78.95%	92.00%	0.895
CNN-4	92.56%	92.86%	91.67%	95.65%	89.47%	93.62%	0.926
CNN-5	82.15%	80.95%	94.12%	69.57%	94.74%	80.00%	0.822

Source: Own authorship.

5.2.1.5 Reavaliação de S1 com classificador de votação por maioria

O desempenho da partição S1 foi novamente medido utilizando os modelos DL propostos (CNN-1, CNN-2, CNN-3, CNN-4 e CNN-5) e os resultados foram comparados entre a decisão tomada em cada *slice* individualmente (instância única) e a decisão tomada pela maioria (decisão por múltiplas instâncias). Os resultados podem ser consultados na Tabela 5.20.

Tabela 5.20. Customized cnn results

Model	B_{ca}	A_{cy}	P_{rn}	S_{yn}	S_{py}	F1	AUC
CNN-1 (<i>S</i>)	96.1%	95.9%	97.9%	94.7%	97.4%	96.3%	0.96
CNN-1 (<i>M</i>)	97.8%	97.6%	100%	95.7%	100%	97.8%	0.98
CNN-2 (<i>S</i>)	96.6%	96.5%	97.5%	96.3%	96.8%	96.9%	0.97
CNN-2 (<i>M</i>)	97.8%	97.6%	100%	95.7%	100%	97.8%	0.98
CNN-3 (<i>S</i>)	94.7%	94.9%	95.0%	95.9%	93.6%	95.4%	0.95
CNN-3 (<i>M</i>)	95.1%	95.1%	95.7%	95.7%	94.4%	95.7%	0.95
CNN-4 (<i>S</i>)	95.3%	95.3%	96.5%	95.0%	95.6%	95.7%	0.95
CNN-4 (<i>M</i>)	100%	100%	100%	100%	100%	100%	1.00
CNN-5 (<i>S</i>)	98.2%	98.2%	98.9%	97.9%	98.5%	98.4%	0.98
CNN-5 (<i>M</i>)	100%	100%	100%	100%	100%	100%	1.00

S - Single decision.

M - Majority voting decision.

Source: Own authorship.

Da mesma forma, o desempenho da partição S1 foi novamente medido utilizando os modelos pré-treinados (VGG-19, VGG-16, ResNet-50 e Inception V3) e os resultados foram comparados entre a decisão tomada em cada *slice* individualmente (instância única) e a decisão tomada pela maioria (decisão por múltiplas instâncias). Os resultados podem ser consultados na Tabela [5.21](#).

Tabela 5.21. Pretrained models results

Model	B_{ca}	A_{cy}	P_{rn}	S_{yn}	S_{py}	F1	AUC
VGG-16 (<i>S</i>)	92.9%	92.9%	93.8%	93.4%	92.1%	93.7%	0.93
VGG-16 (<i>M</i>)	97.2%	97.6%	95.8%	100%	94.4%	97.9%	0.97
VGG-19 (<i>S</i>)	95.7%	95.8%	96.1%	96.3%	95.0%	96.2%	0.96
VGG-19 (<i>M</i>)	97.2%	97.6%	95.8%	100%	94.4%	97.9%	0.97
ResNet (<i>S</i>)	97.5%	97.3%	99.5%	95.7%	99.4%	97.6%	0.98
ResNet (<i>M</i>)	97.8%	97.6%	100%	95.6%	100%	97.8%	0.98
Inception (<i>S</i>)	95.9%	95.9%	96.6%	96.1%	95.6%	96.3%	0.96
Inception (<i>M</i>)	97.8%	97.6%	100%	95.6%	100%	97.8%	0.98

S - Single decision.

M - Majority voting decision.

Source: Own authorship.

As CNNs pré-treinadas foram inicializadas com pesos de treino no conjunto de dados ImageNet, finalizadas com 3 camadas densas e treinadas na totalidade. O ADAM foi utilizado como otimizador e o treino foi efetuado em 100 épocas, com um critério de parada para uma precisão de validação superior a 0,999.

Pode ver-se na Tabela [5.20](#) que a CNN-4 obteve um desempenho de 100% em todas as métricas objectivas utilizadas. Para investigar o desempenho da CNN-4 em relação ao ponto

quiescente (mínimo local) onde o algoritmo de treinamento converge, a rede foi retreinada 10 vezes com pontos de partida aleatórios e todos os modelos treinados foram salvos. Utilizando os modelos obtidos na etapa de treino, o desempenho da rede foi novamente avaliado para cada caso. Os resultados do desempenho são apresentados na Tabela 5.22.

Tabela 5.22. CNN-4 training results

Train	B _{ca}	A _{cy}	P _{rn}	S _{yn}	S _{py}	F1	AUC
1	100%	100%	100%	100%	100%	100%	1.00
2	100%	100%	100%	100%	100%	100%	1.00
3	97.8%	97.6%	100%	95.7%	100%	97.8%	0.98
4	97.2%	97.6%	95.8%	100%	94.4%	97.8%	0.97
5	100%	100%	100%	100%	100%	100%	1.00
6	100%	100%	100%	100%	100%	100%	1.00
7	100%	100%	100%	100%	100%	100%	1.00
8	97.2%	97.6%	95.8%	100%	94.4%	97.8%	0.97
9	100%	100%	100%	100%	100%	100%	1.00
10	100%	100%	100%	100%	100%	100%	1.00
Mean	99.2%	99.3%	99.2%	99.6%	98.9%	99.3%	0.99
Std	1.3%	1.2%	1.8%	1.4%	2.4%	1.1%	0.01

Mean - statistical average.

Std - standard deviation.

Source: Own authorship.

A Tabela 5.22 revela que a CNN-4 alcançou um valor médio de exatidão de 99,3% na distinção do diagnóstico entre CN e AD.

5.2.2 Análise do banco de imagens axial

Conforme descrito no Capítulo 4, foi gerada uma versão de banco de imagem axial e, da mesma forma que no caso da estratégia no coronal, também com com um *slice* central, adicionado de quatro *slices* caudais e quatro *slices* craniais (Figura 4.15). O banco de dados axial foi gerado com as três classes CN, MCI e AD. No entanto, no sentido de facilitar as comparações de desempenho dos bancos axial e coronal, decidiu-se trabalhar com as classes CN e AD. As partições de treino, validação e teste ficaram com a quantidade de imagens descrita na Tabela 5.23.

Tabela 5.23. Split axial dataset

Split	Training	Validation	Test
Axial	1.494 AD images	81 AD images	81 AD images
	1.728 CN images	99 CN images	90 CN images

Source: Own authorship.

A avaliação de desempenho do banco de imagens axial foi feita utilizando os modelos de DL propostos CNN-1, CNN-2, CNN-3, CNN-4 e CNN-5. As métricas de desempenho estão explicitadas na Tabela 5.24.

Tabela 5.24. Axial plan - performance evaluation

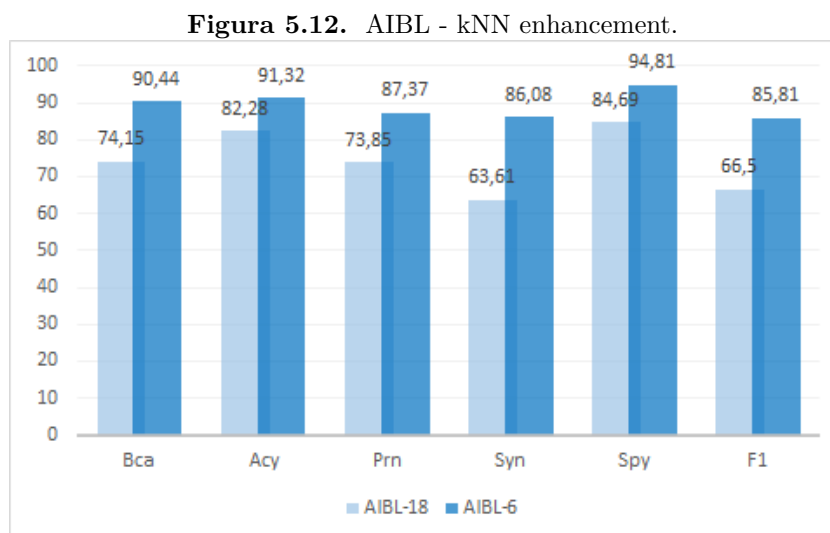
Model	B_{ca}	A_{cy}	P_{rn}	S_{yn}	S_{py}	F1	AUC
CNN-1	55.99%	56.73%	57.27%	70.00%	41.98%	63.00%	0.560
CNN-2	42.41%	43.27%	46.90%	58.89%	25.93%	52.22%	0.424
CNN-3	60.93%	59.06%	88.46%	25.56%	96.3%	39.66%	0.609
CNN-4	59.94%	59.06%	67.24%	43.33%	76.54%	52.70%	0.599
CNN-5	59.01%	59.06%	61.36%	60.00%	58.02%	60.67%	0.590

Source: Own authorship.

5.3 DISCUSSÃO DOS RESULTADOS DA ANÁLISE SOBRE DADOS CLÍNICOS

O modelo kNN avaliado sobre as bases AIBL-18 e AIBL-6 resultou numa melhoria de desempenho da ordem de 22% sobre a métrica B_{ca} . Esse resultado é estatisticamente significativo, correspondente a um valor de $p < 0,05$.

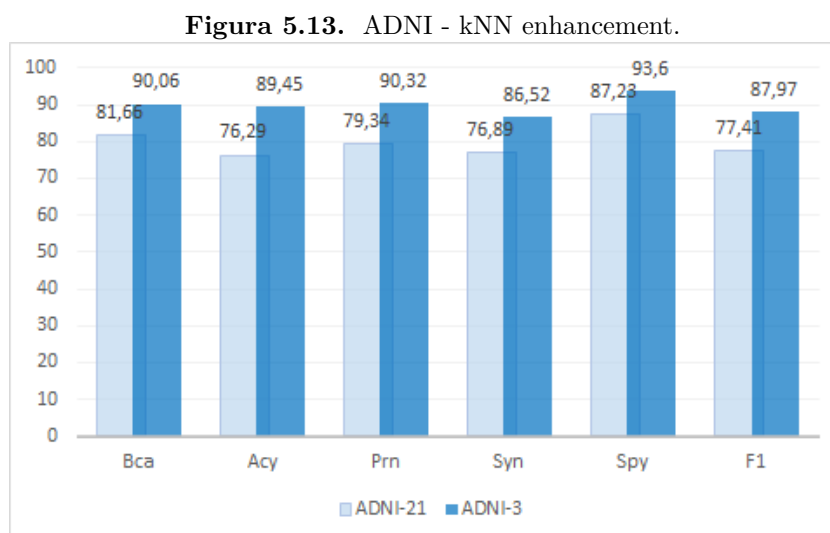
A Figura 5.12 ilustra a melhoria de todas as métricas avaliadas, apresentadas na Tabela 5.3, e mostra o impacto do algoritmo de seleção de atributos FMFS sobre a melhoria nas métricas de avaliação do modelo kNN sobre a base de dados AIBL.



Source: Own authorship.

Do mesmo modo, o modelo kNN avaliado sobre as bases ADNI-21 e ADNI-3 resultou numa melhoria de desempenho da ordem de 10% sobre a métrica B_{ca} , com significância estatística, correspondente a um valor de $p < 0,05$.

A Figura 5.13 ilustra a melhoria de todas as métricas avaliadas, apresentadas na Tabela 5.4, e mostra o impacto do algoritmo de seleção de atributos FMFS sobre a melhoria nas métricas de avaliação do modelo kNN sobre a base de dados ADNI.



Source: Own authorship.

Portanto, pode-se afirmar que os conjuntos de dados AIBL-6 e ADNI-3, submetidos ao modelo kNN, resultaram em melhor desempenho na estimativa do diagnóstico de Alzheimer,

permitiu a redução de dimensionalidade, realçou os atributos mais importantes e reduziu o tempo de treinamento, objetivos desejáveis para o método de seleção de atributos proposto.

O modelo SVM não apresentou melhoria estatisticamente significativa de desempenho, com relação à métrica B_{ca} , comparando os resultados obtidos sobre as bases de dados AIBL-18 e AIBL-4 nem relativo às bases ADNI-21 e ADNI-11. Porém, os objetivos de redução de dimensionalidade, realce dos atributos mais importantes e redução do tempo de treinamento foram atingidos.

O algoritmo FMFS colabora para eliminar os atributos cujos exemplos não contribuem significativamente com o distanciamento das médias entre as classes da base de dados. Essa estratégia tende a facilitar o desempenho do algoritmo kNN, por eliminar o equivalente a ruído no cálculo dos vizinhos mais próximos. O resultado desta combinação foi uma melhoria de 10,29% na acurácia na avaliação da base ADNI e 21,97% na acurácia na avaliação da base AIBL.

Já para o modelo SVM, a consequência do uso do algoritmo FMFS não impacta significativamente o desempenho, pelo fato de a função *kernel* possuir um parâmetro de regularização (conhecido como C), que permite encontrar o equilíbrio ideal entre uma ampla margem e um pequeno número de pontos de dados classificados incorretamente. Controlar C então equivale a controlar *outliers* - C baixo implica que estamos permitindo mais *outliers*, C alto implica que estamos permitindo menos *outliers*. Para todos os modelos de classificação utilizados nesta tese, os hiper-parâmetros foram ajustados no loop interno da validação cruzada aninhada.

A avaliação dos conjuntos de dados por outros 12 métodos de seleção de atributos produziu resultados de acurácia inferiores e superiores aqueles conseguidos com o algoritmo FMFS proposto. A simplicidade no cálculo do FMFS e a a qualidade dos resultados obtidos mostram ser o FMFS uma opção viável para avaliação dos conjuntos de dados tratados nesta tese.

5.3.1 Comparação com outras pesquisas

Os melhores resultados obtidos neste trabalho foram comparados com os melhores resultados em pesquisas anteriores para os três estudos sobre Alzheimer: AIBL, ADNI e OASIS, detalhados nas Tabelas [5-25](#), [5-26](#) e [5-27](#). Todas as referências utilizadas estão compreendidas entre os

anos de 2018 e 2021. Conforme apresentado em (KHAN *et al.*, 2021), o maior número de trabalhos presentes na literatura faz utilização da base de dados ADNI, onde foi encontrado o maior número de referências. A comparação foi feita sobre os resultados de avaliação de desempenho para o caso de três classes: CN, MCI e AD, no caso das bases AIBL e ADNI, e para as classes CDR 0, CDR 0.5 e CDR 12, para a base de dados OASIS.

Nessa comparação entre estudos, foram utilizadas somente as métricas B_{ca} , MAUC, as técnicas de aprendizado de máquina usadas e os modelos de aprendizado de máquina avaliados em cada estudo.

A Tabela 5.25 faz o resumo da comparação dos resultados deste estudo, para a base de dados AIBL, com o estudo (NIYAS; P, 2021b), que também atingiu um valor para a métrica B_{ca} próximo a 92,50%, para o caso do classificador SVM. Nesse caso, é possível afirmar que houve equivalência entre os resultados. Avaliado para o classificador *Gradient Boosting*, o resultado obtido nesta tese também ficou um pouco superior, tendo ficado com 92,88% de B_{ca} .

Tabela 5.25. Comparison with previous study (AIBL dataset)

Reference	ML/DL	Model	B_{ca}	MAUC
Present study	ML	GB	92,88	0,974
		SVM	92,52	0,959
		KNN	82,50	NA
(NIYAS; P, 2021b)	ML	SVM	92,50	NA

NA - not available. GB - Gradient Boosting.

Except for MAUC, all numerical results are expressed as percentages.

Source: Own authorship.

Para a base de dados AIBL, os melhores resultados de aumento de desempenho, utilizando o classificador kNN, foram atingidos pelo uso do algoritmo FMFS, quando a métrica B_{ca} passou de 74,15% para 90,44%, conforme pode ser verificado na Tabela 5.5.

A Tabela 5.26 faz o resumo da comparação dos resultados deste trabalho, para a base de dados ADNI, com os estudos (NIYAS; P, 2021a), (NIYAS; P, 2021b), (NGUYEN M, 2020), (LIU J L, 2020), (ALBRIGHT, 2019), (IDDI *et al.*, 2019), (MOORE *et al.*, 2019), (GHAZI *et al.*, 2019) e (NGUYEN *et al.*, 2018). Neste estudo, a métrica B_{ca} atingiu o valor de 92,61%, para o classificador SVM, e se mostrou 2,9% superior ao melhor resultado reportado pelos outros

estudos. Também o classificador *logistic regression* apresentou resultado de 92,46% para a B_{ca} .

Tabela 5.26. Comparison with previous studies (ADNI dataset)

Reference	ML/DL	Model	B_{ca}	MAUC
		LR	92,46	0,973
Present study	ML			
		SVM	92,61	0,969
(NIYAS; P, 2021a)	ML	RF	87	NA
		KNN	90	0,96
(NIYAS; P, 2021b)	ML			
		SVM	88	0,95
(NGUYEN M, 2020)	DL	RNN	88,7	0,944
(LIU J L, 2020)	DL	RNN	87,93	0.9439
(ALBRIGHT, 2019)	DL	RNN	NA	0,866
(DDI <i>et al</i> , 2019)	ML	RF	86	NA
(MOORE <i>et al</i> , 2019)	ML	RF	73	0,82
(GHAZI <i>et al</i> , 2019)	ML	LDA	NA	0,75
(NGUYEN <i>et al</i> , 2018)	ML	SVM	79	0,86

NA - not available. LDA-linear discriminant analysis. RNN-recurrent neural network. RF-random forest. LR-logistic regression.

Except for MAUC, all numerical results are expressed as percentages.

Source: Own authorship.

Os resultados obtidos nesta tese sobre o diagnóstico de AD fazendo uso das bases de dados ADNI e AIBL, com ênfase em avaliação por dados clínicos, foram aprovados e apresentados no IX Congresso Latino-Americano de Engenharia Biomédica (CLAIB 2022) e no XXVIII Congresso Brasileiro de Engenharia de Engenharia Biomédica (CBEB 2022), por intermédio do artigo intitulado "Seleção simplificada de atributos para auxílio ao diagnóstico da doença de Alzheimer utilizando aprendizado de máquina" (SARAIVA JR. *et al*, 2022). O artigo apresentou o algoritmo de seleção de atributos aqui proposto - FMFS, o impacto da sua utilização na melhoria das métricas de desempenho e a comparação com estudos anteriores.

A base de dados OASIS, juntamente com a base AIBL, também é pouco referenciada na literatura (KHAN *et al*, 2021), porém foi possível comparar o desempenho com o estudo (PAIVA; ESCOVEDO, 2021). A Tabela 5.27 mostra um desempenho superior de 4,5% do melhor resultado da B_{ca} deste trabalho comparado com o melhor resultado obtido em (PAIVA; ESCOVEDO, 2021). Também o classificador *random forest* apresentou desempenho superior, tendo atingido a métrica B_{ca} de 75,30%.

Tabela 5.27. Comparison with previous study (OASIS dataset)

Reference	ML/DL	Model	B_{ca}	MAUC
Present study	ML	RF	75,30	0,881
		DT	76,25	0,842
(PAIVA; ESCOVEDO, 2021)	ML	RF	64,30	NA
		LR	73,00	NA

NA - not available. RF-random forest. DT-decision tree. LR-logistic regression.

Except for MAUC, all numerical results are expressed as percentages.

Source: Own authorship.

5.3.2 Outras considerações sobre a análise em dados clínicos

O uso do algoritmo FMFS sobre a base de dados AIBL atingiu redução em $4.5\times$, de 18 para quatro atributos. Para a base de dados ADNI, a redução chegou a $7\times$, de 21 para três atributos. No entanto, para a base de dados OASIS, o impacto de redução foi de apenas $1.4\times$. Em parte, isso explica o motivo de as métricas de avaliação comparativa de desempenho, expressas na Tabela 5.6, não terem apresentado melhoria significativa.

Enquanto a base de dados AIBL-18 possui 785 instâncias e a ADNI-21 possui 1033 instâncias, a OASIS-7 possui apenas 217 instâncias. Isso explica a maior dificuldade dos modelos de classificação em generalizar as previsões sobre a base de dados OASIS-7, para o caso de avaliação em três e duas classes.

A avaliação do desempenho para duas classes mostrou resultados de diagnóstico da AD muito próximo a 100% de acerto. Particularmente, a base de dados ADNI-11, avaliada entre as classes NC x AD, pelo classificador SVM, apresentou a métrica B_{ca} de 99,88% e AUC de 1.000. Nas mesmas condições de avaliação, a base AIBL-4 apresentou a métrica B_{ca} de 96,76% e a base OASIS-5 apresentou a métrica B_{ca} de 88,04%.

Na avaliação com outros classificadores, além do SVM e do kNN, as Tabelas 5.10 e 5.11 mostram que o classificador SVM manteve bons resultados diante dos demais classificadores, com exceção para o Gradient Boosting. Em parte, isso se explica porque as bases AIBL-4 e ADNI-11 foram obtidas a partir do algoritmo FMFS, porém ajustadas para o classificador SVM. Os resultados foram diferentes para a base de dados OASIS-5, que apresentou o melhor

resultado para o classificador *decision tree*.

O algoritmo FMFS mostrou resultados comparáveis com outros 12 tradicionais métodos de seleção de atributos, como mostram os resultados apresentados na Tabela 5.13, Tabela 5.14 e Tabela 5.15.

5.4 DISCUSSÃO DOS RESULTADOS DA ANÁLISE SOBRE AS IMAGENS

O banco de dados primário *ADNI1:Screening 1.5T* possui 307 pacientes com diagnóstico de CN e 243 pacientes com diagnóstico de AD, conforme explicitado na Tabela 4.1. Esses números mostram por si só que há um desafio em utilizar uma quantidade tão reduzida de MRIs para se produzir uma análise de diagnóstico com base em imagens e usando técnicas de DL, tendo em vista que maiores volumes de dados facilitam o aprendizado das redes neurais profundas.

Os dados de MRI foram disponibilizados em formato NIFTI 3D, com disponibilidade de imagens no plano axial, coronal e sagital. A região de interesse de análise para determinação do diagnóstico de AD é a região do lobo temporal medial. No lobo temporal mesial se localiza o hipocampo, que é uma estrutura duplicada, espelhada na linha mediana do corpo, responsável pela memória, um dos aspectos mais afetados pela AD. Assim, não é razoável se utilizar de imagens no plano sagital, que mostraria apenas o hipocampo esquerdo ou direito a cada imagem. Dessa forma, foram escolhidos os planos coronal e axial para análise.

Todos os bancos de dados de imagem coronal e axial gerados neste trabalho foram utilizados para treinamento das redes propostas, CNN-1, CNN-2, CNN-3, CNN-4 e CNN-5. A rede CNN-1 (Figura 4.25) possui dez camadas *convolutional*, cinco camadas *max pooling*, sete camadas *batch normalization*, cinco camadas *dropout*, uma camada *flatten*, e quatro camadas *dense*. A CNN-2 (Figura 4.26) é uma versão da CNN-1 com redução da dimensionalidade das camadas. A CNN-3 (Figura 4.27) é uma versão da CNN-1 com aumento da dimensão das camadas. A CNN-4 (Figura 4.28) possui o acréscimo de camadas com relação a CNN-1, aumentando a profundidade, e a CNN-5 (Figura 4.29) possui as mesmas camadas da CNN-4, com redução das dimensionalidades das camadas. Também foram utilizados os modelos de redes pré-treinadas VGG-16, VGG-19, ResNet-50 e Inception V3 no processo de avaliação das bases de dados de imagem.

5.4.1 Plano coronal

Para o plano coronal foi possível gerar um banco de dados de imagens com dados de 226 pacientes em CN e 173 pacientes em AD. Portanto, a estratégia de escolha de imagens proposta neste trabalho fez o aproveitamento de 73,6% dos pacientes com diagnóstico CN e 71,2% dos pacientes com diagnóstico de AD da base de dados primária *ADNI1:Screening 1.5T*. Muitas foram as situações de ausência de dados ou de erros no tratamento do arquivo MRI, impossibilitando um aproveitamento maior da base de dados primária. O resultado é que se a quantidade de dados primários já é considerada pequena para ser utilizada em análise por modelos de DL, ficou ainda menor. Assim, foi necessário criar estratégias de aumento de dados.

A seleção das imagens de interesse foi feita com base na evidência da presença do hipocampo no *slice* escolhido. Então, a primeira estratégia de aumento de dados foi determinar uma imagem central e outras de *slices* adjacentes, tendo em vista que o hipocampo é uma estrutura que alcança alguns centímetros. Essa estratégia foi utilizada tanto no plano coronal, como no plano axial.

Então, foram montados dois bancos de imagens. O banco C1 que possui somente o *slice* central e resultou em 226 imagens de CN, e 173 imagens de AD. O banco C2 que possui *slice* central, adicionado de nove *slices* anteriores e nove *slices* posteriores, resultando em 4.294 imagens de CN e 3.287 imagens de AD.

Adicionalmente, foi proposta a montagem de duas partições diferentes para o processamento das imagens, combinando os bancos de dados C1 e C2, que resultou nas partições S1 e S2 (Tabela 5.23).

A partição coronal S1 (Tabela 5.23) foi avaliada por diversos modelos de DL (Tabela 5.17), sendo que a CNN-5 apresentou o melhor resultado para B_{ca} , com o valor de 94,79%. Para esse caso, foi significativo o aumento da profundidade da rede e a redução da dimensionalidade. Também foi feita avaliação com redes as pré-treinadas, VGG-19, VGG-16, ResNet-50 e Inception V3, com resultados inferiores a CNN-5. A quantidade de imagens tanto na partição S1 como na partição S2, não é suficiente para treinar redes tão profundas, como as rede pré-treinadas utilizadas.

A partição coronal S2 foi avaliada pelos modelos CNN-1, CNN-2, CNN-3, CNN-4 e CNN-5.

A CNN-4 apresentou o melhor resultado de B_{ca} , tendo atingido o valor de 97,83%. Essa foi a melhor métrica atingida na classificação de AD por imagens. Assim como a CNN-5, a CNN-4 é uma rede de maior profundidade, mas com dimensionalidade aumentada.

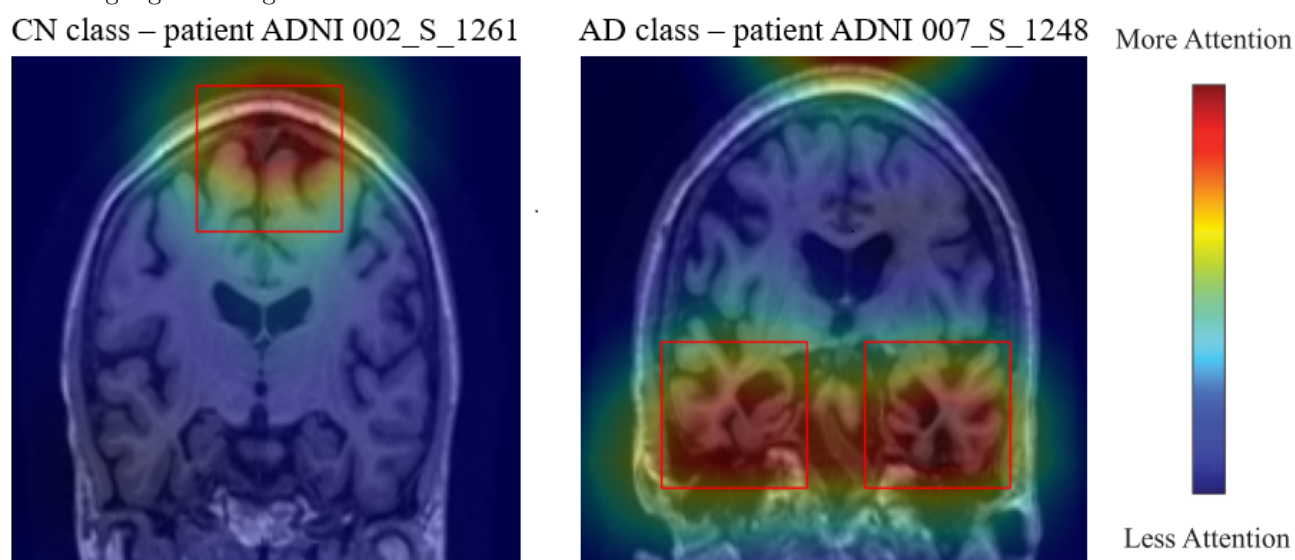
Também foi experimentada a técnica de aumento de dados clássica no tratamento de banco de imagem na partição coronal S2. Um aumento de dados de 2x sobre a partição de treino, usando uma faixa de variação aleatória de [0.99, 1.01] no zoom, faixa de variação aleatória de [0.8, 1.2] no brilho e inversão aleatória horizontal, foi experimentado. Os resultados não mostram evidências de melhora sobre as métricas. Para esse banco de dados e nas condições de avaliação em que foram realizadas, é possível afirmar que a estratégia de realizar aumento de dados utilizando os *slices* laterais foi mais eficiente que utilizar as técnicas clássicas baseadas em zoom, brilho e inversão horizontal.

Em (LIN *et al.*, 2013) foi proposta uma nova rede profunda denominada "Network In Network"(NIN) para tarefas de classificação. Esta nova estrutura consiste em utilizar uma camada *global average pooling* (GAP) como substituto das camadas totalmente conectadas nas CNNs convencionais. Segundo os autores, a camada GAP atua como um regularizador estrutural que impede o sobreajuste global da rede. Adicionalmente, a visualização dos *class activation maps* (CAM) da última camada convolucional da NIN possibilitam a realização de detecção de objetos na imagem de entrada.

A utilização da camada GAP para localização de objetos foi feita em (ZHOU *et al.*, 2015), que concluiu que a rede foi capaz de localizar as regiões discriminativas da imagem. A CAM destaca regiões específicas na imagem em função da classe decidida.

Utilizando estes conceitos, criamos uma rede VGG-16-GAP, acrescentando a camada GAP à uma rede VGG-16, treinamos toda a rede usando as imagens da partição S1 dos dados coronais,, e passamos a avaliar as regiões destacadas pela CAM em função de cada classe decidida pela rede. A Figura 5.14 mostra o *heat map* de dois diferentes pacientes com diagnóstico de CN e AD.

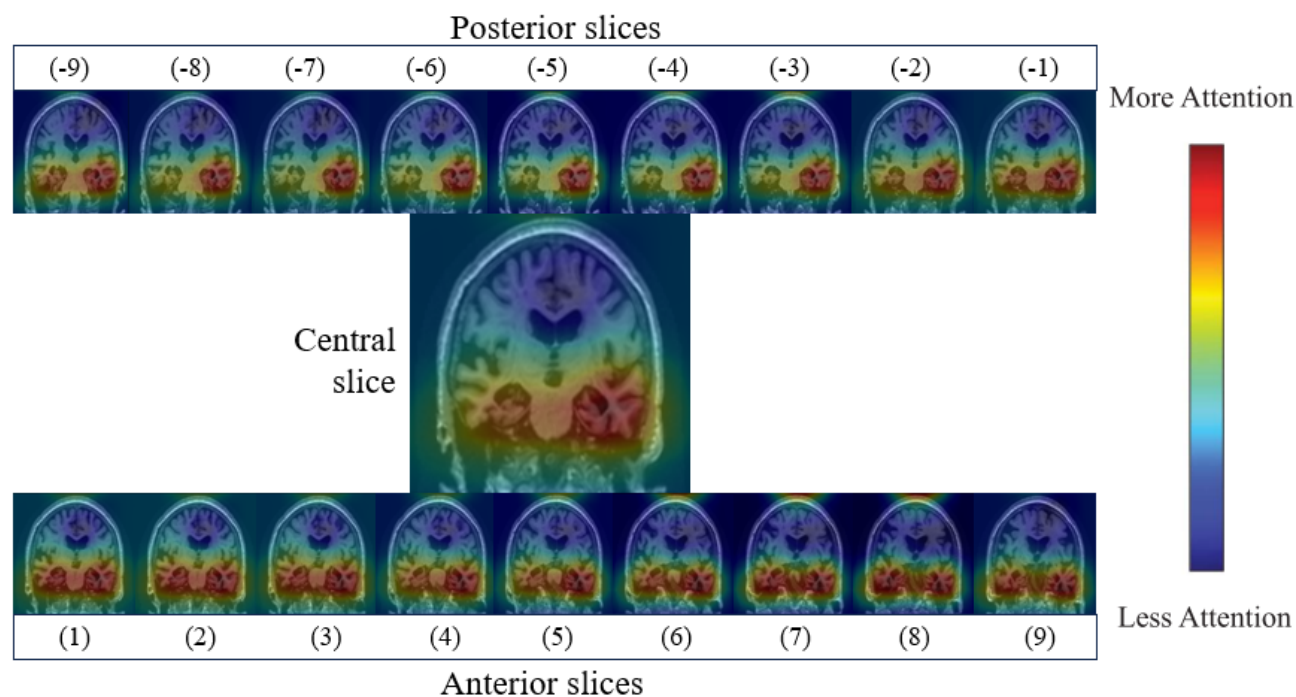
Figura 5.14. CN ADNI 002_S_1261 patient heat map and AD ADNI 007_S_1248 patient heat map. Red boxes highlight the region of interest.



Source: Own authorship.

Nestes casos, mostrados na Figura 5.14, fica evidente que a rede dá destaque à região do hipocampo no paciente com diagnóstico de AD, o que não acontece para o paciente com diagnóstico de CN. A atrofia cortical é uma alteração inespecífica da AD, pois está presente em diversas doenças neurodegenerativas (CIEPEL *et al.*, 2008). Ela pode ser observada em diversos *slices* do grupo AD, mas não foi levada em consideração pela rede neural. Isto demonstra a capacidade da rede neural em identificar padrões mais específicos de demência associada à AD.

Também foram avaliados os *heat maps* para cada uma das 19 imagens do paciente ADNI 007_S_1248, como mostra a Figura 5.15.

Figura 5.15. Heat map of the 19 images of patient ADNI 007_S_1248 diagnosed with AD.

Source: Own authorship.

Nota-se que o *heat map* dos 19 *slices* escolhidos para o paciente ADNI 007_S_1248 possuem uma certa semelhança, ainda que seja perceptível que há maior destaque na região do hipocampo direito nos *slices* posteriores, caminhando para o equilíbrio à medida que se avança na direção dos *slices* anteriores. Neste caso, todos os *slices* foram individualmente decididos como sendo com diagnóstico de AD. A escolha do quantitativo de 19 *slices* por paciente parece ter sido acertada, tendo em vista que a rede conseguiu localizar a região do hipocampo em todos os *slices* do paciente ADNI 007_S_1248.

Os resultados apresentados pela Tabela 5.20 e Tabela 5.21 mostram que a estratégia de utilizar múltiplas instâncias e classificador por maioria é uma boa alternativa em comparação com a decisão em única instância. Além de conquistar melhores valores de métricas, a decisão por maioria mostrou maior robustez na escolha do modelo de CNN.

5.4.2 Plano axial

Para o plano axial foi possível gerar um banco de dados de imagens com dados de 213 pacientes em CN e 184 pacientes em AD. Portanto, a estratégia de escolha de imagens proposta

neste trabalho fez o aproveitamento de 69,4% dos pacientes com diagnóstico CN e 75,7% dos pacientes com diagnóstico de AD da base de dados primária *ADNI1:Screening 1.5T*. Muitas foram as situações de ausência de dados ou de erros no tratamento do arquivo MRI, impossibilitando um aproveitamento maior da base de dados primária.

O banco de imagens gerado para o plano axial resultou em 1.656 imagens de AD e 1.917 imagens de CN e foi montada uma partição única, como pode ser verificado na Tabela 5.23. Diferente do plano coronal onde foi feito o aumento de dados de 19 imagens por paciente, no plano axial foi feito somente o aumento de dados por nove imagens por paciente.

A avaliação do banco de imagens no plano axial pelas redes CNN-1, CNN-2, CNN-2, CNN-3, CNN-4 e CNN-5, produziu métricas de B_{ca} inferiores a 61%. Nos experimentos realizados neste trabalho, é possível afirmar que as imagens no plano coronal resultaram em avaliações muito superiores às realizadas pelas imagens no plano axial.

O banco de imagens axial, gerado nesses experimentos, resultou em apenas 1.656 imagens de AD e 1.917 imagens CN, enquanto a partição S2 do banco de imagens coronal resultou em 2.600 imagens de AD e 3.466 imagens CN. A menor dimensionalidade do banco de imagens axial pode ter contribuído para o menor desempenho na avaliação. Também é razoável supor que a seleção de *slices* no plano axial não gerou as melhores escolhas na determinação das áreas de interesse.

5.4.3 Comparação com outras pesquisas

Os melhores resultados obtidos neste trabalho foram comparados com os melhores resultados em pesquisas anteriores para a classificação entre CN e AD, que utilizaram imagens de MRI oriundas do estudo ADNI, juntamente com alguma técnica de DL. A Tabela 5.28 faz o resumo da comparação dos resultados deste estudo com alguns outros estudos publicados entre 2020 e 2022.

Tabela 5.28. Comparison with previous DL studies

Reference	Dataset	B_{ca}	A_{cy}	P_{rn}	S_{yn}	S_{py}	F1
Present study	ADNI	99.2%	99.3%	99.2%	99.6%	98.9%	99.3%
(HAN <i>et al.</i> , 2020)	ADNI	NA	98,67%	99,66%	98,18%	99,27%	99,46%
(LI <i>et al.</i> , 2022)	ADNI	NA	93,16%	75,00%	83,18%	78,59%	77,38%
(RAZZAK <i>et al.</i> , 2022)	ADNI	NA	87,92%	NA	NA	NA	NA
(HAZARIKA <i>et al.</i> , 2022)	ADNI	NA	97%	96%	NA	NA	97%
(WU <i>et al.</i> , 2022)	ADNI	NA	91,31%	93,8%	88,3%	94,2	91,0%
(SHARMA <i>et al.</i> , 2022)	ADNI	NA	97,33%	NA	100%	97%	NA
(ZHANG <i>et al.</i> , 2023)	AIBL	NA	93,4%	NA	NA	NA	NA
(GAO <i>et al.</i> , 2023a)	ADNI	NA	90,5%	NA	84,6%	95,0%	NA
(GAO <i>et al.</i> , 2023b)	ADNI	NA	90,4%	NA	90,3%	90,5%	NA

NA - not available.

Source: Own authorship.

A exatidão alcançada por este estudo foi melhor do que a dos outros estudos de 2020 a 2023 apresentados na Tabela 5.28. Todos os estudos apresentados tentaram distinguir entre NC e DA, utilizaram imagens de RM dos conjuntos de dados ADNI e AIBL e empregaram técnicas de DL. É preciso considerar que estes estudos comparados podem ter selecionado conjuntos de dados das muitas opções disponíveis na base de dados ADNI que diferem deste estudo, tornando difícil a comparação direta.

5.5 SUGESTÃO DE APLICAÇÃO DAS NOVAS TECNOLOGIAS

Como visto no Capítulo 4, o relatório da Alzheimers Disease International (Gauthier *et al.*, 2021) afirma que 75% das pessoas com demência, por algum motivo, não estão sendo diagnosticadas. Do ponto de vista tecnológico, pode-se afirmar que as técnicas apresentadas neste estudo podem somar na construção de um cenário de ferramental tecnológico útil para o auxílio ao diagnóstico da AD, com potencial para ajudar milhões de pessoas que podem estar sob suspeita dessa doença.

Além dos estudos acadêmicos, alguns revisados neste trabalho, há também plataformas tecnológicas que utilizam tecnologia de ML de última geração para construir ferramentas inovadoras para análise quantitativa automática de imagens médicas, como é o caso do projeto

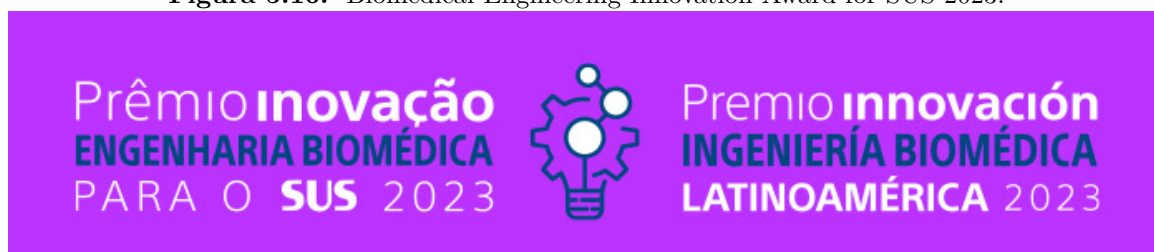
"InnerEye"¹, da Microsoft Research®, que tem apoio do National Health System da Inglaterra.

5.5.1 Sistema Único de Saúde brasileiro - SUS

Segundo definição do portal do Ministério de Saúde brasileiro², o Sistema Único de Saúde (SUS) é um dos maiores e mais complexos sistemas de saúde pública do mundo, abrangendo desde o simples atendimento para avaliação da pressão arterial, por meio da Atenção Primária, até o transplante de órgãos, garantindo acesso integral, universal e gratuito para toda a população do país. Com a sua criação, o SUS proporcionou o acesso universal ao sistema público de saúde, sem discriminação. A atenção integral à saúde, e não somente aos cuidados assistenciais, passou a ser um direito de todos os brasileiros, desde a gestação e por toda a vida, com foco na saúde com qualidade de vida, visando a prevenção e a promoção da saúde.

5.5.2 Premio Inovação Engenharia Biomédica para o SUS 2023

Figura 5.16. Biomedical Engineering Innovation Award for SUS 2023.



Source: Adapted from <https://www.bostonscientific.com/pt-BR/inovacao/premio-de-inovacao-para-o-sus.html>.

Este concurso visa reconhecer e premiar trabalhos técnico-científicos na área da Engenharia Biomédica, com temática relacionada a soluções para o sistema público de saúde, propondo soluções práticas para os problemas enfrentados pelo Sistema Único de Saúde - SUS.

Segundo o edital deste concurso³, as propostas serão avaliadas por membros da:

1. Sociedade Brasileira de Engenharia Biomédica SBEB;

¹<https://www.microsoft.com/en-us/research/project/medical-image-analysis/>

²<https://www.gov.br/saude/pt-br/assuntos/saude-de-a-a-z/s/sus>

³https://sbeb.org.br/wp-content/uploads/2023/06/EDITAL_SBEB_PREMIO_2023_V2-1.pdf

2. Ministério da Saúde MS;
3. Agência Nacional de Vigilância Sanitária ANVISA;
4. Organização Pan-Americana da Saúde OPAS;
5. Federação Internacional de Engenharia Médica e Biológica IFMBE.

5.5.3 Proposta vitoriosa

Este trabalho foi submetido ao Premio e foi reconhecido como vitorioso, tendo sido avaliado sob os seguintes aspectos:

1. INOVAÇÃO - Potencial de inovação do projeto para o SUS;
2. APLICABILIDADE AO SUS - O quanto o projeto pode ser aplicado e impactar o SUS, considerando suas políticas, pilares basilares e demandas da população;
3. QUALIDADE DE VIDA E AMPLITUDE POPULACIONAL - Impacto do projeto na qualidade de vida da população usuária do SUS e no aumento da população atendida;
4. PRAZO PARA BENEFÍCIO - Grau de maturidade do projeto. Tempo necessário para que os benefícios para o SUS sejam alcançados;
5. CUSTO/BENEFÍCIO - Quantidade e complexidade de recursos necessários (tecnológicos, financeiros, humanos, tempo) para consecução/execução do projeto frente os benefícios para o SUS e para a sociedade;
6. IMPACTOS: SOCIAL, TECNOLÓGICO, CIENTÍFICO - Sob a perspectiva da sociedade brasileira, o potencial do projeto em termos de geração de novos avanços científicos, futura geração de empregos, exportações e do fortalecimento do Complexo Industrial de Saúde Brasileiro, diminuindo a dependência do SUS do mercado externo (importações).

Este capítulo apresentou os resultados obtidos pelas análises clínicas e os resultados obtidos pela avaliação de imagens de ressonância magnética. Também foi feita a discussão sobre os resultados e a comparação com resultados atingidos por outros pesquisadores. No próximo capítulo, serão apresentadas as conclusões e as contribuições deste trabalho.

CONCLUSÕES

A utilização de técnicas de inteligência computacional sobre dados clínicos mostrou-se eficiente no auxílio à determinação do diagnóstico da doença de Alzheimer. Este trabalho fez uso de dados clínicos de três renomados estudos conduzidos por universidades e institutos de pesquisa mundiais, realizou diversos experimentos exploratórios e propôs novos recursos tecnológicos que melhor se adequassem aos dados na determinação do diagnóstico, tendo atingido 99,88% de acurácia na previsão da doença, utilizando dados do estudo *Alzheimers Disease Neuroimaging Initiative*, que é a base de dados mais usada na literatura.

Na construção de uma alternativa tecnológica para conduzir os trabalhos sobre os dados clínicos, foi proposto um novo algoritmo de seleção de atributos que apresentou resultados promissores, utilizando uma abordagem simplificada, porém tendo produzido resultados superiores quando comparados com os resultados de outros estudos referenciados na literatura.

É também contribuição do algoritmo de seleção de atributos proposto a redução de dimensionalidade das bases de dados, a redução do tempo de treinamento dos modelos de aprendizado e o realce dos atributos que mais influenciam no diagnóstico da doença de Alzheimer, cujo diagnóstico preciso continua desafiando a ciência.

Também foram aplicadas técnicas de inteligência computacional sobre dados de imagem, obtidas por ressonância magnética. Foram propostas estratégias para a construção de bases de dados de imagem sobre os planos coronal e axial. Estratégias de tratamento das imagens foram realizadas e foi possível atingir 99,3% de acurácia na previsão da doença, utilizando dados de ressonância magnética do estudo *Alzheimers Disease Neuroimaging Initiative*, na versão *ADNI1:Screening 1.5T*.

Dada a necessidade de grandes volumes de dados para processamento dos modelos de aprendizado e a baixa quantidade de ressonâncias magnéticas, foi feita uma proposta de aumento de dados e comparada com uma alternativa clássica de aumento de dados. Para esse caso estudado

o aumento de dados proposto se mostrou mais vantajoso. Também para as condições estudadas neste trabalho, a utilização de imagens no plano coronal apresentou melhor resultado que no plano axial.

Foram também propostas algumas arquiteturas de redes convolucionais, variando a dimensionalidade e a profundidade das camadas, no sentido de melhor encontrar ferramentas computacionais que ajudassem na melhoria dos estimadores para o diagnóstico da doença de Alzheimer.

Os resultados obtidos neste trabalho indicam que a direção tomada por este estudo é promissora e deixa contribuições para o aprimoramento do diagnóstico de Alzheimer baseado em técnicas de inteligência computacional. Poderão ser realizadas novas investigações para automatizar a seleção do *slice* central no plano coronal, em vez da seleção manual realizada neste estudo.

REFERÊNCIAS BIBLIOGRÁFICAS

- ALBRIGHT, J. *Forecasting the Progression of Alzheimer's Disease Using Neural Networks and a Novel Pre-Processing Algorithm*. arXiv, 2019. Disponível em: <https://arxiv.org/abs/1903.07510>. Cited 5 times in pages 15, 17, 18, 107, and 108.
- ALTMANN, A.; TOLOȘI, L.; SANDER, O.; LENGAUER, T. Permutation importance: a corrected feature importance measure. *Bioinformatics*, Oxford University Press, v. 26, n. 10, p. 1340–1347, 2010. Cited in page 30.
- Alzheimers Disease International. World alzheimer report 2019. *Attitudes to dementia*. London: Alzheimers Disease International, 2019. Cited in page 10.
- BALAKRISHNAMA, S.; GANAPATHIRAJU, A. Linear discriminant analysis - a brief tutorial. *Institute for Signal and Information Processing*, 1998. Cited in page 51.
- BALESTRIERI J., N. M. B. G. L. . P. M. N. Structural volume of hippocampus and alzheimer's disease. *Revista da Associacao Medica Brasileira*, Vol. 66(4), p. 512515, 2020. Cited 2 times in pages 2 and 3.
- BLEICH, H. L. Computer evaluation of acid-base disorders. *The Journal of clinical investigation*, Vol. 48,9, p. 1689–1696, 1969. Cited in page 13.
- BREIMAN, L. Random forests. *Machine Learning*, Kluwer Academic Publishers, v. 45, n. 1, p. 5–32, 2001. ISSN 0885-6125. Disponível em: <http://dx.doi.org/10.1023/A%3A1010933404324>. Cited in page 51.
- BRITO, E. N. D.; FIGUEIREDO, B. Q. d.; SOUTO, D. N.; NOGUEIRA, J. F.; MELO, A. L. d. C.; SILVA, I. T. d.; OLIVEIRA, I. P.; ALMEIDA, M. G. d. Artificial intelligence in the diagnosis of neurodegenerative diseases: a systematic literature review. *Research, Society and Development*, v. 10, n. 11, p. e482101120004, Sep. 2021. Disponível em: <https://rsdjournal.org/index.php/rsd/article/view/20004>. Cited in page 15.
- CAWLEY, G. C.; TALBOT, N. L. On over-fitting in model selection and subsequent selection bias in performance evaluation. *J. Mach. Learn. Res.*, JMLR.org, v. 11, p. 20792107, aug 2010. ISSN 1532-4435. Cited in page 14.
- CENTRAL, D. C. *Stages of Alzheimers Dementia: Durations Scales Used to Measure Progression (GDS, FAST CDR)*. 2022. URL: <https://www.dementiacarecentral.com/aboutdementia/facts/stages/>. Cited in page 6.
- CHAWLA, N. V.; BOWYER, K. W.; HALL, L. O.; KEGELMEYER, W. P. SMOTE: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, AI Access Foundation, v. 16, p. 321–357, jun 2002. Disponível em: <https://doi.org/10.1613%2Fjair.953>. Cited in page 28.

- CHEN, T.; GUESTRIN, C. XGBoost. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 2016. Disponível em: <https://doi.org/10.1145%2F2939672.2939785>. Cited in page 51.
- CORTES, C.; VAPNIK, V. Support vector networks. *Machine Learning*, v. 20, p. 273–297, 1995. Cited 2 times in pages 46 and 48.
- COVER, T.; HART, P. Nearest neighbor pattern classification. *IEEE Transactions on Information Theory*, v. 13, n. 1, p. 21–27, 1967. Cited in page 49.
- CUI, Y.; JIA, M.; LIN, T.-Y.; SONG, Y.; BELONGIE, S. *Class-Balanced Loss Based on Effective Number of Samples*. 2019. Cited in page 28.
- DEKEYZER, S.; BOSSCHE, S. V. "unforgettable- a pictorial essay on anatomy and pathology of the hippocampus. *Insights into Imaging*, v. 8, 01 2017. Cited 2 times in pages 61 and 62.
- DHAL, P.; AZAD, C. A comprehensive survey on feature selection in the various fields of machine learning. *Applied Intelligence*, Springer, p. 1–39, 2022. Cited 2 times in pages 28 and 29.
- DING, H.; FENG, P.-M.; CHEN, W.; LIN, H. Identification of bacteriophage virion proteins by the anova feature selection and analysis. *Molecular BioSystems*, Royal Society of Chemistry, v. 10, n. 8, p. 2229–2235, 2014. Cited in page 29.
- ELLIS, K.; BUSH, A.; DARBY, D.; FAZIO, D.; FOSTER, J.; HUDSON, P.; LAUTENSCHLAGER, N.; LENZO, N.; MARTINS, R.; MARUFF, P.; MASTERS, C.; MILNER, A.; PIKE, K.; ROWE, C.; SAVAGE, G.; SZOEKE, C.; TADDEI, K.; VILLEMAGNE, V.; WOODWARD, M.; AMES, D. The australian imaging, biomarkers and lifestyle (aibl) study of aging: Methodology and baseline characteristics of 1112 individuals recruited for a longitudinal study of alzheimers disease. *International psychogeriatrics / IPA*, v. 21, p. 672–87, 06 2009. Cited 3 times in pages 6, 22, and 24.
- FEDOROV, A. Y.; BEICHEL, R. R.; KALPATHY-CRAMER, J.; FINET, J.; FILLION-ROBIN, J.-C.; PUJOL, S.; BAUER, C.; JENNINGS, D. L.; FENNESSY, F. M.; SONKA, M.; BUATTI, J. M.; AYLWARD, S. R.; MILLER, J. V.; PIEPER, S. D.; KIKINIS, R. 3d slicer as an image computing platform for the quantitative imaging network. *Magnetic resonance imaging*, v. 30 9, p. 1323–41, 2012. Cited in page 58.
- FREUND, Y.; MASON, L. The alternating decision tree learning algorithm. In: *ICML*. [s.n.], 1999. p. 124–133. Disponível em: <http://www.lsmason.com/papers/ICML99-AlternatingTrees.pdf>. Cited in page 51.
- FRIEDMAN, J. H. Greedy function approximation: A gradient boosting machine. *The Annals of Statistics*, Institute of Mathematical Statistics, v. 29, n. 5, p. 1189–1232, 2001. ISSN 00905364. Disponível em: <http://www.jstor.org/stable/2699986>. Cited in page 61.
- GAO, X.; CAI, H.; LIU, M. A hybrid multi-scale attention convolution and aging transformer network for alzheimer’s disease diagnosis. *IEEE Journal of Biomedical and Health Informatics*, IEEE, 2023. Cited 3 times in pages 17, 20, and 116.
- GAO, X.; LIU, H.; SHI, F.; SHEN, D.; LIU, M. Brain status transferring generative adversarial network for decoding individualized atrophy in alzheimer’s disease. *IEEE Journal of Biomedical and Health Informatics*, IEEE, 2023. Cited 3 times in pages 17, 20, and 116.

- Gauthier S.; P., R.-N.; JA, M.; C., W. World alzheimer report 2021. *Journey through the diagnosis of dementia*. London, England: *Alzheimers Disease International*, 2021. Cited 2 times in pages 10 and 16.
- GHAZI, M. M.; NIELSEN, M.; PAI, A.; CARDOSO, M. J.; MODAT, M.; OURSELIN, S.; SØRENSEN, L. Training recurrent neural networks robust to incomplete data: Application to alzheimer's disease progression modeling. *Medical Image Analysis*, Elsevier BV, v. 53, p. 39–46, apr 2019. Disponível em: <<https://doi.org/10.1016%2Fj.media.2019.01.004>>. Cited 5 times in pages 15, 17, 19, 107, and 108.
- GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. *Deep Learning*. [S.l.]: MIT Press, 2016. <<http://www.deeplearningbook.org>>. Cited 2 times in pages 13 and 14.
- GOUBRAN, M.; NTIRI, E. E.; AKHAVEIN, H.; HOLMES, M.; NESTOR, S.; RAMIREZ, J.; ADAMO, S.; OZZOUDE, M.; SCOTT, C.; GAO, F. *et al. Hippocampal segmentation for brains with extensive atrophy using three-dimensional convolutional neural networks*. [S.l.], 2020. Cited in page 54.
- HAN, R.; CHEN, C. L. P.; LIU, Z. A novel convolutional variation of broad learning system for alzheimers disease diagnosis by using mri images. *IEEE Access*, v. 8, p. 214646–214657, 2020. Cited 3 times in pages 17, 19, and 116.
- HART, P. E.; STORK, D. G.; DUDA, R. O. *Pattern classification*. [S.l.]: Wiley Hoboken, 2000. Cited in page 29.
- HAZARIKA, R. A.; KANDAR, D.; MAJI, A. K. A deep convolutional neural networks based approach for alzheimers disease and mild cognitive impairment classification using brain images. *IEEE Access*, v. 10, p. 99066–99076, 2022. Cited 3 times in pages 17, 19, and 116.
- HE, H.; BAI, Y.; GARCIA, E. A.; LI, S. Adasyn: Adaptive synthetic sampling approach for imbalanced learning. p. 1322–1328, 2008. Cited in page 28.
- HE, K.; ZHANG, X.; REN, S.; SUN, J. *Deep Residual Learning for Image Recognition*. arXiv, 2015. Disponível em: <<https://arxiv.org/abs/1512.03385>>. Cited in page 75.
- HOQUE, N.; BHATTACHARYYA, D.; KALITA, J. Mifs-nd: A mutual information-based feature selection method. *Expert Systems with Applications*, v. 41, n. 14, p. 6371–6385, 2014. ISSN 0957-4174. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0957417414002164>>. Cited in page 29.
- HUGHES C. P., B. L. D. W. L. C. L. A. . M. R. L. A new clinical scale for the staging of dementia. *The British journal of psychiatry : the journal of mental science*, Vol. 140, p. 566572, 1982. Cited 2 times in pages 1 and 22.
- IDDI, S.; D, L.; PS, A.; MS, R.; WK, T.; MC, D. Predicting the course of alzheimers progression. *Brain Informatics*, Vol. 6(1), 2019. Cited 5 times in pages 15, 17, 18, 107, and 108.
- JIN, X.; XU, A.; BIE, R.; GUO, P. Machine learning techniques and chi-square feature selection for cancer classification using sage gene expression profiles. In: SPRINGER. *Data Mining for Biomedical Applications: PAKDD 2006 Workshop, BioDM 2006, Singapore, April 9, 2006. Proceedings*. [S.l.], 2006. p. 106–115. Cited in page 29.

- KHAN, P.; KADER, M. F.; ISLAM, S. M. R.; RAHMAN, A. B.; KAMAL, M. S.; TOHA, M. U.; KWAK, K.-S. Machine learning and deep learning approaches for brain disease diagnosis: Principles and recent advances. *IEEE Access*, v. 9, p. 37622–37655, 2021. Cited 6 times in pages [15](#), [16](#), [17](#), [21](#), [107](#), and [108](#).
- KNOPMAN D. S., P. R. C. . J. C. R. J. A brief history of "alzheimer disease". *Neurology*, Vol. 92(22), p. 10531059, 2019. Cited in page [6](#).
- KRIZHEVSKY, A.; SUTSKEVER, I.; HINTON, G. E. Imagenet classification with deep convolutional neural networks. In: PEREIRA, F.; BURGESS, C.; BOTTOU, L.; WEINBERGER, K. (Ed.). *Advances in Neural Information Processing Systems*. Curran Associates, Inc., 2012. v. 25. Disponível em: <https://proceedings.neurips.cc/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf>. Cited in page [13](#).
- KURSA, M. B.; RUDNICKI, W. R. Feature selection with the boruta package. *Journal of statistical software*, v. 36, p. 1–13, 2010. Cited in page [30](#).
- LAROBINA, M.; MURINO, L. Medical image file formats. *Journal of digital imaging*, 12 2013. Cited in page [58](#).
- LAVALLEY, M. Logistic regression. *Circulation*, v. 117, p. 2395–9, 06 2008. Cited in page [51](#).
- LECUN, Y.; BENGIO, Y.; HINTON, G. Deep learning. *Nature*, v. 521, p. 436–44, 05 2015. Cited in page [74](#).
- LECUN, Y.; BOTTOU, L.; BENGIO, Y.; HAFFNER, P. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, v. 86, n. 11, p. 2278–2324, 1998. Cited in page [75](#).
- LEDIG, C.; SCHUH, A.; GUERRERO, R.; HECKEMANN, R. A.; RUECKERT, D. Structural brain imaging in alzheimers disease and mild cognitive impairment: biomarker analysis and shared morphometry database. *Scientific reports*, Nature Publishing Group, v. 8, n. 1, p. 1–16, 2018. Cited 2 times in pages [53](#) and [54](#).
- LI, J.; WEI, Y.; WANG, C.; HU, Q.; LIU, Y.; XU, L. 3-d cnn-based multichannel contrastive learning for alzheimers disease automatic diagnosis. *IEEE Transactions on Instrumentation and Measurement*, v. 71, p. 1–11, 2022. Cited 3 times in pages [17](#), [19](#), and [116](#).
- LI, X.; MORGAN, P. S.; ASHBURNER, J.; SMITH, J.; RORDEN, C. The first step for neuroimaging data analysis: Dicom to nifti conversion. *Journal of Neuroscience Methods*, v. 264, p. 47–56, 2016. ISSN 0165-0270. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0165027016300073>. Cited in page [58](#).
- LIN, M.; CHEN, Q.; YAN, S. Network in network. *CoRR*, abs/1312.4400, 2013. Disponível em: <https://api.semanticscholar.org/CorpusID:16636683>. Cited in page [112](#).
- LIU J LI, P. C. X. Modeling disease progression with deep neural networks. *The Fourth International Symposium on Image Computing and Digital Medicine (ISICDM 2020)*, p. 32–34, 2020. Cited 4 times in pages [17](#), [19](#), [107](#), and [108](#).
- LUNDBERG, S. M.; LEE, S.-I. A unified approach to interpreting model predictions. *Advances in neural information processing systems*, v. 30, 2017. Cited in page [30](#).

- MARCUS, D. S.; WANG, T. H.; PARKER, J.; CSERNANSKY, J. G.; MORRIS, J. C.; BUCKNER, R. L. Open Access Series of Imaging Studies (OASIS): Cross-sectional MRI Data in Young, Middle Aged, Nondemented, and Demented Older Adults. *Journal of Cognitive Neuroscience*, v. 19, n. 9, p. 1498–1507, 09 2007. ISSN 0898-929X. Disponível em: <<https://doi.org/10.1162/jocn.2007.19.9.1498>>. Cited 3 times in pages 6, 21, and 24.
- MARINESCU, R. V.; OXTOBY, N. P.; YOUNG, A. L.; BRON, E. E.; TOGA, A. W.; WEINER, M. W.; BARKHOF, F.; FOX, N. C.; KLEIN, S.; ALEXANDER, D. C.; CONSORTIUM, t. E. Tadpole challenge: Prediction of longitudinal evolution in alzheimers disease. *F1000Research*, 2018. Cited 4 times in pages 6, 23, 24, and 59.
- MELBOURNE, T. U. of. *QUANTUM BOOST FOR MEDICAL IMAGING*. 2018. URL: <https://pursuit.unimelb.edu.au/articles/quantum-boost-for-medical-imaging>. Cited 2 times in pages 56 and 57.
- MILLER-THOMAS, M. M.; SIPE, A. L.; BENZINGER, T. L. S.; MCCONATHY, J.; CONNOLLY, S.; SCHWETYE, K. E. Multimodality review of amyloid-related diseases of the central nervous system. *RadioGraphics*, v. 36, n. 4, p. 1147–1163, 2016. PMID: 27399239. Disponível em: <<https://doi.org/10.1148/rg.2016150172>>. Cited 2 times in pages 52 and 61.
- MOORE, P. J.; LYONS, T. J.; GALLACHER, J.; INITIATIVE, A. D. N. Random forest prediction of alzheimers disease using pairwise selection from time series data. *Plos One*, Vol. 14(2), 2019. Cited 5 times in pages 13, 17, 18, 107, and 108.
- NGUYEN, M.; SUN, N.; ALEXANDER, D. C.; FENG, J.; YEO, B. T. Modeling alzheimers disease progression using deep recurrent neural networks. p. 1–4, 2018. Cited 6 times in pages 13, 17, 19, 46, 107, and 108.
- NGUYEN M., H. T. A. L. A. D. C. F. J. Y. B. . A. D. N. I. Predicting alzheimers disease progression using deep recurrent neural networks. *NeuroImage*, Vol. 222, 2020. Cited 5 times in pages 13, 17, 18, 107, and 108.
- NIYAS, M.; P, T. Alzheimer’s classification using dynamic ensemble of classifiers selection algorithms: A performance analysis. *Biomedical Signal Processing and Control*, v. 68, p. 102729, 2021. ISSN 1746-8094. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S1746809421003268>>. Cited 4 times in pages 13, 17, 107, and 108.
- NIYAS, M.; P, T. Feature selection using efficient fusion of fisher score and greedy searching for alzheimers classification. *Journal of King Saud University - Computer and Information Sciences*, 2021. Cited 8 times in pages 13, 17, 21, 23, 31, 46, 107, and 108.
- NORFRAY, J. F.; PROVENZALE, J. M. Alzheimer’s disease: Neuropathologic findings and recent advances in imaging. *American Journal of Roentgenology*, v. 182, n. 1, p. 3–13, 2004. PMID: 14684506. Disponível em: <<https://doi.org/10.2214/ajr.182.1.1820003>>. Cited in page 52.
- ONU. World population prospects 2022: Summary of results. *United Nations Department of Economic and Social Affairs, Population Division*, UN DESA/POP/2022/TR/NO. 3, 2022. Cited in page 4.
- PAIVA, N.; ESCOVEDO, T. Detecção precoce de alzheimer usando machine learning. *XV Congresso Brasileiro de Inteligência Computacional*, 2021. Disponível em: <https://sbic.org.br/eventos/cbic_2021/>. Cited 3 times in pages 23, 108, and 109.

- PARMERA, J. B.; NITRINI, R. Demências: da investigação ao diagnóstico. *Revista de Medicina*, v. 94, n. 3, p. 179–184, dez. 2015. Disponível em: <https://www.revistas.usp.br/revistadc/article/view/108748>>. Cited in page 8.
- PARVANDEH, S.; YEH, H.-W.; PAULUS, M.; MCKINNEY, B. Consensus features nested cross-validation. *Bioinformatics (Oxford, England)*, v. 36, 01 2020. Cited 2 times in pages 44 and 45.
- PATEL, K. P.; WYMER, D. T.; BHATIA, V. K.; DUARA, R.; RAJADHYAKSHA, C. D. Multimodality imaging of dementia: Clinical importance and role of integrated anatomic and molecular imaging. *RadioGraphics*, v. 40, n. 1, p. 200–222, 2020. PMID: 31917652. Disponível em: <https://doi.org/10.1148/rg.2020190070>>. Cited in page 52.
- RASCHKA, S. *STAT 479: Machine Learning - Lecture Notes*. [S.l.]: Department of Statistics - University of Wisconsin, 2018. <http://stat.wisc.edu/~sraschka/teaching/stat479-fs2018/>>. Cited in page 50.
- RAZZAK, I.; NAZ, S.; ALINEJAD-ROKNY, H.; NGUYEN, T.; KHALIFA, F. A cascaded multiresolution ensemble deep learning framework for large scale alzheimer’s disease detection using brain mris. *IEEE/ACM transactions on computational biology and bioinformatics*, PP, 11 2022. Cited 3 times in pages 17, 19, and 116.
- REPORT, A. A. Alzheimer’s disease facts and figures. *Alzheimer’s dementia : the journal of the Alzheimer’s Association*, Vol. 17(3), p. 327–406, 2021. Cited in page 8.
- RISH, I. An empirical study of the naïve bayes classifier. *IJCAI 2001 Work Empir Methods Artif Intell*, v. 3, 01 2001. Cited in page 51.
- ROGACHESKI, .; MAZER, S.; RODRIGUES, D.; BITTENCOURT, P. R. M. D. Análises visual e volumétrica por ressonância magnética das formações hipocampais em um grupo de pacientes com diagnóstico clínico de epilepsia do lobo temporal. *Arquivos de Neuro-Psiquiatria*, Academia Brasileira de Neurologia - ABNEURO, v. 56, n. Arq. Neuro-Psiquiatr., 1998 56(3A), Sep 1998. ISSN 0004-282X. Disponível em: <https://doi.org/10.1590/S0004-282X1998000300011>>. Cited in page 53.
- SAEYS, Y.; ABEEL, T.; PEER, Y. Van de. Robust feature selection using ensemble feature selection techniques. In: SPRINGER. *Machine Learning and Knowledge Discovery in Databases: European Conference, ECML PKDD 2008, Antwerp, Belgium, September 15-19, 2008, Proceedings, Part II 19*. [S.l.], 2008. p. 313–325. Cited in page 30.
- SALIH, Y.; MD-ESA, W. bt.; MALIK, A. S.; SAAD, N. Tone mapping of hdr images: A review. In: *2012 4th International Conference on Intelligent and Advanced Systems (ICIAS2012)*. [S.l.: s.n.], 2012. v. 1, p. 368–373. Cited in page 69.
- SARAIVA JR., R. G.; CARVALHO, J. L. A.; PEIXOTO, E.; NASCIMENTO, F. A. de O. Seleção simplificada de atributos para auxílio ao diagnóstico da doença de alzheimer utilizando aprendizado de máquina. XXVIII Congresso Brasileiro de Engenharia de Engenharia Biomédica (CBEB 2022), 2022. Cited in page 108.
- SCHACHTER MD; KENNETH L. DAVIS, M. A. S. Alzheimers disease. *Dialogues in Clinical Neuroscience*, Vol. 2(2), p. 91–100, 2000. Cited in page 7.

- SCHAPIRE, R. E. Explaining adaboost. In: *Empirical inference*. [S.l.]: Springer, 2013. p. 37–52. Cited in page 51.
- SEJO-PARDO, B.; BOLÓN-CANEDO, V.; ALONSO-BETANZOS, A. On developing an automatic threshold applied to feature selection ensembles. *Information Fusion*, Elsevier, v. 45, p. 227–245, 2019. Cited in page 30.
- SHAFER, J. C.; AGRAWAL, R.; MEHTA, M. Sprint: A scalable parallel classifier for data mining. In: *Proceedings of the 22th International Conference on Very Large Data Bases*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1996. (VLDB '96), p. 544555. ISBN 1558603824. Cited in page 51.
- SHARMA, R.; GOEL, T.; TANVEER, M.; SUGANTHAN, P.; RAZZAK, I.; MURUGAN, R. Conv-ervfl: Convolutional neural network based ensemble rvfl classifier for alzheimer's disease diagnosis. *IEEE Journal of Biomedical and Health Informatics*, IEEE, 2022. Cited 3 times in pages 17, 20, and 16.
- SIMONYAN, K.; ZISSERMAN, A. *Very Deep Convolutional Networks for Large-Scale Image Recognition*. arXiv, 2014. Disponível em: <<https://arxiv.org/abs/1409.1556>>. Cited in page 75.
- SWAFFE, K. *What the hell happened to my brain? Living Beyond Dementia*. London: Jessica Kingsley Publishers, 2016. 391 p. ISBN 978-1-84905-6-083. Cited in page 1.
- SZEGEDY, C.; VANHOUCHE, V.; IOFFE, S.; SHLENS, J.; WOJNA, Z. *Rethinking the Inception Architecture for Computer Vision*. arXiv, 2015. Disponível em: <<https://arxiv.org/abs/1512.00567>>. Cited in page 75.
- TAHIRA, A. C.; VERJOVSKI-ALMEIDA, S.; FERREIRA, S. T. Dementia is an age-independent risk factor for severity and death in covid-19 inpatients. *Alzheimer's & Dementia*, v. 17, n. 11, p. 1818–1831, 2021. Disponível em: <<https://alz-journals.onlinelibrary.wiley.com/doi/abs/10.1002/alz.12352>>. Cited in page 9.
- TAN, M.; LE, Q. *EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks*. 2019. Cited in page 75.
- TEIPEL, S. J.; MEINDL, T.; GRINBERG, L.; HEINSEN, H.; HAMPEL, H. Novel mri techniques in the assessment of dementia. *European journal of nuclear medicine and molecular imaging*, Springer, v. 35, p. 58–69, 2008. Cited in page 13.
- URBANOWICZ, R. J.; MEEKER, M.; CAVA, W. L.; OLSON, R. S.; MOORE, J. H. Relief-based feature selection: Introduction and review. *Journal of biomedical informatics*, Elsevier, v. 85, p. 189–203, 2018. Cited in page 29.
- VARMA, S.; SIMON, R. Bias in error estimation when using cross-validation for model selection. *BMC Bioinformatics*, v. 7, 02 2006. Cited in page 45.
- WELLER J., B. A. Current understanding of alzheimers disease diagnosis and treatment. *F1000Research*, Vol. 7, 2018. Cited in page 1.
- WU, Y.; ZHOU, Y.; ZENG, W.; QIAN, Q.; SONG, M. An attention-based 3d cnn with multi-scale integration block for alzheimer's disease classification. *IEEE Journal of Biomedical and Health Informatics*, IEEE, v. 26, n. 11, p. 5665–5673, 2022. Cited 3 times in pages 17, 20, and 16.

YAMASHITA, R.; NISHIO, M.; DO, R.; TOGASHI, K. Convolutional neural networks: an overview and application in radiology. *Insights into Imaging*, v. 9, 06 2018. Cited in page [174](#).

YANG H. D., K. D. H. L. S. B. . Y. L. D. History of alzheimer's disease. *Dementia and neurocognitive disorders*, Vol. 15(4), p. 115121, 2016. Cited in page [6](#).

ZHANG, D.; SHEN, D.; INITIATIVE, A. D. N. Multi-modal multitask learning for joint prediction of multiple regression and classification variables in alzheimers disease. *NeuroImage*, Vol. 59(2), p. 895907, 2011. Cited 2 times in pages [15](#) and [17](#).

ZHANG, Z.-C.; ZHAO, X.; DONG, G.; ZHAO, X.-M. Improving alzheimer's disease diagnosis with multi-modal pet embedding features by a 3d multi-task mlp-mixer neural network. *IEEE Journal of Biomedical and Health Informatics*, IEEE, 2023. Cited 3 times in pages [17](#), [20](#), and [116](#).

ZHOU, B.; KHOSLA, A.; LAPEDRIZA, A.; OLIVA, A.; TORRALBA, A. *Learning Deep Features for Discriminative Localization*. 2015. Cited in page [112](#).