Programa de Doctorado en Empresa, Internet y Tecnologías de las Comunicaciones (EmITIC)

Doctoral Thesis

Intelligent computing-based systems for diagnosing Mild Cognitive Impairment and Alzheimer's Disease. Towards an e-Health solution

Tesis Doctoral

Sistemas basados en computación inteligente, de ayuda al diagnóstico del Deterioro Cognitivo Leve y la Enfermedad de Alzheimer. Hacia una solución de e-Salud

Ylermi Cabrera León

2025
Las Palmas de Gran Canaria, España

Resumen

El aumento de la población envejecida conlleva más personas con enfermedades no transmisibles asociadas al envejecimiento, como es la demencia, siendo la Enfermedad de Alzheimer (EA) su forma más prevalente. El estado cognitivo en el Deterioro Cognitivo Leve (DCL) está entre la EA y el cerebro sano [Petersen, 2004]. Afecta también a la memoria del paciente, pero no a sus actividades diarias.

Ambas enfermedades constituyen el mayor reto sociosanitario al que se enfrentan nuestras sociedades, siendo su diagnóstico y manejo, un problema complejo de la medicina clínica. De hecho, actualmente el diagnóstico *ante mortem* de la EA no es posible, considerándose únicamente el *post mortem* como certero. Como dichos diagnósticos carecen de fiabilidad y certeza, investigadores de muy diversos campos han trabajado conjuntamente para tratarlo. Una de las posibilidades ha sido el uso de herramientas de diagnóstico asistidas por computador, o Computer-Aided Diagnosis (CAD). Dentro de ellas ha emergido la alternativa basada en métodos computacionales, esencialmente los métodos de computación inteligente. Y es en esta línea donde hemos desarrollado esta tesis doctoral, la cual pretende alcanzar los siguientes objetivos específicos, que pueden considerarse objetivos SMART (específicos, medibles, asignables, realistas y relacionados con el tiempo), y que podrán desembocar en posibles productos de transferencia tecnológica en el ámbito de la salud. Estos objetivos son los siguientes:

1) Sistema de ayuda al diagnóstico de DCL basado en computación inteligente.

   Incluye la caracterización del DCL, importante por ser una enfermedad difícil de diferenciar de la EA debido a la similitud de sus síntomas. Este sistema facilitaría el diagnóstico diferencial del DCL, especialmente relevante en Atención Primaria, ya que es donde los pacientes o sus familiares acuden inicialmente. Para ello, consideramos necesario un clasificador multiclase capaz de diferenciar a los sujetos sanos de los afectados por DCL. Entre las múltiples opciones posibles, consideramos que dicho clasificador podría basarse en una Red Neuronal Artificial (RNA) híbrida y ontogénica.

2) Sistema de ayuda al diagnóstico de EA basado en computación inteligente.

   Tanto éste como el anterior son objetivos de gran interés clínico, ya que permitirían identificar los criterios diagnósticos representativos de ambas neuropatologías, aunque este es especialmente deseable dado que la EA es la demencia más extendida. Al igual que el anterior, ambos objetivos generan conocimiento no solo a nivel clínico, sino también computacional. En este caso, la razón es que dicho sistema posiblemente se base en nuevos modelos de arquitecturas neuronales con capacidad para trabajar con datos faltantes y desequilibrados, y clasificación multiclase (EA vs. DCL vs. sanos). Esto coincide con lo indicado por investigadores del ámbito biológico/clínico especialistas en este campo [Petersen, 2004; Hampel et al., 2011]. Para ello, consideramos viable el uso de la fusión

de datos y las arquitecturas neuronales. Se desarrollarán nuevas arquitecturas neuronales para mejorar el rendimiento de los sistemas actuales de aprendizaje automático.

3) Diseño y desarrollo de nuevas arquitecturas neuronales capaces de cumplir los requisitos y abordar las tareas descritas en los objetivos anteriores.

   Estas nuevas propuestas provendrán fundamentalmente de mejoras en arquitecturas neuronales paradigmáticas. Deben ser capaces de impulsar el desarrollo de sistemas inteligentes adecuados para el diagnóstico de DCL y EA. Estos sistemas no sólo deben ser precisos y fiables, sino también fáciles de usar, accesibles y rápidos en el ámbito clínico, lo que permitirá avances en la medicina traslacional.

4) Encontrar un conjunto de criterios de diagnóstico que sean suficientes y apropiados para el diagnóstico diferencial de EA, DCL y cerebros envejecidos sanos.

   Actualmente no existe un conjunto de criterios específicos (NINCDS-ADRDA, CAMDEX, DSM-IV y ICD-10) que hayan sido ampliamente verificados, adoleciendo de baja sensibilidad y especificidad en general. A menudo se afirma que la EA sólo puede diagnosticarse con total fiabilidad en la autopsia, considerándose el diagnóstico *ante mortem* como posible como mucho [McKhann et al., 1984]. Se realizará un estudio tanto de los biomarcadores que han comenzado a mostrar resultados prometedores en la última década como de las diversas formas de combinarlos.

5) Hacia una solución de e-Salud para el diagnóstico de la EA y el DCL.

   Integrará los sistemas mencionados anteriormente basados en computación inteligente. La solución de e-Salud contará con un perfil para Atención Primaria y otro para Atención Especializada, por lo que constituye un objetivo de gran interés clínico y sociosanitario. Además de facilitar el acceso a una atención médica adecuada, diagnóstico, tratamiento y seguimiento a cualquier paciente afectado por estas enfermedades, sin depender en parte de la disponibilidad horaria y geográfica de un especialista,

   La incorporación de sistemas de personalización, participación y capacidades de interconsulta completará esta solución de e-Salud.

   Dispondrá de funcionalidades que agilizarán y harán fiable la gestión del médico (proceso de diagnóstico).

En esta tesis proponemos dos nuevas arquitecturas neuronales híbridas, la Modular Hybrid Growing Neural Gas (MyGNG) y la Supervised Reconfigurable Growing Neural Gas (SupeRGNG), ambas basadas en la también red neuronal artificial ontogenética Growing Neural Gas (GNG).

La MyGNG que hemos usado en esta tesis es una versión mejorada y más simple de la descrita en [Sosa-Marrero et al., 2021], y fue introducida en [Cabrera-León et al., 2024b]. La conforman dos módulos, Figure 1: el primero está construido con una red ontogénica no supervisada y autoorganizada desarrollada por Fritzke, la GNG [Fritzke, 1995]; mientras que el segundo, está basado en una RNA supervisada muy popular, el perceptrón monocapa [Rosenblatt, 1961; Widrow and Lehr, 1990].

MyGNG está organizada jerárquicamente a propósito: primero se realiza un clusterizado de los datos de entrada y luego se realiza el etiquetado. En otros modelos este clusterizado no existe y la ventaja que proporciona es que se proyectan los datos de entrada en un espacio con

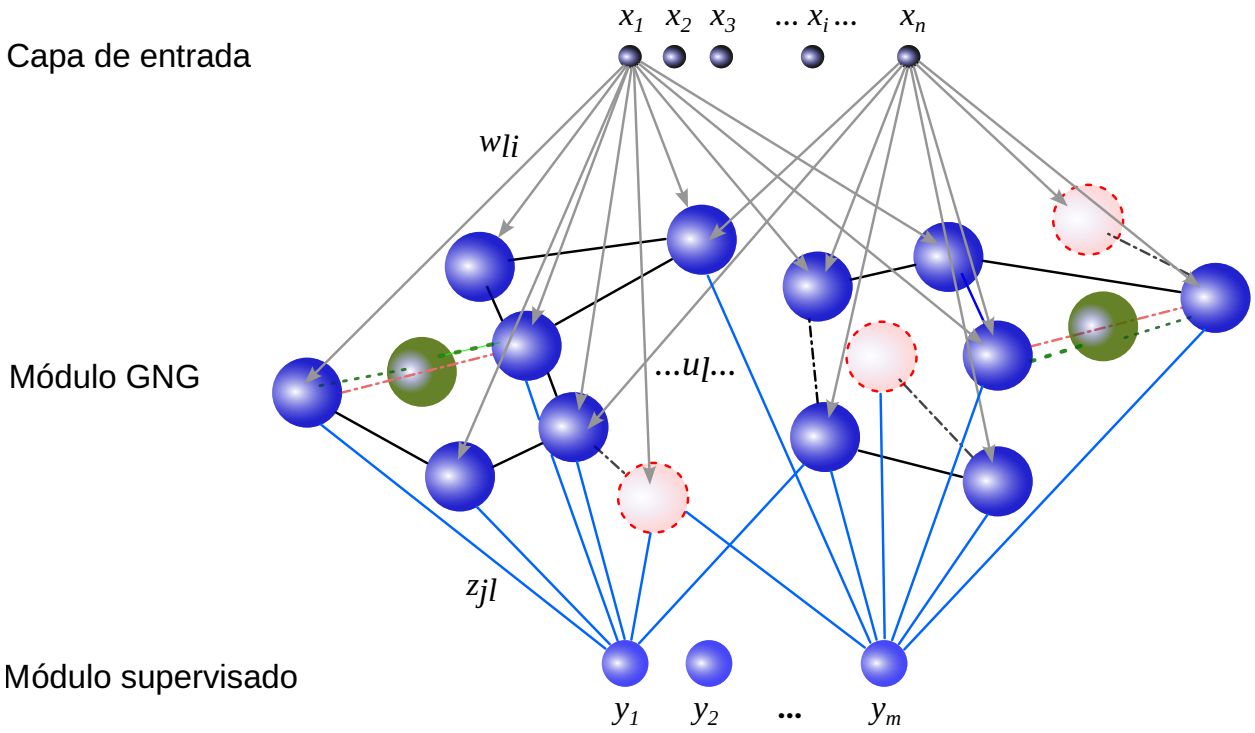


Figure 1: Estructura de la Modular Hybrid Growing Neural Gas, donde x_i es la i-ésima componente del vector de entrada; w_{li}, los pesos de la neurona li; y u_l, una neurona/unidad del módulo GNG

más dimensiones mientras se preserva su topología para facilitar la clasificación hecha luego en el módulo de etiquetado [Cabrera-León et al., 2024b]. Menos tiempo de entrenamiento y mayor rendimiento son esperados al hacer eso, como ocurre en otras redes híbridas como es la Counterpropagation Network [Cabrera-León et al., 2018a].

Estos dos módulos aprenden secuencialmente. Esto es, se hace en cascada: el segundo módulo (llamado "Supervisado" en Figure 1) usa para entrenar las etiquetas de los datos y la salida del primer módulo (denominado "GNG") después de que éste haya entrenado. En esta figura los colores indican los dos procesos relacionados con la biología que ocurren: neurogénesis y apoptosis neuronal. En azul las neuronas que se adaptan; en verde, las nuevas (creadas entre la neurona con el mayor error y su vecina con mayor error); y en rojo, las que van a ser borradas (son las que se quedan sin conexiones).

En [Fritzke, 1995, 1997b] describen cómo entrena la GNG. Empieza con 2 neuronas, se adapta, se encoge y crece, consiguiendo un aprendizaje topológico.

La adaptación de los pesos sigue la Equation 1 donde ε_b y ε_n son los ratios de aprendizaje de la ganadora y sus vecinas, respectivamente [Fritzke, 1995],

$$\begin{aligned} \Delta w_{s_1} &= \varepsilon_b(\xi - w_{s_1}) \\ \Delta w_n &= \varepsilon_n(\xi - w_n) \text{ para toda vecina directa } n \text{ de } s_1 \end{aligned} \tag{1}$$

Las regiones con mayor error se eligen en base a una variable local de error en cada neurona. La neurogénesis ocurre en esa zona y con una cierta frecuencia λ, y dicho error se reduce tras cada inserción en una proporción β.

Cambios en las conexiones alterna la topología de la red, ya sea creando nuevas o borrando las que sean más antiguas que a_{max}, lo que puede provocar apoptosis neuronal si la neurona se queda aislada.

Un perceptrón monocapa, supervisado, aprende según la "regla del perceptrón", Equation 2, que indica como sus pesos se actualizan [Widrow and Lehr, 1990]. En esta ecuación, $x(k)$ es una entrada; $\omega(k)$, pesos; ρ, ratio de aprendizaje; y $\tilde{e}(k) = d(k) - y(k) = d(k) - \text{sgn}[\omega^T(k) \cdot x(k)]$, siendo sgn la función signo; $d(k)$, la salida deseada para una entrada $x(k)$; y $y(k)$, la salida obtenida dada esa entrada.

$$\omega(k+1) = \omega(k) + \rho \cdot \frac{\tilde{e}(k)}{2} \cdot x(k) \tag{2}$$

La SupeRGNG, como su nombre indica, posee un mecanismo de reconfiguración de naturaleza supervisada, cuyo propósito es refinar los clústeres obtenido de manera no supervisada y que puede dividirse en dos procesos, que se realizan con la misma frecuencia y secuencialmente: desconexión de neurona con etiquetas diferentes pero que fueron erróneamente conectadas, y reconexión, en el cual una nueva conexión une neuronas con misma etiqueta pero que estuvieran erróneamente separadas. Además, posee un mecanismo de parada temprana que acorta el tiempo de entrenamiento y evita la posibilidad de sobreajuste. Al contrario que en [Cabrera-León et al., 2023], donde ambos procesos se hacían cuando el aprendizaje de la GNG terminaba, en esta versión de SupeRGNG se pueden llevar a cabo dinámicamente, tras cualquier *epoch* inicial (η) y con la frecuencia deseada (ι, en iteraciones) tras esa *epoch* η, siendo ambos valores elegidos por el usuario.

En Figure 2 se observa la estructura, destacando la neurona central que fue etiquetada erróneamente y que pasará a reconfigurarse.

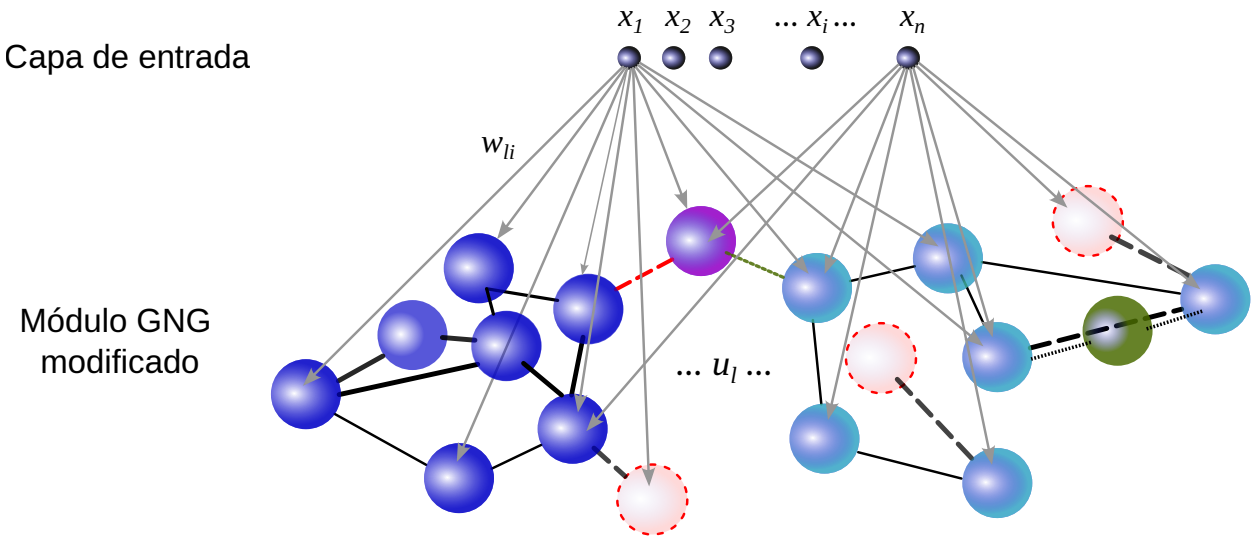


Figure 2: Topología de la Supervised Reconfigurable Growing Neural Gas. En rojo, neuronas que van a morir; verde, nuevas; negro y línea discontinua, desconexiones no supervisadas; en rojo y línea discontinua, desconexiones supervisadas, y en verde y línea discontinua, reconecciomes supervisadas.

El aprendizaje de la SupeRGNG es similar al de la GNG pero, aparte de guardar las etiquetas de los datos, se realiza la reconfiguración cuando indique η y ι. Consiste en 4 pasos: 1) asignar una clase a cada clúster/componente conectada, 2) localizar todas las

Table 1: Comparativa entre la SupeRGNG y métodos de aprendizaje automático neuronales y no neuronales populares (CN vs EA).

Dataset	Método	Accu	Sens	Spec	AUC	CUI+	CUI-
"Invasivo", Con AGE, 2PC, MaxAbsScaler, NCA identity	DT	0.99	0.99	0.99	0.99	0.98	0.98
	RF	0.99	0.99	0.99	0.99	0.98	0.98
	NB	0.99	0.99	0.99	0.99	0.99	0.99
	SVM	1	1	1	1	1	1
	MLP	0.97	0.95	0.98	0.97	0.96	0.95
	CPN	1	1	1	1	1	1
	ResNet-18	0.99	1	0.99	1	0.98	0.99
	ResNet-50	0.99	1	0.99	1	0.98	0.99
	ParallelNet	0.99	0.99	0.99	0.99	0.98	0.99
	SupeRGNG	1	1	1	1	1	1
"Invasivo", Sin AGE, 2PC, MaxAbsScaler, NCA identity	DT	0.99	0.99	0.99	1	0.98	0.98
	RF	1	1	1	1	0.99	0.99
	NB	1	1	1	1	1	1
	SVM	1	1	1	1	1	1
	MLP	0.94	0.88	0.99	0.93	0.97	0.90
	CPN	1	1	1	1	1	1
	ResNet-18	0.99	1	0.97	1	0.98	0.99
	ResNet-50	0.99	0.99	0.99	1	0.98	0.99
	ParallelNet	0.99	0.99	0.99	0.99	0.98	0.98
	SupeRGNG	1	1	1	1	1	1
"Invasivo", Con AGE, 2PC, RobustScaler, NCA identity	DT	0.99	0.99	0.99	0.99	0.98	0.98
	RF	0.99	0.99	0.99	0.99	0.98	0.98
	NB	0.99	0.99	0.99	0.99	0.99	0.99
	SVM	1	1	1	1	0.99	0.99
	MLP	1	1	1	1	0.99	0.99
	CPN	1	1	1	1	1	1
	ResNet-18	0.99	1	0.99	1	0.99	0.99
	ResNet-50	0.99	1	0.99	1	0.99	0.99
	ParallelNet	0.99	1	0.99	0.99	0.99	0.99
	SupeRGNG	1	1	1	1	1	1
"No-invasivo", Con AGE, 2PC, RobustScaler, NCA identity	DT	1	1	1	1	0.99	0.99
	RF	0.99	0.99	0.99	1	0.99	0.99
	NB	0.99	0.99	0.99	0.99	0.98	0.98
	SVM	1	1	1	1	1	1
	MLP	0.98	0.99	0.96	0.98	0.91	0.96
	CPN	1	1	1	1	1	1
	ResNet-18	0.84	0.73	0.93	0.90	0.64	0.76
	ResNet-50	0.85	0.76	0.92	0.90	0.67	0.77
	ParallelNet	0.66	0.21	1	0.71	0.27	0.63
	SupeRGNG	1	1	1	1	1	1

Acrónimos: Accu (accuracy), AUC (Area Under the Curve), CUI (Clinical Utility Index), DT (Decision Tree), MLP (Multilayer Perceptron), MyGNG (Modular Hybrid Growing Neural Gas), NB (Naïve Bayes), RF (Random Forest), Sens (sensitivity), Spec (specificity), SupeRGNG (Supervised Reconfigurable Growing Neural Gas), SVM (Support Vector Machine).

Table 2: Comparativa entre la SupeRGNG y métodos de aprendizaje automático neuronales y no neuronales populares (CN vs DCL).

Dataset	Método	Accu	Sens	Spec	AUC	CUI+	CUI-
	DT	0.74	0.74	0.74	0.73	0.54	0.54
	NB	0.79	0.79	0.79	0.75	0.62	0.62
	RF	0.80	0.80	0.80	0.74	0.65	0.65
6 componentes	SVM	0.84	0.84	0.84	0.79	0.71	0.71
proyectadas,	MLP	0.79	0.79	0.79	0.77	0.63	0.63
RobustScaler,	CPN	0.86	0.91	0.76	0.91	0.80	0.63
t-SNE con	ResNet-18	0.83	0.88	0.74	0.87	0.75	0.58
perplexity=100	ResNet-50	0.83	0.89	0.74	0.88	0.76	0.59
	ResNet-101	0.83	0.89	0.72	0.87	0.75	0.57
	ParallelNet	0.82	0.89	0.70	0.87	0.75	0.55
	SupeRGNG	0.86	0.86	0.86	0.88	0.74	0.74

Acrónimos: Accu (accuracy), AUC (Area Under the Curve), CUI (Clinical Utility Index), DT (Decision Tree), MLP (Multilayer Perceptron), MyGNG (Modular Hybrid Growing Neural Gas), NB (Naïve Bayes), RF (Random Forest), Sens (sensitivity), Spec (specificity), SupeRGNG (Supervised Reconfigurable Growing Neural Gas), SVM (Support Vector Machine).

y bajo coste. Estas características las hacen especialmente interesantes para la Atención Primaria.

En comparación con otras soluciones de aprendizaje automático, SupeRGNG se comportó de forma similar, obteniendo mejores resultados en la mayoría de las métricas, excepto en AUC. Al comparar estos valores con la arquitectura híbrida CPN, la diferencia no es estadísticamente significativa. Los valores de sensibilidad con CPN fueron los mejores, aunque presentaron una especificidad mucho menor y un CUI-. Los mejores valores de CUI+ de SupeRGNG podrían indicar que esta red es una buena opción para la medicina traslacional.

La mayoría de las arquitecturas comparadas arrojaron buenos resultados, siendo la SupeRGNG la que presentó valores buenos, aunque equilibrados, en los pares sensibilidad-especificidad y CUI+-CUI-. La sensibilidad, AUC y CUI+ en CPN y Deep Neural Networks (DNNs) fueron superiores o ligeramente superiores a las de la SupeRGNG. Sin embargo, sus valores de especificidad y CUI- fueron bastante inferiores, en el caso de ParallelNet hasta 0,16 y 0,19 menos, respectivamente. Esto confirma que las arquitecturas ontogenéticas como la propuesta son buenas soluciones computacionales para distinguir a los sujetos con DCL de aquellos cognitivamente normales.

Por otro lado, considerando que solo se ha propuesto un conjunto de características y se han obtenido buenos resultados, se puede concluir que el conjunto de características propuesto (es decir, ABETA, AGE, MMBALLDL, MMDAY, NPIL y MMYEAR) puede recomendarse a los médicos como los criterios clínicos óptimos para la tarea de clasificación del DCL-CN. Dado que incluye ABETA, que requiere la realización de una extracción invasiva de líquido cefalorraquídeo, no se recomienda para Atención Primaria.

El modelo SupeRGNG superó a los demás modelos por un amplio margen, logrando el mismo rendimiento que la variante anterior [Cabrera-León et al., 2023], seguido del MyGNG. El Multilayer Perceptron (MLP) y las DNNs no mostraron un buen rendimiento,

Table 3: Comparativa entre la SupeRGNG, la MyGNG y métodos de aprendizaje automático neuronales y no neuronales populares (DCL vs EA).

Dataset	Método	Accu	Sens	Spec	AUC	CUI+	CUI-
2PC, Unscaled, PCA	DT	0.91	0.91	0.91	0.92	0.83	0.83
	NB	0.91	0.91	0.91	0.92	0.83	0.83
	RF	0.90	0.90	0.90	0.92	0.81	0.81
	SVM	0.94	0.94	0.94	0.94	0.88	0.88
	MLP	0.89	0.89	0.89	0.90	0.79	0.79
	CPN	0.87	0.87	0.87	0.96	0.64	0.82
	ResNet-18	0.90	0.77	0.96	0.94	0.68	0.86
	ResNet-50	0.89	0.80	0.93	0.94	0.66	0.85
	ResNet-101	0.90	0.73	0.97	0.95	0.67	0.87
	ParallelNet	0.89	0.73	0.96	0.95	0.65	0.85
	MyGNG	0.93	0.97	0.83	0.96	N/A	N/A
	SupeRGNG	0.98	0.98	0.98	0.97	0.96	0.96

Acrónimos: Accu (accuracy), AUC (Area Under the Curve), CUI (Clinical Utility Index), DT (Decision Tree), MLP (Multilayer Perceptron), MyGNG (Modular Hybrid Growing Neural Gas), N/A (Not Available), NB (Naïve Bayes), RF (Random Forest), Sens (sensitivity), Spec (specificity), SupeRGNG (Supervised Reconfigurable Growing Neural Gas), SVM (Support Vector Machine).

especialmente en cuanto a la sensibilidad del último. Esto confirma que las arquitecturas ontogenéticas como la propuesta son buenas soluciones computacionales para distinguir entre sujetos con DCL y con EA.

Por otro lado, considerando que solo se ha propuesto un conjunto de características y que se han obtenido muy buenos resultados, se puede concluir que el conjunto de características propuesto (es decir, MMSCORE, MMDATE, MMBALLDL, ADAS_Q7, MMYEAR y FAQSHOP) puede recomendarse a los profesionales como los criterios clínicos óptimos para la tarea de clasificación de DCL-EA. Otra ventaja de este conjunto de características es que sólo se incluyen funciones económicas, no invasivas y fáciles de obtener, lo que hace interesante su aplicación incluso en Atención Primaria.

En los dos escenarios analizados, el método SupeRGNG mostró un buen rendimiento, superando a los métodos de aprendizaje automático. Si bien sus valores de AUC fueron similares o ligeramente superiores a los de los métodos CPN y los cuatro métodos de Deep Learning (DL), los valores del resto de las métricas no lo fueron. Por el contrario, el método MyGNG fue superado por redes profundas y CPN, pero obtuvo resultados de rendimiento similares a los de los clasificadores de aprendizaje automático más populares.

Además, el conjunto de características propuesto (es decir, VENTRICLES, ABETA, FAQTOTAL, MMSCORE y AGE) ha demostrado ser un buen conjunto de características para el problema de clasificación CN-DCL-EA. Por lo tanto, es adecuado para su presentación a médicos con fines diagnósticos, aunque la presencia de ABETA hace que este conjunto no sea el más recomendable para Atención Primaria.

Tanto la MyGNG como la SupeRGNG han demostrado obtener muy buenos resultados diagnosticando EA y DCL, superando a algoritmos de aprendizaje automático populares

Table 4: Comparativa entre la SupeRGNG, la MyGNG y métodos de aprendizaje automático neuronales y no neuronales populares (CN vs DCL vs EA).

Dataset	Método	Accu	Sens	Spec	AUC	CUI+	CUI-
Con AGE, 3PC, StandardScaler, NCA identity	DT	0.86	0.86	0.86	0.90	0.73	0.73
	NB	0.87	0.87	0.87	0.88	0.76	0.76
	RF	0.88	0.88	0.88	0.89	0.77	0.77
	SVM	0.88	0.88	0.88	0.90	0.77	0.77
	MLP	0.71	0.71	0.71	0.75	0.51	0.51
	CPN	0.93	0.93	0.93	0.95	0.88	0.88
	ResNet-18	0.91	0.91	0.91	0.94	0.85	0.85
	ResNet-50	0.91	0.91	0.91	0.94	0.85	0.85
	ResNet-101	0.91	0.91	0.91	0.93	0.85	0.85
	ParallelNet	0.92	0.92	0.92	0.94	0.86	0.86
	MyGNG	0.83	0.89	0.78	0.83	N/A	N/A
	SupeRGNG	0.88	0.88	0.88	0.95	0.77	0.77
Con AGE, 4PC, RobustScaler, NCA identity	DT	0.86	0.86	0.86	0.86	0.74	0.74
	NB	0.86	0.86	0.86	0.89	0.74	0.74
	RF	0.88	0.88	0.88	0.91	0.77	0.77
	SVM	0.87	0.87	0.87	0.90	0.76	0.76
	MLP	0.70	0.70	0.70	0.53	0.50	0.50
	CPN	0.93	0.93	0.93	0.94	0.88	0.88
	ResNet-18	0.92	0.92	0.92	0.94	0.86	0.86
	ResNet-50	0.91	0.91	0.91	0.93	0.85	0.85
	ResNet-101	0.91	0.91	0.91	0.94	0.84	0.84
	ParallelNet	0.91	0.91	0.91	0.94	0.85	0.85
	MyGNG	0.86	0.89	0.79	0.83	N/A	N/A
	SupeRGNG	0.89	0.89	0.89	0.94	0.79	0.79

Acrónimos: Accu (accuracy), AUC (Area Under the Curve), CUI (Clinical Utility Index), DT (Decision Tree), MLP (Multilayer Perceptron), MyGNG (Modular Hybrid Growing Neural Gas), NB (Naïve Bayes), RF (Random Forest), Sens (sensitivity), Spec (specificity), SupeRGNG (Supervised Reconfigurable Growing Neural Gas), SVM (Support Vector Machine).

y a técnicas de aprendizaje profundo. Se han obtenido, también, conjuntos de criterios diagnósticos óptimos para las tareas de clasificación tratadas, lo que posibilitará a los clínicos establecer un protocolo de criterios diagnósticos estándar en el diagnóstico de la EA.

Su integración en un entorno clínico virtual de ayuda al diagnóstico nos permitirá proponer una solución de e-Salud que dará lugar al diagnóstico universal en el ámbito de las demencias y de las patologías neuropsicológicas en general. Ello supondrá una mejora en la calidad de vida del paciente por la posibilidad de acceder a un diagnóstico, la rapidez del mismo, así como su fiabilidad. Entre las posibles soluciones se encuentra Clinical Virtual Environment to aid diagnosis and prognosis of Alzheimer's Disease and other dementias (EDEVITALZH).

Considerando todo lo anterior, los sistemas inteligentes desarrollados para cumplir con los objetivos primero y segundo de esta tesis doctoral pueden utilizarse como motores de detección inteligente en un esquema único o múltiple en el módulo de diagnóstico de EDEVITALZH (Figure 3). De esta forma, se proporciona una solución completa de e-Salud en los campos de Atención Primaria, Neurología y Gerontología, capaz de abordar todas las etapas de la enfermedad en el continuo de la EA.

Podemos concluir, en general, que las contribuciones principales de esta tesis doctoral son sus objetivos en sí, que han sido completamente cubiertos. De los desarrollos de esta tesis se obtienen las siguientes conclusiones, exponiéndose también las principales aportaciones:

1. La capacidad de las arquitecturas ontogénicas, como el SupeRGNG, para abordar problemas complejos del mundo real, como el diagnóstico diferencial de la EA en sus etapas iniciales, en comparación con las arquitecturas profundas, los ensembles neuronales y otros enfoques tradicionales de aprendizaje automático.

2. La posibilidad de obtener diagnósticos de alta precisión utilizando criterios clínicos multimodales, no invasivos y de uso apropiado en Atención Primaria.

3. La posibilidad de lograr un diagnóstico universal tanto en Atención Primaria como especializada (neurología y geriatría).

4. Se ha creado un conjunto de datos integrando datos longitudinales de diferentes modalidades, como pruebas neuropsicológicas, datos demográficos, biomuestras, genética, neuroimagen cuantitativa y líquido cefalorraquídeo. Estos datos se extrajeron de múltiples archivos de la base de datos Alzheimer's Disease Neuroimaging Initiative (ADNI), e incluyen datos de sujetos cognitivamente normales (CN), con DCL y con EA.

5. Se han propuesto dos sistemas inteligentes para el diagnóstico de todas las etapas del continuo de la EA, basados en dos arquitecturas neuronales novedosas. Además, se ha demostrado que las RNAs son buenos enfoques para problemas complejos, y las dos que se desarrollaron, también con conjuntos de datos no balanceados, una característica frecuente en medicina.

6. Se ha propuesto una nueva RNA ontogenética híbrida, la MyGNG. Consta de dos módulos: una GNG — no supervisada — para la clusterización de los datos de entrada, seguido de un perceptrón monocapa — supervisado — para mejorar la clusterización. Entre sus principales características, se incluye un módulo basado en perceptrón, posterior a la GNG, cuyo propósito es mejorar la agrupación realizada por la GNG.

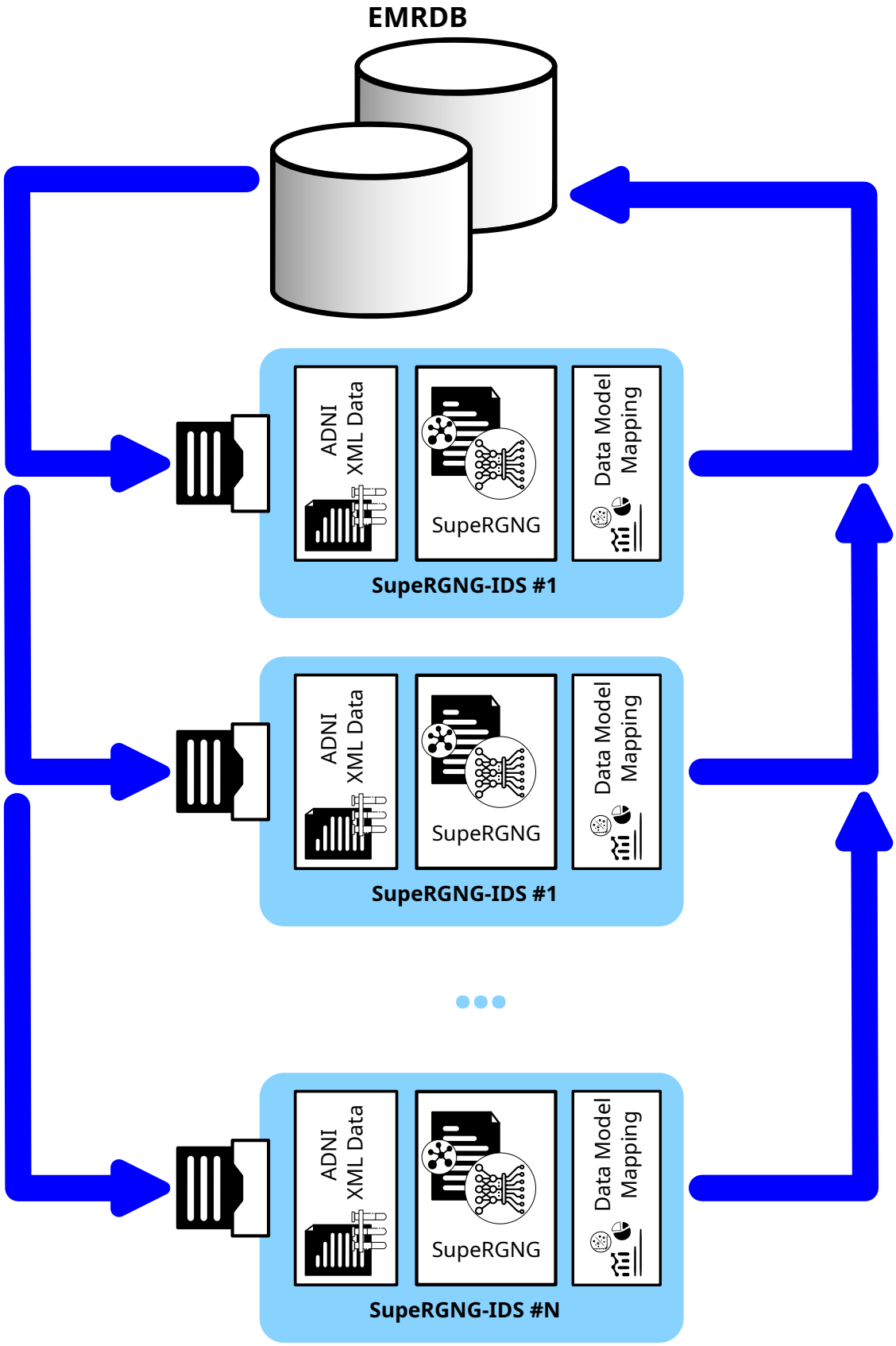


Figure 3: Ejemplo de un multi-esquema basado en SupeRGNG funcionando como motor de detección inteligente en el módulo de diagnóstico de EDEVITALZH.

7. Se ha desarrollado una novedosa RNA ontogénica supervisada, la SupeRGNG. Sus principales características son la desconexión y reconexión dinámica de lo clústeres, lo que mejora la clusterización realizada por la GNG en la que se basa. Estos procedimientos se pueden realizar de forma dinámica, permitiendo al usuario elegir cuándo iniciarlos y con qué frecuencia ejecutarlos. Al igual que otras RNAs, se incluyó un mecanismo de detención temprana para acortar el tiempo de entrenamiento. La SupeRGNG funciona de forma fiable con conjuntos de datos extremadamente desequilibrados (hasta un 90%) y con conjuntos de datos no gaussianos y no linealmente separables. Además, es capaz de producir muy buenos resultados de rendimiento con conjuntos de datos con una alta superposición de clases.

8. Los resultados de rendimiento con MyGNG y SupeRGNG en varios problemas de clasificación binaria relacionados con el diagnóstico de las diferentes etapas del continuo de la EA (CN vs. DCL, CN vs. EA y DCL vs. EA) fueron excepcionales, especialmente con SupeRGNG. SupeRGNG superó al resto de clasificadores de aprendizaje automático, incluyendo a una Convolutional Neural Network (CNN) paralela y a varias variantes de Residual Network (ResNet).

9. Los resultados de rendimiento con MyGNG en el problema multiclase CN vs. DCL vs. EA fueron similares a los de los modelos de aprendizaje automático más populares.

10. Los resultados de rendimiento con SupeRGNG en la tarea de clasificación CN vs. DCL vs. EA fueron similares a los de ResNets y una CNN paralela, y superiores a los de los restantes modelos de aprendizaje automático más usados.

11. Para cada tarea de clasificación binaria y multiclase que los médicos e investigadores consideraron de interés, se obtuvo el conjunto mínimo y óptimo de criterios diagnósticos, el cual puede considerarse suficiente y apropiado para el diagnóstico diferencial de la EA, el DCL y el envejecimiento cerebral sano.

12. Los dos métodos de computación neuronal desarrollados pueden integrarse en una solución de e-Salud en forma de sistemas inteligentes para facilitar el diagnóstico de la EA. Pueden utilizarse en prácticamente cualquier solución de e-Salud. De hecho, pueden utilizarse como motores de detección en el módulo de diagnóstico de EDEVITALZH. De esta manera, pueden proporcionar una solución completa de e-Salud en los ámbitos de la Atención Primaria, la Neurología y la Gerontología, que es capaz de abordar todas las etapas en el continuo de la EA. Además, EDEVITALZH es accesible de forma segura con dispositivos económicos, en cualquier momento y desde cualquier lugar, lo que permite a diferentes profesionales clínicos e investigadores colaborar e intercambiar información. Todo esto indica que nuestra propuesta es capaz de proporcionar un diagnóstico universal de la demencia cortical.

Abstract

The elderly population in developed countries has grown in the last decades. This has also entailed an increased prevalence of aging diseases such as strokes, cancers and dementia. Many types of dementia have been stated, being Alzheimer's Disease (AD) the most common one. Mild Cognitive Impairment (MCI) was considered a prodromal stage of AD, which is the most common dementia, albeit now AD is seen as a continuum. Machine Learning (ML) methods started to being used in the last decades for the detection or prognosis of both MCI and AD. Artificial Neural Networks (ANNs) is a family of ML algorithms inspired by biological neural networks, mainly human brain and neurons.

Increasing aging population involves more presence of non-communicable, geriatric and chronic diseases. An example of such illnesses is dementia, being AD its more representative example: around 70% of the cases [Zhu and Sano, 2006].

AD is a neurodegenerative disorder that affects memory and later the rest of cognitive areas [McKhann et al., 1984], producing dependence and disability states. Hence, patients with AD have a high social-economic costs for health systems, *e.g.* annually more than 28,000€ per patient in the Canary Islands (Spain) [Lopez-Bastida et al., 2006]. Every $5-10$ years a person gets older, the probability of having AD doubles [Zhu and Sano, 2006]. In AD we can distinguish between Early Onset AD and Late Onset AD, which depends on the patient's age: less than 65, or more or equal 65, respectively. Late Onset AD is the most common one, with 95% of the cases. AD develops steadily and slowly, over years, unlike other dementia. Consequently, some stages have been stated [Reitz and Mayeux, 2014], especially when distinguishing non-clearly normal aging from non-exactly dementia [Scinto and Daffner, 2000]: "benign senescent forgetfulness", "age-associated memory impairment", "pathological aging", "cognitive impairment, no dementia", and "Mild Cognitive Impairment". McKhann et al. [McKhann et al., 2011] wrote that AD dementia is a continuum, and Dubois at al. [Dubois et al., 2016] added to this that it cannot be separated in the 3 different clinically-defined stages. More recently, Jack at al. [Jack et al., 2018] defined AD as a set of neuropathological changes, hence not defined by clinical symptoms but *in vivo* by biomarkers and postmortem by autopsies.

MCI is a construct with memory complaints between Cognitively Normal (CN) and AD subjects. It was considered a prodromal stage of AD [Petersen, 2004], but after new findings, it started to be seen as a precursor of any dementia, or even reversible to cognitively healthy state [Reitz and Mayeux, 2014]. Patients with MCI have a probability of annually converting to dementia of $5-10\%$ [Mitchell and Shiri-Feshki, 2009] or $5-20\%$ [Schott and Petersen, 2015], much higher than the $1-2\%$ incidence in CN [Schott and Petersen, 2015]. Based on the severity of the symptoms, MCI is divided into Early Mild Cognitive Impairment (EMCI) and Late Mild Cognitive Impairment (LMCI).

As both AD and MCI were deemed to be related, many diagnostic methods have been used

for both. A wide selection of medical examinations, such as neuroimaging, neuropsychological scales and biomarkers, have been developed for this purpose [Cabrera-León et al., 2024a]. Recently, they have been used with a vast variety of computational solutions, from Support Vector Machines (SVMs) [Arimura et al., 2008; Vemuri et al., 2008; Magnin et al., 2009] to Artificial Neural Networks and Deep Learning [Liu et al., 2018a; Litjens et al., 2017; Ning et al., 2018; Senanayake et al., 2018]. More often than not, such solutions lack applicability in primary care, mostly because they require expensive or uncomfortable procedures, such as neuroimaging and Cerebrospinal Fluid (CSF) [Arimura et al., 2008], respectively. The importance of including the detection and diagnosis of cognitive impairments in Primary Care is based on two facts [Alberca Serrano and López-Pousa, 2011]: the longitudinal characteristic of Primary Care (which facilitates longitudinal studies), and the continuous growth of the elderly population (which increments the prevalence and incidence of cognitive disorders). However, there are several barriers that have to be dealt with in Primary Care [Alberca Serrano and López-Pousa, 2011]. Low confidence in diagnosis and lack of time per patient are the most common ones.

Taking these issues into account, in this thesis we propose the development of two novel neural architectures, the Modular Hybrid Growing Neural Gas (MyGNG) and the Supervised Reconfigurable Growing Neural Gas (SupeRGNG). Both ANNs are ontogenic as they are based on the Growing Neural Gas (GNG). They can tackle both the barriers that models may have when working in primary care (they are fast and provide good performance) and perform well with high unbalanced datasets, as is common with disease-related ones.

This doctoral thesis aims to meet the following specific goals, which can be considered SMART (specific, measurable, actionable, realizable, and time-bound[1] objectives, and that could lead to potential technology transfer products in the health field. These objectives are the following:

1) Diagnostic aid system of MCI based on intelligent computing.

   Including the characterization of MCI, which is important in that it is an ailment difficult to differentiate from AD due to its similar symptoms. This system would help in the differential diagnosis of MCI, which is especially relevant in Primary Care, since it is where patients or their relatives go initially. To achieve this, we consider necessary a multiclass classifier capable of differentiating healthy subjects from those affected with MCI. Among the many possible options, we consider that such a classifier could be based on hybrid and ontogenic ANNs.

2) Diagnostic aid system of AD based on intelligent computing.

   Both this and the previous one are objectives of high clinical interest since they would allow identifying the diagnostic criteria representative of both neuropathologies, although this one is particularly desirable to satisfy since AD is the most widespread dementia. As in the previous one, both objectives generate knowledge not only at the clinical level but also computationally. In this case, the reason for this is that such a system is possibly based on new models of neural architectures with the ability to work with missing and unbalanced data, and multiclass classification (AD vs. MCI vs. healthy). This is in line with what has been indicated by researchers in the biological/clinical field who are

[1] Other meanings of the acronym have been proposed such as "specific, measurable, assignable, realistic and time-related").

specialists in this field [Petersen, 2004; Hampel et al., 2011]. For this purpose, we consider the usage of data fusion and neural architectures feasible. New neural architectures will be developed to improve the performance of current Machine Learning systems.

3) Design and development of new neural architectures able to both fulfill the requirements and tackle the tasks described in previous goals.

   These new proposals will essentially come from improvements in paradigmatic neural architectures. They must be capable of advancing the development of intelligent systems suitable for the diagnosis of MCI and AD. These systems must not only be accurate and reliable, but also user-friendly, accessible, and fast in the clinical setting, providing advances in translational medicine.

4) Finding a set of diagnostic criteria that is sufficient and appropriate for the differential diagnosis of AD, MCI and healthy aging brains.

   Currently there is no set of specific criteria (NINCDS-ADRDA, CAMDEX, DSM-IV and ICD-10) that have been extensively verified, suffering from low sensitivity and specificity in general. It is often stated that AD can only be diagnosed with complete reliability at *postmortem*, with diagnosis at *antemortem* being considered possible at best [McKhann et al., 1984]. A study of both the biomarkers that have begun to show promising results over the past decade and the various ways of combining them would be carried out.

5) Towards an e-Health solution for the diagnosis of AD and MCI.

   It will integrate the above systems based on intelligent computing. The e-Health solution will have a profile for primary care and another for specialized care, which is why it is an objective of great clinical and socio-healthcare interest. In addition to facilitating access to adequate medical care, diagnosis, treatment and follow-up to any patient affected by these diseases, partly ignoring the time and geographical availability of a specialist. The incorporation of customization, participation, and inter-consultation capabilities will complete this eHealth solution. It will have functionalities that will speed up and make the physician's management (diagnosis process) reliable.

This PhD thesis is organized as follows. In chapter 1 we have briefly introduced the topics of this thesis, showing how it is structured, and will continue indicating the goals, explaining several aspects of dementia of the Alzheimer type and a milder stage, and reviewing the state of the art in diagnosis and prognosis of these diseases. In chapter 2 we describe ANNs, their characteristics and biological inspiration, and explain some paradigmatic models. In chapter 3, for each ANN developed, the Supervised Reconfigurable Growing Neural Gas and the Modular Hybrid Growing Neural Gas, a formal description and an analysis of its hyperparameters with synthetic datasets are included. In chapter 4 applications of both ANNs with real data are shown, with a description of each dataset used according to the classification task tackled, results yielded, a comparative with other Machine Learning algorithms, and finishing describing the integration of the ANNs developed in an e-Health solution is proposed, including the description of an example of such solution, Clinical Virtual Environment to aid diagnosis and prognosis of Alzheimer's Disease and other dementias (EDEVITALZH). The conclusions emanated from this research and several future works that can be derived from it are exposed in chapter 5. Finally, the appendices contain a list of acronyms in Appendix A, some recommendations from other authors on the values of Growing Neural

Gas hyperparameters in Appendix B, and a summary on statistical significance and how to choose the most appropriate statistical test Appendix C. After the appendices comes the bibliography used followed by two lists of publications and communications related to this thesis in which the PhD candidate has participated.

Chapter 1

Introduction and Background

1.1 Introduction

Dementia and other chronic diseases have become more prevalent with the increasing number of older adults. Alzheimer's Disease (AD), as the most common type of dementia, flourishes too. Seen as a prelude of AD, Mild Cognitive Impairment (MCI) began to be studied, despite being a construct and, due to being characterized by a cognitive degradation between Cognitively Normal (CN) and AD, a severity stage more challenging to differentiate.

Diagnosis and prognosis of AD and MCI are very complex due to the lack of specific biomarkers and the similarities between both or with other diseases. Neural computation methods have been used to tackle these difficulties, generally with very good performance results because some of them are able to work with data with noise, missing values and even in an unsupervised fashion. Inspired by their biological counterparts, many Artificial Neural Networks (ANNs) have been proposed for many different tasks, being dementia ones just one of the fields where they have been used.

Our research has been principally focused on the proposal of original and effective solutions for the diagnosis of AD and MCI. Both models, the Supervised Reconfigurable Growing Neural Gas and the Modular Hybrid Growing Neural Gas, belong to the family of supervised and ontogenic shallow ANNs. Furthermore, finding an adequate yet reduced set of features for each classification task that was considered relevant for medics was another of our goals. Confirming the possibility of including any of theses models into an e-Health solution was our last objective.

1.2 Alzheimer's Disease and Mild Cognitive Impairment

Albeit memory impairment is the main problem brought by AD, suffering it does not necessarily mean that a person has AD or even any type of non-AD dementia such as Frontotemporal dementia, Vascular dementia, Parkinson's Disease Dementia and Dementia with Lewy bodies, among others[1]. Indeed, many health problems also affect memory, thinking and other cognitive processes [Ronson, 2011]: depression, alcoholism, vitamin B12 deficit,

[1]Several tables and figures about the characteristics of different types of dementia can be found in [Korolev, 2014].

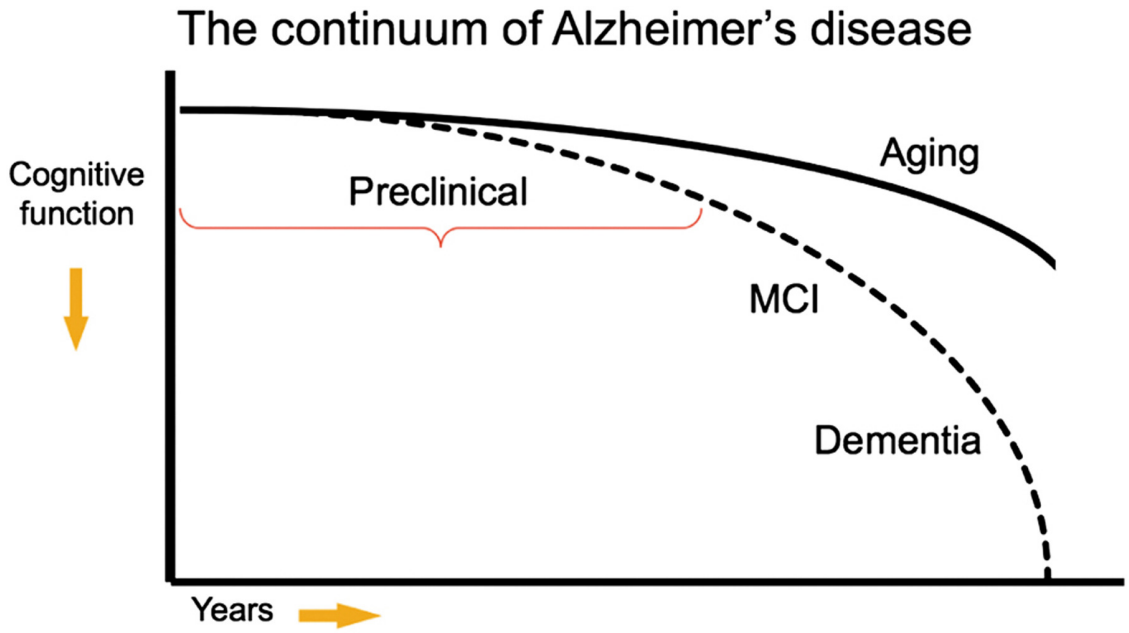


Figure 1.1: Hypothetical model of the decline in cognitive function along the AD continuum compared to normal aging. Source: Sperling et al. [2011]

stroke, delirium... Even worse, the recently recognized disease entity Limbic-predominant Age-related TDP-43 Encephalopathy is characterized by very similar symptoms, culminates with dementia and can only be diagnosed with certainty at autopsy [Nelson et al., 2019]. For this reason, a differential diagnosis is required, and it is still a problem to deal with [Bagyinszky et al., 2014]. The main difference with most of these health problems is that the cognitive function in AD impairs gradually, Figure 1.1.

MCI is a construct that is characterized by an intermediate cognitive impairment: between CN and AD subjects [Petersen, 2004]. Unlike in AD, the impairment in MCI patients does not affect their routine daily activities [Petersen, 2004; McKhann et al., 2011]. Several divisions have been proposed for MCI subjects [Cabrera-León et al., 2024a]: "reverters" (between consecutive visits their cognitive state improve, and so does their diagnostic label), "converters" (the opposite, as those that go from MCI to AD, which some researchers label as progressive Mild Cognitive Impairment (pMCI)), Early Mild Cognitive Impairment (EMCI), Late Mild Cognitive Impairment (LMCI), amnestic Mild Cognitive Impairment (aMCI), nonamnestic Mild Cognitive Impairment (naMCI), single-domain characteristics, and multi-domain characteristics.

It should be noted that AD diagnosis is a difficult task not only because other diseases having similar symptoms [Cabrera-León et al., 2024a], but also the severity of AD symptoms may vary seasonally: they are likely to be most pronounced during winter and early spring [Lim et al., 2018]. This is rarely taken into account, especially in longitudinal studies, and may be responsible of some "reverters" and "converters" cases. In order to reduce this complexity, solutions need to be sought. The use of Computer-Aided Diagnosis tools is one of them. They have been widely used to assist clinicians not only in the diagnosis or prognosis of these diseases but also with interpreting clinical criteria, especially neuroimaging.

1.2.1 Causes, risk factors and protective factors

The exact causes of AD are still unknown [Alberca Serrano and López-Pousa, 2011], although certain genes are considered to be involved in a small percentage of cases [Reitz and Mayeux, 2014]. Many hypotheses have tried to explain the origin of AD and how it gradually develops over time, usually with the goal of its early detection.

Some of the hypotheses that try to explain AD are:

- Amyloid hypothesis: after Amyloid Precursor Protein (APP) gets broken down, molecules of the Amyloid beta (Aβ) peptide are generated as a residue. Deposits of Aβ outside the cells form insoluble amyloid plaques [Mudher and Lovestone, 2002; Lewis and Trempe, 2014]. Senile plaques, as they are also called, may have different types [Scinto and Daffner, 2000]. These plaques negatively affect the functions of tissues and organs, although this disruption is not well known. On the other hand, after just one night of sleep deprivation, Aβ significantly accumulates in right hippocampus and thalamus, regions associated with AD [Shokri-Kojori et al., 2018]. Several neuroimaging methods prepared to detect amyloid, and concentrations of Cerebrospinal Fluid (CSF) Aβ42 or the CSF Aβ42/40 ratio are commonly used by researchers and medics [Jack et al., 2018].
- Tau hypothesis: in normal conditions, the tau (an abbeviation of tubulin associated unit) proteins are in charge of stabilizing the microtubules of cells, and are quite numerous in the neurons of the central nervous system, especially in the cerebral cortex. They are located in the cytoplasm of a cell [Scinto and Daffner, 2000]. Hyperphosphorylated tau proteins are not able to stabilize the microtubules correctly and start to get in pairs, which later forms neurofibrillary tangles (NFTs) inside the neurons. The neuron's cytoskeleton gets destroyed when the microtubules disintegrate due to the NFTs. After this, synapses start to malfunction and, at a later stage, neurons die [Mudher and Lovestone, 2002]. Usually, total tau and hyperphosphorylated tau concentrations in CSF and some neuroimaging techniques able to detect tau are used in dementia research [Jack et al., 2018].
- Cholinergic hypothesis: low acetylcholine synthesis. Abnormal cholinergic values may facilitate the deposition of neuritic plaques, and changes in behavior and cognition [Terry and Buccafusco, 2003].
- Genetic: many genes seem to increase the risk of developing AD, especially the Early Onset AD [Reitz and Mayeux, 2014; Scinto and Daffner, 2000], a type of AD that affects individuals before the age of 65. Simultaneous presence of Apolipoprotein E (APOE) ε4 and a variant of RNF219 genes in subjects with MCI or AD produced lower scores in some neuropsychiatric scales [Mosca et al., 2018].

Apart from lower brain volume due to having less neurons in several places of their brain, each neuron of a subject with AD usually has less number of dendrites, and many of these neurons have alterations in their mitochondria and fragmentations in their Golgi complexes [Baloyannis et al., 2019].

More recently, it was found that subjects with AD had more human herpesvirus 6A and human herpesvirus 7 than CN subjects, although causality could not be concluded [Readhead et al., 2018]. Regulatory relationships between APP processing genes and quantity of these viruses were observed.

Nowadays, these are considered some of the risk factors for the development of AD [Baumgart et al., 2015; Reiss et al., 2022; Reitz and Mayeux, 2014; Scinto and Daffner, 2000]:

- Age: AD is more common in people older than 64. Every 5–10 years a person gets older, the probability of having AD doubles [Zhu and Sano, 2006].
- Trisomy 21, commonly known as Down syndrome.
- Head injury, especially traumatic brain injury.
- Family history: genetics (hereditary) and environmental factors may affect same family members.
- Environmental factors: Killin et al. [2016] grouped them in six categories (air quality, toxic heavy metals, other metals, other trace elements, occupational-related exposures, and miscellaneous environmental factors). Moderate evidence was found for air pollution, aluminum, silicon, selenium, pesticides, vitamin D deficiency, and electric and magnetic fields [Killin et al., 2016].
- Genetics: many researchers affirm that having two APOE $\varepsilon 4$ alleles is a genetic risk factor not only for AD but also for other health problems such as stroke, sleep apnea and Parkinson's disease [Mahley et al., 2006; Mosca et al., 2018; Reitz and Mayeux, 2014]. Conversely, two APOE $\varepsilon 2$ alleles or one $\varepsilon 2$ and one $\varepsilon 3$ is beneficial, protecting from cognitive decline and neurodegeneration [Mahley et al., 2006; Wu and Zhao, 2016]. For instance, a person carrying two $\varepsilon 4$ is almost 15 times more likely to develop AD, whereas people who carry two $\varepsilon 2$ are 40% less prone to have it [Wu and Zhao, 2016]. Moreover, half of subjects with AD have at least an $\varepsilon 4$ allele, while it is only present in 20% of the general population [Wu and Zhao, 2016].
- Race: albeit the reason for this is still not well known, Hispanic/Latinos are up to 1.5x times more likely to have any type of dementia compared with non-Hispanic white older adults, whereas in the case of Black/African-American individuals, it is twice as likely as non-Hispanic white ones [Alzheimer's Association, 2025].
- Vascular problems: includes elevated blood pressure [Mantzavinos and Alexiou, 2017], and cerebrovascular diseases. More related to Vascular dementia (VaD) than with AD though.
- Patients with higher levels of LDL cholesterol ("bad") have a greater risk of early development of AD [Wingo et al., 2019], whereas no link with HDL cholesterol ("good") was found.
- Type 2 diabetes: it doubles the risk of AD. It can also increase the conversion risk from amnestic MCI to AD [Cooper et al., 2015].
- Female gender: more females suffer from AD and have a higher risk to develop it, although this gender bias is still not well understood. There seems to be a significant interaction between APOE $\varepsilon 2$ and sex, a protective role exclusively in males [Wu and Zhao, 2016]. On the other hand, females with APOE $\varepsilon 4$ have much more risk conversion from both CN to MCI and MCI to AD than males [Wu and Zhao, 2016].
- Mid-life obesity.
- Sleep apnea.
- Poor diet, as inadequate amounts and forms of nutrients may affect health. Nutritional deficits, such as low levels of some vitamins (B12, B3 and B1) and glucose, are frequent in subjects with dementia, albeit this relationship, from a practical point of view, should be considered bidirectional [Alberca Serrano and López-Pousa, 2011].
- Lack of exercise.
- Infection of the brain, such as meningitis and syphilis.
- Depression is associated with a higher risk of dementia and AD [Santabárbara et al., 2019].

A systematic review and meta-analysis mentioned that different levels of evidence were found regarding risk factors and interventions [Yu et al., 2020]. These authors found that there is strong evidence with the following risk factors: hyperhomocysteinaemia, diabetes, poor Body Mass Index management, hypertension in midlife, orthostatic hypotension, reduced education, head trauma, stress, less cognitive activity, and depression.

On the other hand, some protective factors are [Reitz and Mayeux, 2014]: Mediterranean diet [Cooper et al., 2015], physical exercise, and frequent intellectual activities. Caloric restriction mitigates pathologic aging of the brain and reduces age-related memory problems [Martí-Nicolovius and Arévalo-García, 2018]. High folate, a.k.a. folic acid or vitamin B9, levels are beneficial too, as there is lower conversion from any type of MCI to any kind of dementia [Cooper et al., 2015]. Vitamin E may hamper functional decline produced by AD, albeit not preventing conversion from MCI to dementia nor improving cognitive functions in both types of patients. [Farina et al., 2017].

AD is still an incurable and unstoppable illness [Lewis and Trempe, 2014]. Many researchers and pharmacological companies have investigated and tried to discover efficient medicines intended if not to cure, to stop or slow down dementia progression. One of the goals of finding these medications is the reduction of the economic impact of dementia if the patient does not progress to more severe stages of the disease. In the majority of pharmaeconomic studies, this economic impact was reduced after administration of some medicine as there were lower direct medical costs (*e.g.* hospitalization, medication, medics, nurses, adult daycare, home health aides…) and cost of caregiver time [Zhu and Sano, 2006]. Another reason for early diagnosis of AD is the more impact that treatments may achieve [Scinto and Daffner, 2000].

1.3 Criteria for the diagnosis and staging of Alzheimer's Disease

For years, the most used diagnostic criteria for AD was proposed by National Institute of Neurological and Communicative Disorders and Stroke and Alzheimer's Disease and Related Disorders Association (NINCDS-ADRDA) in 1984 [McKhann et al., 1984], Table 1.1. It was necessarily updated in 2007 and 2011 after new findings, especially those obtained with functional neuroimaging techniques and genetics, were made [Dubois et al., 2007; McKhann et al., 2011], Table 1.2 and Table 1.3.

Another prevailing diagnostic criteria was the Diagnostic and Statistical Manual of Mental Disorders (DSM)-IV-TR, a "text revision" of the DSM-IV [Pichot and López-Ibor Aliño, 1998]. It was published in 2000 by the American Psychiatric Association (APA) [Aliño et al., 2008]. In the most recent version, DSM-5 (2013), dementia started to be called "major neurocognitive disorder", whereas early stages of cognitive impairment, "mild neurocognitive disorder" [Alzheimer's Australia, 2017], a term equivalent to both prodromal AD and MCI, Table 1.4.

Unlike the previous ones, which are intended for clinical care, there are other diagnostic criteria that are recommended in research and clinical trials, such as International Working Group-1 (IWG-1), International Working Group-2 (IWG-2), and National Institute on Aging and Alzheimer's Association (NIA-AA) [Vos et al., 2015].

The NIA-AA Research Framework seems to be the most periodically updated one [Cummings, 2018; Jack et al., 2018]. It recommends the use of both neurodegeneration and amyloid

Table 1.1: NINCDS-ADRDA (1984) diagnostic criteria for AD (based on [McKhann et al., 1984]).

Definite AD	Clinical criteria for probable AD + Brain autopsy or biopsy
Probable AD	Clinical examinations. + Deficiencies in ≥ 2 areas of cognition. + Gradual deterioration of memory and other cognitive functions. + Consciousness not altered or lost. + Patient is from 40 to 90 years old, commonly > 65. + No other cognitive-affecting disease or disorder.
	Supported by: Gradual deterioration of specific cognitive functions. + Daily living activities affected and behavior changes. + Similar problems in other family members. + Laboratory results: normal CSF, normal EEG of brain activity, and cerebral atrophy in serial CT scans.
	After excluding other diseases: • Stagnation of illness progression. • Normal CT results for patient's age. • Depression, sexual disorders, weight loss, insomnia, physical and verbal outbursts, hallucinations... • Seizures (in advanced AD). • Neurological abnormalities (in advanced AD) such as gait alteration, myoclonus, and increased muscle tone.
	Unlikely or uncertain Probable AD if: Abrupt onset. + Early in the course of the disease: seizures, sensory loss, loss of coordination, or gait alteration.
Possible AD	Dementia syndrome, no other disorders that produce dementia, and variations in clinical course, onset and presentation. When there is another brain or systemic disorder that usually produces dementia but which is not considered to be *the* cause. In research, when there is a severe and progressive cognitive decline and other cause is unknown.

Table 1.2: NINCDS-ADRDA (2007) diagnostic criteria for AD (based on [Dubois et al., 2007]).

Definite AD	Clinical evidence. + Brain autopsy or biopsy.
	Clinical evidence. + Genetic evidence (mutations in chromosomes 1, 14 or 21).
Probable AD	A plus ≥ 1 supportive features B, C, D, or E: A. Early, significant and gradual memory impairment over ≥ 6 months, reported by patient or informants, which can be later confirmed with tests. Other cognitive changes can be present too. B. Medial temporal lobe atrophy: volume reduction of hippocampus, entorhinal cortex and amygdala. C. Abnormal CSF, such as low Aβ1–42, high total tau or high phospo-tau concentrations. D. Lower glucose metabolism in bilateral temporal parietal regions, or amyloid presence stated by ligands. E. Family members with AD autosomal dominant mutation.
Exclusion criteria	Medical history: sudden outset, early seizures, prompt gait alteration, or early changes in behavior. Clinical features: sensory loss, or extrapyramidal side effects. Other medical disorders that may affect cognitive functions such as depression, non-AD dementia, infections, toxics, and cerebrovascular disease.

markers in clinical settings, with whom prognosis is the most accurate [Vos et al., 2015]. Three biomarker classes[2] are included in this framework:

- Amyloid (A): obtained via amyloid Positron Emission Tomography (PET), CSF Aβ42, or Aβ42/40 ratio.
- Tau (T): from tau PET, or CSF phosphorylated tau.
- Neurodegeneration (N): indicated by Fluorodeoxyglucose-Positron Emission Tomography (FDG-PET), CSF total tau or Magnetic Resonance Imaging (MRI).
- Cognitive symptoms (C): after results from objective cognitive tests.

For these biomarker classes it has appeared a special nomenclature: AT(N)(C) "biomarker profile" [Jack et al., 2018], albeit it is most commonly seen as AT(N). The different combinations of these biomarker classes, whether it happens or not (binary), give all the eight possible categories that indicate presence of AD, non-AD dementia or healthy subjects, Table 1.5. The reason why neurodegeneration and cognitive symptoms are inside parenthesis in the AT(N)(C) profile is that, unlike amyloid and tau, they are not specific to AD [Jack et al., 2018]. In fact and according to Table 1.5, the presence of amyloid is enough to indicate AD continuum, no tau or neuronal injuries required. Presence of AD is indicated by the existence of at least tau or neural damage (or both), given abnormal amyloid quantities were detected too.

On the other hand, according to the NINCDS-ADRDA diagnostic criteria, several cognitive domains, also called areas of cognition, can be negatively affected in AD [McKhann et al., 1984]: time and place orientation, memory, language skills, praxis (motor abilities), attention,

[2]Cognitive symptoms is not considered a biomarker class.

Table 1.3: NINCDS-ADRDA (2011) diagnostic criteria for all-cause dementia and AD (based on [McKhann et al., 2011]).

All-cause dementia	Daily activities interfered. Functional decline over time. No other major psychiatric disorder Cognitive impairment in ≥ 2 of these domains: a. Remember new information. b. Judgment, reasoning and complex activities. c. Visuospatial abilities. d. Language. e. Changes in behavior.
Probable AD dementia	Previous for all-cause dementia plus the next ones: A. Gradual appearance. B. Unambiguous worsening over time. C. Most relevant cognitive deficits occur in more than one of the cognitive domains: amnestic, language, visuospatial, or executive presentations. D. More certainty if: a. Documented decline on successive evaluations. b. Genetic mutation. c. Biomarker evidence of the AD pathophysiological process: Aβ protein deposition (low Aβ42 and positive PET amyloid imaging), and neuronal injury (high CSF tau, reduced FDG uptake on PET in temporo-parietal cortex; and extreme atrophy on sMRI in medial, basal, and lateral temporal lobe, and medial parietal cortex) E. Unlikely probable AD dementia if there is: a. Cerebrovascular disease related to the onset of cognitive decline. b. Features from other dementia. c. Evidence for other neurological disease, or treatment that affects cognition.
Possible AD dementia	Atypical: sudden onset, progressive decline not enough documented... Evidence of the characteristics that reduce the likeliness of probable AD, previously mentioned.
Proved AD dementia	Previous cognitive and clinical criteria are met. + Demonstrated via neuropathological examination (autopsy or biopsy).
Unlikely AD dementia	Criteria for AD dementia was not met. Sufficient evidence for other diseases that may overlap with AD, or both Aβ and neural degeneration biomarkers are negative.

Table 1.4: DSM-5 (2013) criteria for MCI and dementia (differences in bold) (based on [Alzheimer's Australia, 2017]).

Mild neurocognitive disorder	**Modest** cognitive decline change in ≥ 1 cognitive domains. + Daily living activities **unaffected**. + No other mental disorder.
Major neurocognitive disorder	**Significant** cognitive decline change in ≥ 1 cognitive domains. + Daily living activities **affected**. + No other mental disorder.

Table 1.5: NIA-AA (2018) Research Framework (based on [Cummings, 2018; Jack et al., 2018]).

Amyloid	Tau	Neural damage	AT(N) profiles	Biomarker category	
Negative	Negative	Negative	A-T-(N)-	CN	
Positive	Negative	Negative	A+T-(N)-	AD pathologic change	AD continuum
Positive	Positive	Negative	A+T+(N)-	AD	AD continuum
Positive	Positive	Positive	A+T+(N)+	AD, or AD and other brain disorders	AD continuum
Positive	Negative	Positive	A+T-(N)+	AD and concomitant suspected non-AD pathologic change	AD continuum
Negative	Positive	Negative	A-T+(N)-		
Negative	Negative	Positive	A-T-(N)+	Non-AD pathologic change	
Negative	Positive	Positive	A-T+(N)+		

visual perception, problem-solving skills, and social function (activities of daily living). These cognitive domains are similar to those indicated in DSM-5 [Alzheimer's Australia, 2017]: complex attention, executive function, learning and memory, language, perceptual-motor function, and social cognition.

As observed, language is one of the cognitive domains frequently damaged by AD. Indeed, language differences among individuals in the AD spectrum have been detected in a big cohort of participants from distinct classes [Liampas et al., 2022].

1.3.1 Clinical diagnostic techniques

In spite of being more difficult, and frequently having lower accuracy [Galasko et al., 1998], the earlier AD is diagnosed, the better [Scinto and Daffner, 2000]. Also, the sooner some therapies and treatments are applied to an AD patient, the most effective they are [Bagyinszky et al., 2014; Scinto and Daffner, 2000].

A biomarker is any measurable characteristic or substance that allows the diagnosis or prognosis of an illness. A biomarker is deemed clinically valuable in this field if it provides values of sensitivity, specificity, precision, and Negative Predictive Value above 90% [Huynh and Mohan, 2017]. A proposal of how AD biomarkers progressively become abnormal is the following one, supported by the Alzheimer's Disease Neuroimaging Initiative: lower Aβ $\rightarrow$ higher tau $\rightarrow$ lower glucose metabolism (damaged neurons) $\rightarrow$ damage to brain structure $\rightarrow$ cognitive impairment. Significantly lower Aβ and higher tau concentrations in CSF in AD patients compared with CN subjects were experimentally confirmed by Galasko at al. [Galasko et al., 1998]. However, Furcila et al. [Furcila et al., 2018] declared: "the disease course of AD is highly variable, and neuropathological changes are not homogeneous", and also that "AD is not a unique entity even within the same neuropathological stage, since the microanatomical/neurochemical changes that occur in the hippocampus greatly vary from one patient to another". According to this, choosing values related to the hippocampus may not be an optimal option due to its variability within the same disease stage. Nevertheless, it should be noticed that a little number of patients was used in [Furcila et al., 2018], so generalization is not recommended and more research should be made to confirm or not

those points.

Most current diagnostic techniques are slow (neuropsychological tests), costly (brain imaging), or risky and invasive (CSF) [Laske et al., 2015]. Researchers are looking for novel diagnostic tests which should be characterized by a low price, being little to non-invasive, and with whom results can be obtained quickly. Potential biomarkers that have been tested more recently almost always comply with the previous conditions [Bagyinszky et al., 2014; Scinto and Daffner, 2000]. Some of them are obtained from the patient's genome [Huynh and Mohan, 2017], blood [DeMarshall et al., 2016; Huynh and Mohan, 2017], urine, serum, saliva, skin, and pupillary response. Other potential biomarkers include vascular disorders, protein dysfunctions, metal ions, and oxidative stress [Mantzavinos and Alexiou, 2017].

Several biomarkers have been proposed in the last decades for the early diagnosis of AD [Budelier and Bateman, 2020; Frisoni et al., 2017; Laske et al., 2015; Reitz and Mayeux, 2014], which will be described below. In Table 1.6 certain relevant characteristics of the most popular diagnostic tools are summarized.

Table 1.6: Characteristics of the most used diagnostic tools for AD diagnosis and prognosis.

Diagnostic tool	Invasiveness	Monetary cost	Temporal cost	Required knowledge	Diagnostic precision
EEG	Low-High	Medium	High	Medium	Medium
fNIRS	Low	Low-High	High	Medium	Medium
MRI	Low-Medium	Medium-High	High	High	Medium-High
PET	Medium	High	High	High	Medium-High
SPECT	Medium	Medium-High	Medium-High	High	Medium-High
B	Low	Low	Medium	Medium-High	Medium
CSF	High	Medium	Medium	High	High
HG/M/S	None-Low	Low-Medium	High	Medium-High	Medium
NT	None-Low	Low	Low-Medium	Low	Medium
OCTA	Low	Medium-High	Medium-High	High	Medium-High

Abbreviations: B = Blood; CSF = Cerebrospinal Fluid; EEG = Electroencephalography; fNIRS = Functional Near-Infrared Spectroscopy; HG/M/S = Human gait, movements or speech; MRI = Magnetic Resonance Imaging; NT = Neuropsychological tests; OCTA = Optical Coherence Tomography Angiography; PET = Positron Emission Tomography; SPECT = Single-Photon Emission Computed Tomography

1.3.1.1 Cerebrospinal Fluid

Cerebrospinal Fluid is a colorless biofluid in the brain and spinal cord. It is obtained via lumbar puncture, a quite invasive method where a needle is inserted between certain lumbar vertebrae. Several biomarkers of interest for dementia diagnosis can be found in CSF, usually Aβ42, the ratio Aβ42/40, total tau, and hyperphosphorylated tau. Combining CSF with neuroimaging — or, in general, with more than one biomarker — suggests better classification accuracy [Lista et al., 2014]. CSF is useful for the differential diagnosis of dementia between autopsy-confirmed AD and non-AD [Niemantsverdriet et al., 2018].

1.3.1.2 Transcranial Magnetic Stimulation

Transcranial Magnetic Stimulation (TMS) is a non-invasive brain stimulation method that allows not only the detection of early synaptic impairment and pathophysiological changes in cortical circuits, but also the prediction of AD progression [Lazzaro, 2018; Motta et al.,

2018]. Motta et al. found that long-term potentiation (LTP)-like cortical plasticity was a significant predictor of AD progression and cognitive decline while an excellent method for AD detection, achieving Area Under the Curve (AUC) values of 0.9 in the latter task [Motta et al., 2018].

1.3.1.3 Neuropsychological tests

There are many neuropsychological tests that cover different domains:

- Functional domain refers to how well the patient is able to do daily-living activities and how independent the patient is doing them. Examples of such activities are showering, shopping, dressing, and using public transport. Functional Activities Questionnaire (FAQ) is a test specialized in this domain [Pfeffer et al., 1982; Teng et al., 2010]. Others are Katz's index [Katz et al., 1963], Barthel's index [Mahoney and Barthel, 1965], and Lawton-Brody's index [Lawton and Brody, 1969].
- Cognitive domain includes language, recall, orientation, calculation, and attention, among others. Memory in general is analyzed. Examples of tests specialized in this domain are: Cognitive Change Index (CCI) [Rattanabannakit et al., 2016], Mini-Mental State Examination (MMSE) [Folstein et al., 1975], Montreal Cognitive Assessment (MoCA) [Nasreddine et al., 2005], Rey Auditory Verbal Learning Test (RAVLT) [Rosenberg et al., 1984], Wechsler Memory Scale (WMS) [Wechsler, 1945]...
- Behavioral domain stands for changes in mood and conduct. Depression, aggressiveness, or anxiety are some of the elements that are included in this group. Geriatric Depression Scale (GDS) is specialized in depression detection in geriatric environments [Yesavage and Sheikh, 1986]. Neuropsychiatric Inventory (NPI) is able to detect 10 different behavioral problems [Cummings et al., 1994; Cummings, 1997]. A shorter variant, Neuropsychiatric Inventory-Questionnaire (NPI-Q), also exists and was validated in [Kaufer et al., 2000].

Some neuropsychological assessments are prepared to test multiple domains: cognitive and functional, with Cognitive-Functional Composite (CFC) [Jutten et al., 2018] or Everyday Cognition (ECog) [Farias et al., 2008]; and cognitive and behavioral, with Alzheimer's Disease Assessment Scale (ADAS) [Rosen et al., 1984]. Special mention is Clinical Dementia Rating (CDR), another multiple domain — cognitive and functional — neuropsychological test, used for staging dementia [Hughes et al., 1982]. Values go from 0 for CN subjects to 3 for those with severe dementia. A variant exists called Clinical Dementia Rating Scale Sum of Boxes (CDR-SB), which seems to be useful for early stages [Cedarbaum et al., 2013]. Combined with hippocampal volume, CDR-SB was able to predict MCI to AD conversion [Borgio et al., 2012].

Multiple-domain amnestic MCI subtype can be detected with a set of neuropsychological tests, unlike single-domain amnestic and non-amnestic MCI subtypes [Klekociuk and Summers, 2014].

In moderate to severe AD patients, results obtained with neuroimaging and neuropsychological tests in autopsy-confirmed AD patients were similar to those which where just clinically diagnosed with AD [Fearing et al., 2007].

Regarding the validation of neuropsychological scales, only Wechsler Adult Intelligence Scales and MMSE were reported to be valid in up to 80% of the European countries included in [Maruta et al., 2011]. As only a few scales were available in all countries, these authors concluded that the use of these scales require more consensus in order to facilitate both the

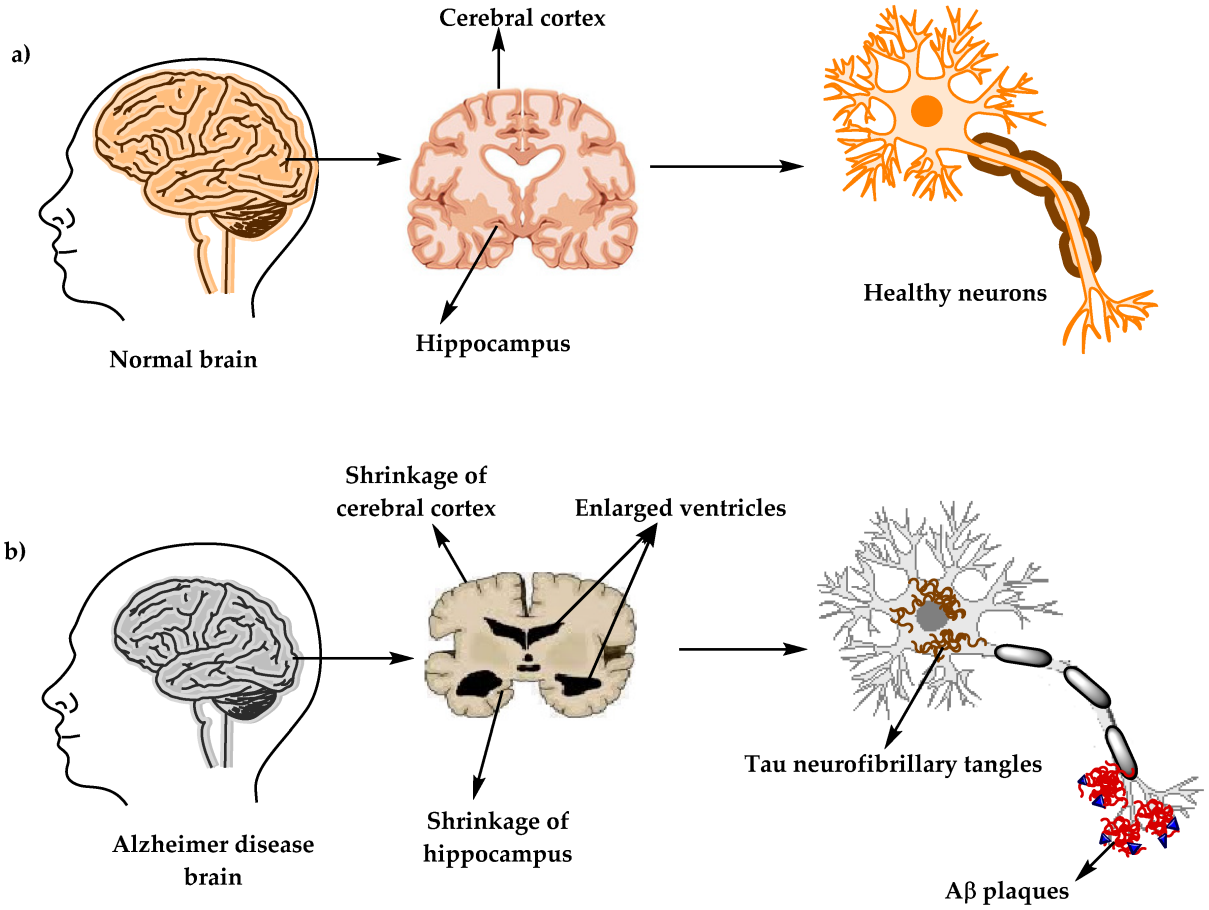


Figure 1.2: Differences in the physiological structure of the brain and neurons of a) a Cognitively Normal individual, and b) a patient with Alzheimer's Disease. Source: Breijyeh and Karaman [2020]

creation of international cooperative studies and the harmonization of the data gathered [Maruta et al., 2011].

1.3.1.4 Neuroimaging

As it can be seen in Figure 1.2 and Figure 1.3, the brain of a CN subject is different from that of a patient with MCI and AD: there is a gradual atrophy of the brain [Johnson et al., 2012], especially shrinkage of both hippocampus and cerebral cortex, and enlargement of ventricles. This brain volume loss can be explained by the neural apoptosis implied by some of the hypotheses on the origin of AD, subsection 1.2.1: high proliferation of abnormal Aβ and tau. Henceforth, the use of neuroimaging, *i.e.* images of the brain, is useful for the *in vivo* detection of AD and other dementia [Barkhof et al., 2011]. It is less invasive than brain biopsies, mostly made in autopsies though, which has been considered the only certain way to confirm the presence of definite AD [McKhann et al., 1984; Fearing et al., 2007], Table 1.1 and Table 1.2.

Neither the volume of the whole brain nor that of some of its parts are the same in every person of the same age and sex, so using only the values of volumes without taking into accounts those factors might be inadequate. Several normalization techniques can be used to

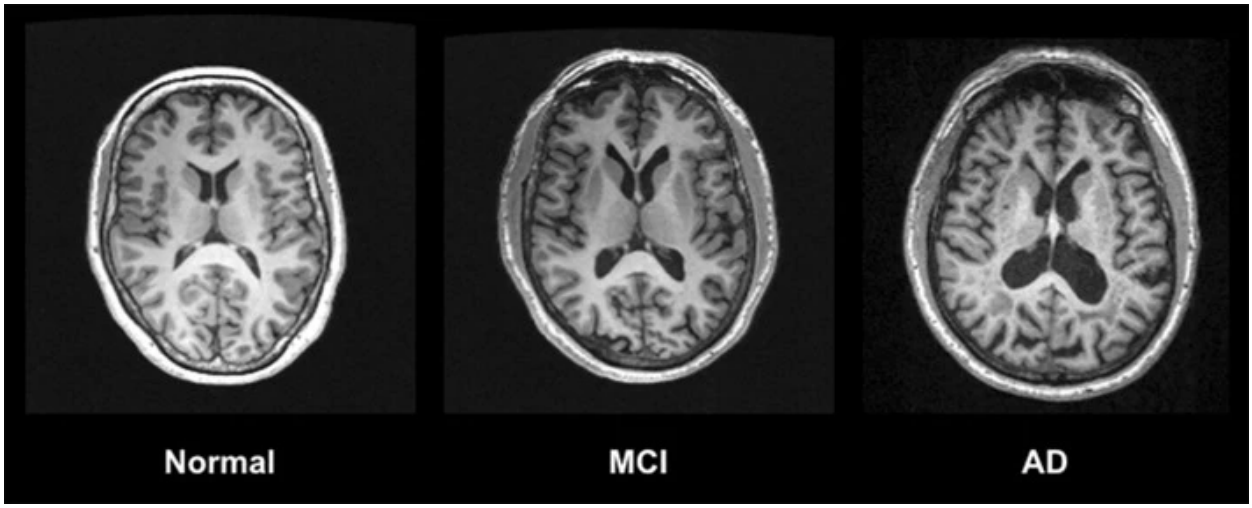


Figure 1.3: Comparison between the MRI images of the brains of a Cognitively Normal person (left), a Mild Cognitive Impairment subject (middle) and an Alzheimer's Disease individual (right). Source: Chandra et al. [2019]

counter this. For example, using the Intracranial Volume (ICV) to adjust for different head sizes is safe [Jenkins et al., 2000], or care has to be taken [Voevodskaya et al., 2014].

Neuroimaging has played different roles in the last four decades and each technique has its own advantages and disadvantages [Johnson et al., 2012]. There is an abundant quantity of neuroimaging techniques, which have been widely used not only in the diagnosis and prognosis of AD [Barkhof et al., 2011; Frisoni et al., 2010; Frizzell et al., 2022; Suárez-Araujo et al., 2024] but also of other diseases [Litjens et al., 2017; Zhang et al., 2021]:

1.3.1.4.1 Computed Tomography Computed Tomography (CT) is based on X-rays technology taken from many angles so it provides cross-sectional images (that is, that looks like "virtual slices") of the body. Radiation exposure and contrast reaction are the main inconveniences of this technology.

Other types of CT are [Davison and O'Brien, 2013]:

- Positron Emission Tomography is based on nuclear technology as the system detects a radionuclide ligand, paragraph 1.3.1.4.5, previously administered to the patient. It allows observing metabolic activity of parts of the body. It is, by far,the most popular CT technique [Mathotaarachchi et al., 2017; Nakamura et al., 2018b; Palmqvist et al., 2015; Zhou et al., 2014].
- Single-Photon Emission Computed Tomography (SPECT) is based on gamma rays and provides 3D and cross-sectional images. It requires that a gamma-emitting radioligand is given to the patient [Livieratos, 2012]. Sensitivity and specificity values of SPECT lower than PET were reported in [Davison and O'Brien, 2013]

1.3.1.4.2 Brain signals: Electroencephalography and Magnetoencephalography While being probably the simplest and more inexpensive neuroimaging technique, Electroencephalography (EEG) is still useful in differential diagnosis of dementia, and from other diseases [Ronson, 2011]. It monitors electrical activity of the brain according to different cerebral signals measured from diverse parts of the scalp. EEG is commonly applied with multiple non-invasive electrodes placed on the scalp [McBride et al., 2014], albeit invasive ones

exist too. It is prone to artifacts (wrong signals) from both non-cerebral or environmental origins. Unlike other methods, EEG can be used while performing other activities due to its more portable nature. However, it is generally used while the patient is in a resting state [Cassani et al., 2018].

Magnetoencephalography (MEG) maps brain activity by recording magnetic fields, which are naturally generated by electrical currents in the brain, with magnetometers. Due to the low values of these magnetic fields, MEG needs to be measured in magnetically shielded rooms in order to avoid interferences from, among others, the Earth's magnetic field [Alberdi et al., 2016]. This technique provides better spatial resolution and is more sensitive to superficial cortical activity than EEG. MEG is not dependent on head geometry, unlike EEG. Lower results due to its reduced discriminative potential might be the reason for its low popularity compared to EEG [Alberdi et al., 2016].

Several MEG signatures can be useful as unique biomarkers for the predementia stages of AD [Nakamura et al., 2018a]. These authors found that increase in the alpha band power in medial temporal regions was associated with Aβ deposition. Also, delta band power augmentation in the same areas was correlated with decrease in glucose metabolism and entorhinal atrophy, and related to AD progression. They observed that global theta power increase was not specific to AD due to the fact that it also happened without Aβ deposition, but with hippocampal atrophy and general cognitive decline.

1.3.1.4.3 Magnetic Resonance Imaging MRI is based on strong magnetic fields so patients are not exposed to ionizing radiation. It is not recommended to be used by patients with implants or non-removable internal metallic objects. This technology does not require compounds, although some contrast agents are used in order to get better resolution. MRI can be affected by visual artifacts, which can be produced by the patient, the device, or the signal processing [Erasmus et al., 2004]. Therefore, some kind of standardized protocols are required and similar equipments are recommended when data from different sites are used, especially when MRI images are analyzed with automated software, which are less tolerant to visual artifacts [Jack et al., 2008]. MRI techniques are usually divided into Structural Magnetic Resonance Imaging (sMRI) and Functional Magnetic Resonance Imaging (fMRI), whether they report the brain anatomical structures or activity, respectively. fMRI is based on the blood flow as higher flow indicates that certain part is in use at that moment.

Arterial Spin Labeling (ASL) is a fMRI technique where magnetic spin of arterial blood water is tracked and blood flow is measured. It does not need contrast agents as water is used for that purpose.

On the other hand, quantitative MRI can be seen as a complementary way of using MRI data where, apart from the "qualitative" visual-spatial information, each voxel of the image comes with a value of certain physicochemical, pre-selected, parameter [Bonny, 2005].

MRI, especially the structural one, is the most common diagnostic tool in the last decade [Cabrera-León et al., 2024c].

Diffusion Tensor Imaging (DTI) is a special kind of Diffusion-Weighted Magnetic Resonance Imaging (DW-MRI), where white matter connectivity of the brain is mapped. DW-MRI is based on the diffusion of water within tissues.

1.3.1.4.4 Functional Near-Infrared Spectroscopy Also known as "near-infrared imaging" or "optical topography" [Ferrari and Quaresima, 2012], or even simply as "near-infrared spectroscopy", Functional Near-Infrared Spectroscopy (fNIRS) is a functional neuroimaging

technique that uses near-infrared light to roughly calculate cortical hemodynamic activity produced by neural activity. This is done by quantifying changes in the concentrations of both deoxyhemoglobin and oxyhemoglobin. fNIRS is only able to gauge shallow cortical regions, although deep brain structures are possible on babies [Ferrari and Quaresima, 2012]. It is as non-invasive as EEG, and its cost of the equipment and portability depends on the fNIRS technology [Ferrari and Quaresima, 2012]. fNIRS has similar or slightly worse spatial resolution than EEG, whereas its temporal one is worse [Takeda et al., 2015].

Not as many researches were found that used fNIRS [Zhang et al., 2023a; Park, 2024; Yang et al., 2019, 2020]. As in some variants of EEG, fNIRS can be used while the patient performs different mental tasks [Yang et al., 2019, 2020]. In the former work they discovered that some subregions instead of the whole prefrontal cortex allowed Convolutional Neural Networks (CNNs) to achieve better performance values.

1.3.1.4.5 Compounds for neuroimaging methods In order to ease or even to allow that some receptor (that is, a particular substance, cell type, or protein, among others, that wants to be detected) become visible to a neuroimaging technology, a compound may be required. The way a compound achieves this is by pairing (also called, sticking or binding) to the receptor, hence called "ligand". Several compounds exist nowadays:

- Fluorodeoxyglucose (FDG) is an analog of glucose that is used as a tracer in PET scans. Concentrations of FDG indicate metabolic process in certain parts of the body while those parts are consuming glucose.
- Pittsburgh compound B (PiB) is a radioligand analog of thioflavin. It binds well with Aβ plaques, and it is used in PET.
- Florbetapir F 18 amyloid (AV-45) is a PET scanning radiopharmaceutical compound.

1.3.1.5 Human posture, gait and body parts movement

Some illnesses, such as AD, Parkinson's Disease (PD), stroke, and cerebral palsy, negatively affect the ability of the patients to keep certain posture and to control their body parts [Cabrera-León et al., 2024a]. In extreme cases, these diseases even impede it.

Individuals with greater cognitive impairment are characterized with slower gaits [Laske et al., 2015]. These authors recommended to analyze gait while the subject is performing another cognitive or motor task due to the fact that gait problems might be easier to detect as they tend to stay unnoticed when only one task is done [Laske et al., 2015]. In [Gillain et al., 2016] gait speed and variability were considered as possible markers of MCI patients converting to AD. The latter metric may be a better predictor of cognitive decline in CN subjects [Byun et al., 2018].

Many families of sensors and methodologies have been used to analyze human posture, gait and movements [Wong et al., 2007], among others photogrammetry, video analysis, eye trackers, accelerometers and gyroscopes. Thanks to the availability of smartphones and webcams, these methodologies have become more popular [Cabrera-León et al., 2024a], albeit more specialized equipment and controlled environments might be required for some experiments. Teaching the caregivers or the patients might also be needed when these devices are going to be used at home or in other unsupervised environments [Cabrera-León et al., 2024a].

1.3.1.6 Speech

As cognition gets impaired by AD, so does language [Pulido et al., 2020]. Therefore, writing and speech are negatively affected so the patients cannot communicate with or even understand others [Cabrera-León et al., 2024a].

Speech impairment increases gradually with the stage of AD [Cabrera-León et al., 2024a]:

- Early: bad word recalling, poor verbal fluency, and common use of word fillers.
- Moderate: impossibility to follow a conversation, word repetition, and incorrect use of words.
- Severe: incoherent speech, repeating other people, and use of unrevealed or illogical words or sentences.

Different characteristics of writing and speech can help in the diagnosis of AD and MCI [Gosztolya et al., 2019]. Bigger phonation and reading times are brought along by cognitive decline [Ivanova et al., 2022]. The quantities of voice breaks and periods of voice can help differentiate between individuals with AD, those with MCI and those with CN [Meilán et al., 2014].

Therefore, speech may be considered an early indicator of AD. After being analyzed with Machine Learning (ML) techniques during the last decade, it has been concluded that speech may be useful as a fast, cheap and non-invasive complementary diagnostic method [Pulido et al., 2020].

1.3.1.7 Blood

Blood-based biomarkers for AD are one of the most recent diagnostic tools [Cabrera-León et al., 2024a], and they have lower cost and are less invasive than CSF for obtaining APOE and serum Aβ. Their use in a primary care setting was recommended in [Leuzy et al., 2022], where different mixes of blood biomarkers were also advised depending on the AD stage. For preclinical cases, Aβ42/40 or glial fibrillary acidic protein combined with plasma P-tau217. For prodromal ones, cognitive scales and plasma P-tau217. The latter alone can be used for patients with AD dementia.

1.4 Prevalence and costs of dementia

Patients with dementia, including AD, have high socioeconomic costs for health systems. For example, annually the cost of each demented patient is more than 28,000 € in the Canary Islands (Spain) [Lopez-Bastida et al., 2006], ranging between 27,000 € and 37,000 € in all Spain [Prieto Jurczynska et al., 2011], where around 87% of these direct and indirect costs are borne by the families of these patients. However, this cost changes depending on the severity level of the disease, going from more than 18,000 € in mild stages to more than 52,000 € in severe ones [Prieto Jurczynska et al., 2011].

The incidence of AD grows with age, even doubling every 5 to 10 years [Zhu and Sano, 2006]. Also, its prevalence raises exponentially with age: almost 50% in individuals older than 84 years old, whereas 3% in those between 65 and 74 [Zhu and Sano, 2006]. It was also found that prevalence and incidence rates increase with age in the European population [Niu et al., 2017]. In 2010 in the Unites States 14.7% of people older than 70 years old had dementia, and the cost of each of these people was up to $56,290 per year [Hurd et al., 2013].

In 2015 the cost per person with dementia varied depending on the country where that patient lived: $43,680 in G7 countries[3], $20,187 in G20 countries[4], and $6,757 in non-G7 and non-G20 countries [Wimo et al., 2017].

Between 1995 and 2005, the prevalence of AD in Europe raised, unlike its incidence [Niu et al., 2017]. Both the prevalence and the incidence of AD were greater in women (7.13% and 13.25 per 1000 person-years) than in men (3.31% and 7.02 per 1000 person-years). In 2018 around 5.7 million US inhabitants had Alzheimer's dementia [Alzheimer's Association, 2018].

Besides, dementia brings along high burdens to both the physical and mental health of caregivers [Prieto Jurczynska et al., 2011]: stress, insomnia, anxiety, depression, feeling of social isolation, higher intake of psychotropic medicines, and worse self-perceived health status.

According to Nichols et al. [2022], in 2019 there were 57.4 million people with dementia globally and they estimated that in 2050 there will be 152.8 million cases, over 2.66 times more. In Spain these quantities are 826,686, 1,516,523 and 1.83, respectively. Sex prevalence will remain stable, with more women with dementia as currently occurs: a female-to-male ratio of 1.69. This increase in people with dementia will not be the same in all regions [Nichols et al., 2022]: the largest percentage changes in the number of projected dementia will be in north Africa and the Middle East and eastern sub-Saharan Africa, whereas the smallest in high-income Asia Pacific and western Europe. Such changes can be attributed to population growth (especially in sub-Saharan Africa) and population aging (notably in east Asia).

1.5 Neural computing methods for detecting Alzheimer's Disease and Mild Cognitive Impairment

1.5.1 History and recent advances

ANNs have been widely applied in very different fields such as spam filtering [Cabrera-León et al., 2018c,b] and chemicals detection [García Báez et al., 2011]. Regarding the topics of this thesis, in the last 30 years, ANNs have been used in the analysis, diagnosis, early detection and outcome prediction of many psychiatric disorders and health problems such as dementia, AD, myocardial infarction, appendicitis, several types of tumors, and diabetes [Galletly et al., 1996]. Examples of early researches on the diagnosis of dementia with ANNs are the following. In the case of dementia, Mulsant and Servan-Schreiber [Mulsant and Servan-Schreiber, 1988] implemented a connectionist neural network with 2 hidden layers which, in spite of being able to request more information about a particular attribute of a patient, could only correctly classify 61% of the cases. One of the first cases where neuroimaging was used together with an ANN for the CN-AD classification problem was one work of Kippenhan at al. [Kippenhan

[3]The Group of Seven (G7) is formed by the seven countries with the largest economies in the world: Canada, France, Germany, Italy, Japan, the United Kingdom, and the United States.

[4]The Group of Twenty (G20) is composed of governments and central banks of the countries with the most industrialized or developing economies in the world: Argentina, Australia, Brazil, Canada, China, the European Union, France, Germany, India, Indonesia, Italy, Japan, Mexico, Russia, Saudi Arabia, South Africa, South Korea, Turkey, the United Kingdom, and the United States.

et al., 1992]. Their backpropagation ANN achieved up to 0.85 AUC whereas 0.89 a human expert. An incremental learning, unsupervised multilayer neural network was another pioneer in terms of using neuroimaging and an ANN [Chan et al., 1993]. It used cerebral blood flow obtained via SPECT and had problems distinguishing CN from AD.

In the recent years, more complex ANNs have appeared, although, apart from Deep Learning (DL) techniques with neuroimaging, they have not become preponderant. Thanks to the availability of accessible datasets, such as Alzheimer's Disease Neuroimaging Initiative (ADNI)[5] and Open Access Series of Imaging Studies (OASIS)[6], more researchers have become interested in AD, and the reproducibility of their experiments is far easier than before, when private datasets were common.

In Table 1.7, Table 1.8, Table 1.9 and Table 1.10 we can see examples of other authors' works, the characteristics of the system and the results obtained, grouped by the type of model: CNNs, other DL methods and shallow ANNs, respectively.

[5] http://adni.loni.usc.edu/

[6] https://www.oasis-brains.org/

Table 1.7: Characteristics and results of Deep Learning methods of the CNN family for detecting AD and MCI.

Method	Dataset	Nº of subjects	Features	Best results	Research
CNN	ADNI	100 CN, 76 pMCI, 128 sMCI, 93 AD	MRI, PET.	0.96 AUC, 0.93 A in CN-AD; 0.88 AUC, 0.83 A in CN-pMCI	[Liu et al., 2018a]
		229 CN, 167 pMCI, 236 sMCI, 198 AD	MRI, PET.	0.93 AUC in CN-AD; 0.73 AUC in s-MCI-pMCI; 0.76 AUC in MCI-NC	[Li et al., 2014]
		161 CN, 193 MCI, 161 AD	MRI, up to 35 neuropsychological test scores	$<$ 0.75 A in CN-MCI; $>$ 0.75 A in AD-MCI; $<$ 0.8 A in CN-AD	[Senanayake et al., 2018]
		61 CN, 77 EMCI, 43 LMCI, 50 AD	MRI	0.66 AUC, 0.64 A in AD-EMCI; 0.61 AUC, 0.62 A in AD-LMCI; 0.88 AUC, 0.79 A in CN-AD; 0.47 AUC, 0.56 A in LMCI-EMCI; 0.67 AUC, 0.63 A in CN-LMCI; 0.57 AUC, 0.54 A in CN-EMCI	[Korolev et al., 2017]
	OASIS	74	MRI	0.75 A	[Alkabawi et al., 2017]
	CADDementia, ADNI	70 CN, 70 MCI, 70 AD.	sMRI.	0.948 A in AD-MCI-CN; 0.957 A in (AD+MCI)-CN; 0.993 A in AD-CN; 1 A in AD-MCI; 0.942 A in MCI-CN	[Hosseini-Asl et al., 2018]
CNN+CAE	ADNI	229 CN, 199 AD	MRI	0.93 AUC, 0.88 A, 0.91 Se, 0.84 Sp in CN-AD	[Li et al., 2017]
CNN+MNLR	Chinese dataset	120 CN, 120 MCI, 110 AD	ASL (fMRI)	Median of diagnostic errors of DL approach is less than other ML methods'	[Huang et al., 2017]

A = Accuracy; ASL = Arterial Spin Labeling; CAE = Convolutional Autoencoder; DL = Deep Learning; fMRI = Functional MRI; MNLR = Multi-Nominal Logistic Regression; MRI = Magnetic Resonance Imaging; pMCI = MCI converters; Se = Sensitivity; sMCI = MCI non-converters; sMRI = Structural MRI; Sp = Specificity

Table 1.8: Characteristics and results of Deep Learning methods other than CNNs for detecting AD and MCI.

Method	Dataset	Nº of subjects	Features	Best results	Research
SAE	Sydney Memory and Aging Study	505 CN, 332 MCI	Up to 35 neuropsychological test scores	0.85 A, 0.89 AUC in CN-MCI	[Senanayake et al., 2017]
	ADNI, CADDementia	171 CN, 232 MCI, 101 AD	sMRI	Up to 0.58 A in CN-MCI-AD	[Dolph et al., 2017]
	ADNI	52 CN, 43 pMCI, 56 sMCI, 43 AD.	MRI, PET, CSF, MMSE, ADASCog	0.9 A in CN-AD; 0.74 A in MCI-CN; 0.6 A in pMCI-sMCI	[Suk and Shen, 2013]
		77 CN, 67 pMCI, 102 sMCI, 65 AD.	MRI	0.88 A, 0.89 Se, 0.87 Sp in CN-AD; 0.77 A, 0.74 Se, 0.78 Sp in MCI-CN	[Liu et al., 2014]
RNN	TADPOLE (ADNI)	N/A	ADASCog13, ventricles volume	0.86 AUC	[Nguyen et al., 2018]
MLP+RNN	ADNI	229 CN, 198 AD.	sMRI	0.9 A, 0.87 Se, 0.93 Sp in CN-AD	[Cui et al., 2018]

A = Accuracy; ADAS = Alzheimer's Disease Assessment Scale; ADNI = Alzheimer's Disease Neuroimaging Initiative; CSF = Cerebrospinal Fluid; MMSE = Mini-Mental State Examination; MRI = Magnetic Resonance Imaging; PET = Positron Emission Tomography; pMCI = MCI converters; RNN = Recurrent Neural Network; SAE = Stacked Autoencoder; Se = Sensitivity; sMCI = MCI non-converters; sMRI = Structural MRI; Sp = Specificity

Table 1.9: Characteristics and results of shallow neural computing methods for detecting AD and MCI (Table 1 of 2).

Method	Dataset	Nº of subjects	Features	Best results	Research
HUMANN-S	Private	267 consultations on 30 patients	MMSE, FAST, Katz's index, Barthel's index, and Lawton-Brody's index	0.87 Se, 0.93 Sp with AD	[García Báez et al., 2010]
				0.88 Se, 0.82 Sp with mild CI; 0.97 Se, 0.96 Sp with severe CI	[García Báez et al., 2009]
				0.96 Se, 0.93 Sp with AD	[García Báez et al., 2012]
Ensembles of HUMANN-S				0.97 Se, 0.9 Sp with AD	[García Báez et al., 2010]
				0.78 Se, 1 Sp with mild CI; 1 Se, 0.99 Sp with severe CI	[García Báez et al., 2009]
				0.97 Se, 0.9 Sp with AD	[García Báez et al., 2012]
CPN	ADNI	166 CN, 159 EMCI, 137 LMCI.	Age, years of education, FAQ, MoCA, NPI	0.657AUC, 0.545 A in CN-EMCI-LMCI	[Cabrera-León et al., 2018a]
		203 CN, 128 MCI.	Age, years of education, FAQ, MMSE, GDS	0.88 AUC, 0.88 A, 0.8 Se, 0.94 Sp in CN-MCI	[García Báez et al., 2015]
				0.95 AUC, 0.87 A, 0.9 Se, 0.85 Sp in CN-MCI	[Suárez Araujo et al., 2017]

A = Accuracy; EMCI = Early Mild Cognitive Impairment; FAQ = Functional Activities Questionnaire; FAST = Functional Assessment Staging scale; GDS = Geriatric Depression Scale; HUMANN = Hybrid Unsupervised Modular Adaptive Neural Network; LMCI = Late Mild Cognitive Impairment; MMSE = Mini-Mental State Examination; MoCA = Montreal Cognitive Assessment; NPI = Neuropsychiatric Inventory; P = Precision; Se = Sensitivity; Sp = Specificity

Table 1.10: Characteristics and results of shallow neural computing methods for detecting AD and MCI (Table 2 of 2).

Method	Dataset	Nº of subjects	Features	Best results	Research
Linear feed-forward	AlzBiomarker	35 CN, 45 AD	From CSF: albumin ratio, Aβ40, Aβ42, Total tau, tau-phospho	0.955 Se, 0.91 Sp in CN-AD	[Aljović et al., 2016]
ANN, DT, NB, UCS	"nep499" (Rdatasets)	242 CN, 257 AD	8 SNP, APOE	1 P, 0 FPR, 1 TPR	[Salazar et al., 2018]
BP-ANN	Chinese dataset	60 CN, 60 AD	Age, Activities of Daily Living, aluminium, creatinine, 5-hydroxytryptamine, dopamine	0.93 AUC, 0.925 A, 0.9 Se, 0.95 Sp	[Tang et al., 2013]
	OASIS	135 CN, 69 very mild AD, 29 mild AD, 2 moderate AD.	MRI (hippocampus), age, gender, education, socio-economic status, MMSE	0.87 A	[Raut and Dalal, 2017]
2-hidden layers ANN	ADNI	225 CN, 358 MCI, 138 AD	MRI (volumes of 16 ROIs), SNP	0.992 AUC in CN-AD; 0.835 AUC in MCI to AD conversion	[Ning et al., 2018]
RF, NB, MLP, k-NN, BP-ANN...	Private	40 CN, 40 Late Onset AD	30 biochemical and genetic biomarkers	0.94 A in CN-AD	[Coppedè et al., 2013]

A = Accuracy; AUC = Area Under the Curve; BP-ANN = Backpropagation Artificial Neural Network; CI = Cognitive Impairment; CSF = Cerebrospinal Fluid; FPR = False Positive Rate; k-NN = k-Nearest Neighbors; MLP = Multilayer Perceptron; MMSE = Mini-Mental State Examination; MRI = Magnetic Resonance Imaging; NB = NB; Open Access Series of Imaging Studies = OASIS; P = Precision; RF = Random Forest; ROI = Region of Interest; Se = Sensitivity; SNP = Single Nucleotide Polymorphism; Sp = Specificity; TPR = True Positive Rate; UCS = Supervised Classifier System

1.5.2 Surveys, overviews and reviews

Several similar surveys, overviews and reviews exist in the literature, mainly related with AD diagnosis. The number of such review-like articles has increased greatly in the last decade, even doubling each year between 2021 and 2023. Those that included works related with prognosis and longitudinal studies were uncommon [Cassani et al., 2018; Liu et al., 2018b; Jo et al., 2019; Arya et al., 2023; Khaliq et al., 2023]. It was rare to find reviews that analyzed a very limited number of articles [Bevilacqua et al., 2023; Crary, 2023]. Different types of analysis were carried out by some researchers: bibliometrically [Liu et al., 2023], bibliographically [Wu et al., 2023], or with visibility graphs [Sulaimany and Safahi, 2023]. The extended study in [Tsoi et al., 2023] studied applications of Artificial Intelligence (AI) for cognitive screening and training, dementia diagnosis and prognosis, and care and treatment. The main differences between reviews were related to the illnesses, modalities and computational methods that were analyzed.

According to illnesses, reviews where solely AD-related works were included were extremely common [Alberdi et al., 2016; Rathore et al., 2017; Cassani et al., 2018; Liu et al., 2018b; Jo et al., 2019; Pulido et al., 2020; Logan et al., 2021; Zeng et al., 2021; Zhao et al., 2021; Frizzell et al., 2022; Gao and Lima, 2022; Mirzaei and Adeli, 2022; Qu et al., 2022; Saleem et al., 2022; Shastry et al., 2022; Zhang et al., 2023c; Cabrera-León et al., 2024c; Malik et al., 2024; Ozkan et al., 2024; Suárez-Araujo et al., 2024; Vimbi et al., 2024]. Diagnosis and prognosis of MCI was found in [Liu et al., 2023], whereas prediction of MCI conversion in [Arya et al., 2023]. Few of them also studied non-AD dementia works [Zheng et al., 2015; Battineni et al., 2022; Li et al., 2022], and some focused on AD and PD [Alfalahi et al., 2023; Khaliq et al., 2023]. Reviews where several illnesses were analyzed were also found [Litjens et al., 2017; Goceri and Goceri, 2017; Noor et al., 2020; Zhang et al., 2020, 2021; Ardalan and Subbian, 2022; Iqbal et al., 2024], including cognitive impairment detection [Fei et al., 2017; Pellegrini et al., 2018; Bevilacqua et al., 2023], different neurological disorders [Zhang et al., 2023b], several degenerative nerve diseases [Bhachawat et al., 2023], or neurodegenerative neuropathologies [Crary, 2023]. Sulaimany and Safahi [Sulaimany and Safahi, 2023], by means of visibility graphs, wrote a scope review in neuroscience in general.

Regarding modalities, the majority only discussed works that used neuroimaging [Rathore et al., 2017; Pellegrini et al., 2018; Jo et al., 2019; Zhang et al., 2020; Logan et al., 2021; Zeng et al., 2021; Ardalan and Subbian, 2022; Mirzaei and Adeli, 2022; Aberathne et al., 2023; Arya et al., 2023; Zhao et al., 2023; Ozkan et al., 2024]. Medical imaging, which also includes neuroimaging, were reviewed by fewer authors [Zheng et al., 2015; Litjens et al., 2017; Goceri and Goceri, 2017; Gao and Lima, 2022], some even with radiomics from medical images [Bevilacqua et al., 2023]. In [Liu et al., 2018b] multimodality imaging were studied. Multimodality articles where neuroimaging together with other modalities were analyzed in [Mirzaei and Adeli, 2022], whereas multimodality of any type in [Alberdi et al., 2016; Elazab et al., 2024; Suárez-Araujo et al., 2024]. Few reviews focused solely on MRI [Noor et al., 2020; Zhang et al., 2021; Frizzell et al., 2022] or sMRI [Zhao et al., 2021]. Works based on non-neuroimaging techniques were reviewed in [Cabrera-León et al., 2024a]. EEG works were collated in [Cassani et al., 2018; Liu et al., 2023]. Articles about automatic speech analysis for AD were reviewed in [Pulido et al., 2020]. Assessment of spontaneous speech was studied in [Parsapoor, 2023; Qi et al., 2023], and also of language disorders in [Parsapoor, 2023]. Works that extracted data from wearables or sensors were studied in [Alfalahi et al., 2023]. Researches that used data from non-invasive and portable techniques were summarized in

[Zhang et al., 2023c].

Finally, many reviews only included studies based on deep neural architectures [Litjens et al., 2017; Goceri and Goceri, 2017; Jo et al., 2019; Noor et al., 2020; Zhang et al., 2020; Logan et al., 2021; Gao and Lima, 2022; Saleem et al., 2022; Iqbal et al., 2024]. However, many others preferred wider families, whether AI [Liu et al., 2018b; Zhao et al., 2021; Zhang et al., 2021; Battineni et al., 2022; Frizzell et al., 2022; Li et al., 2022; Aberathne et al., 2023; Tsoi et al., 2023; Zhang et al., 2023c], ML [Zheng et al., 2015; Alberdi et al., 2016; Cassani et al., 2018; Pellegrini et al., 2018; Ardalan and Subbian, 2022; Mirzaei and Adeli, 2022; Arya et al., 2023; Bhachawat et al., 2023; Khaliq et al., 2023; Parsapoor, 2023; Zhao et al., 2023; Elazab et al., 2024; Malik et al., 2024], or ANNs [Cabrera-León et al., 2024a; Suárez-Araujo et al., 2024]. On the contrary, other authors narrowed the analysis to specific neural methods, such as Generative Adversarial Network (GAN)-based DL methods [Qu et al., 2022], Graph Neural Networks (GNNs) [Zhang et al., 2023b], or Local Interpretable Model-agnostic Explanations (LIME) and SHapley Additive exPlanations (SHAP) frameworks [Vimbi et al., 2024]. Others took different approaches, such as comparing Deep Neural Networks (DNNs) with other ML or DL modular systems [Shastry et al., 2022], or studying the importance of transfer learning to DNNs [Ardalan and Subbian, 2022].

Among the most recent reviews, two that we have authored, whether published or under review, will be summarized in the next two subsections [Cabrera-León et al., 2024a, 2025]. Researches included in any of them made use of ANNs, DNNs, a combination of any of them or with other ML models. Works were grouped into neuroimaging and non-neuroimaging because, in a prospective bibliography collation, it was found that the popularity of the neural computation methods, the databases used and use of multimodality varied a lot depending on these families of modalities.

1.5.2.1 State of the art of works based on non-neuroimaging techniques

In [Cabrera-León et al., 2024a] 42 works based on non-neuroimaging techniques were reviewed. They were extracted from the PubMed database and a selection criteria was required to be complied for the articles to be included.

Figure 1.4 shows that the number of cross-sectional studies on this topic that fulfill the selection criteria has almost multiplied by 5 in the last two years. There has been almost the same number of studies based on deep and shallow ANNs in the last two years. Genes or blood have become more frequently used by ANN-based methods in that period of time. Conversely, in the last decade modalities being used have changed, with novel ones appearing while others stayed with a similar level.

In Figure 1.5 how frequent the different classification tasks have been in the selected cross-sectional studies is shown. Almost 54% correspond to the CN-AD and CN-MCI classification tasks. Multimodality, with almost 32% of the cross-sectional studies, was more common for tackling binary classification problems or with shallow ANNs. Of the selected studies based on DNNs, none tackled MCI-AD.

Private databases were the most popular (over 52% of the times a database was used), followed by far by Gene Expression Omnibus (GEO) (almost 15%) and ADReSS (over 12%), Figure 1.9. ADNI and Pitt Corpus follow them with above 7% each. Reasons for such results might be derived from the characteristics of public and private databases: lower number of participants but more specific biomarkers and clinical criteria in the latter.

Cross-sectional studies were five times more common than longitudinal ones, and it

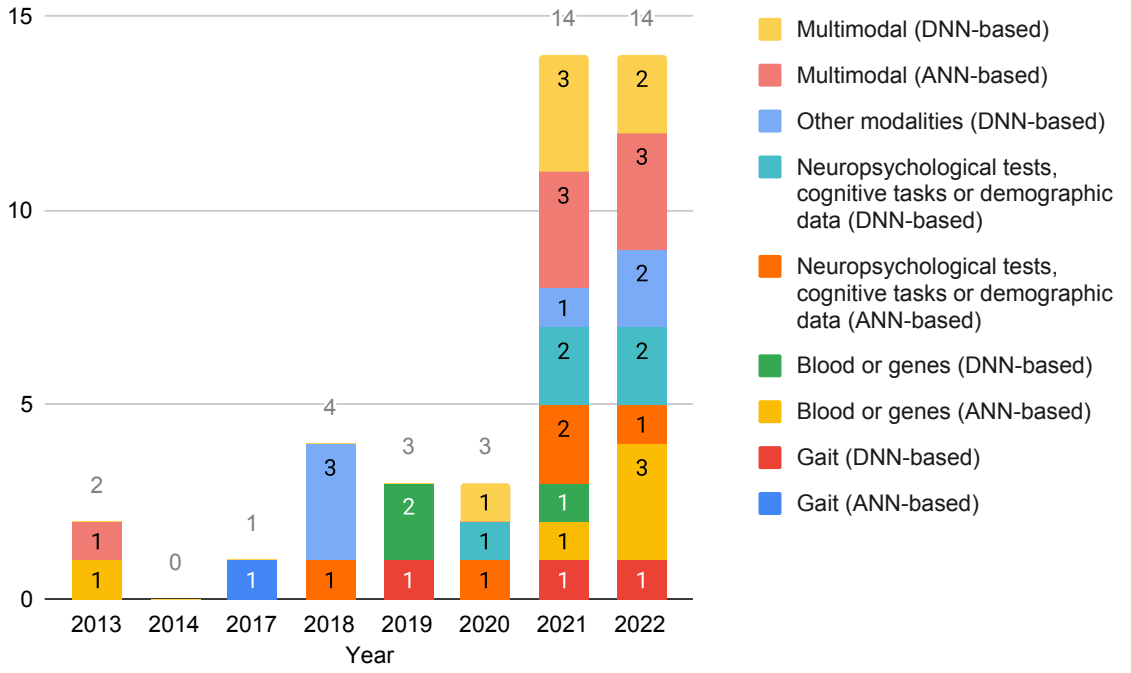


Figure 1.4: Quantity of cross-sectional studies that used data not from neuroimaging per year, grouped by modality and neural family. Source: Cabrera-León et al. [2024a]

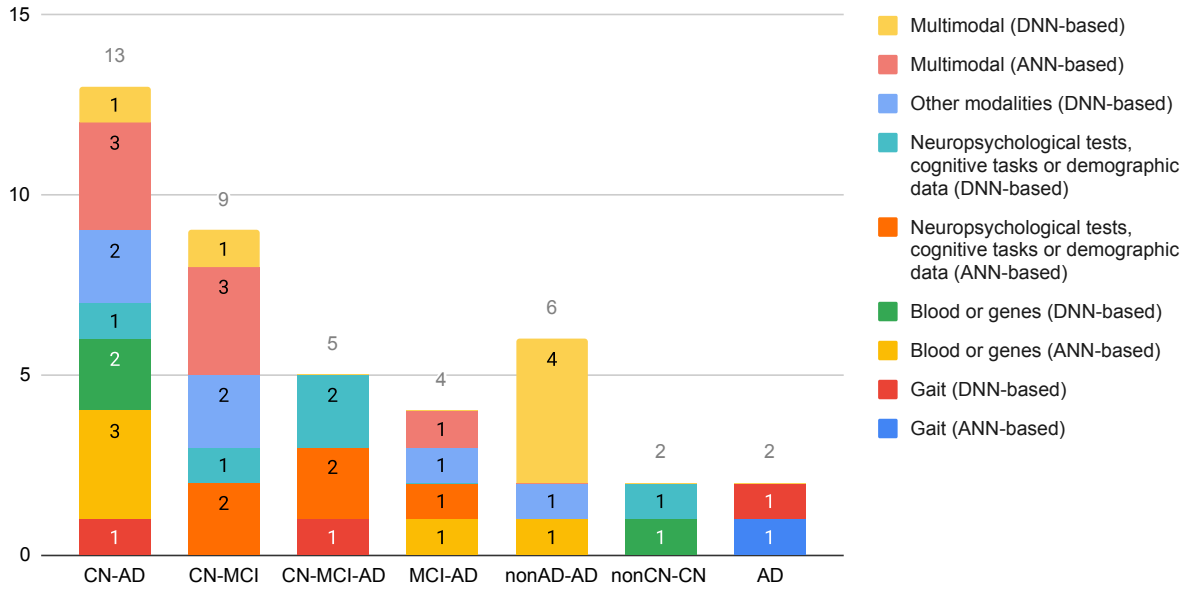


Figure 1.5: Quantity of cross-sectional studies that used data different to neuroimaging per classification task, grouped by modality and neural family. Source: Cabrera-León et al. [2024a]

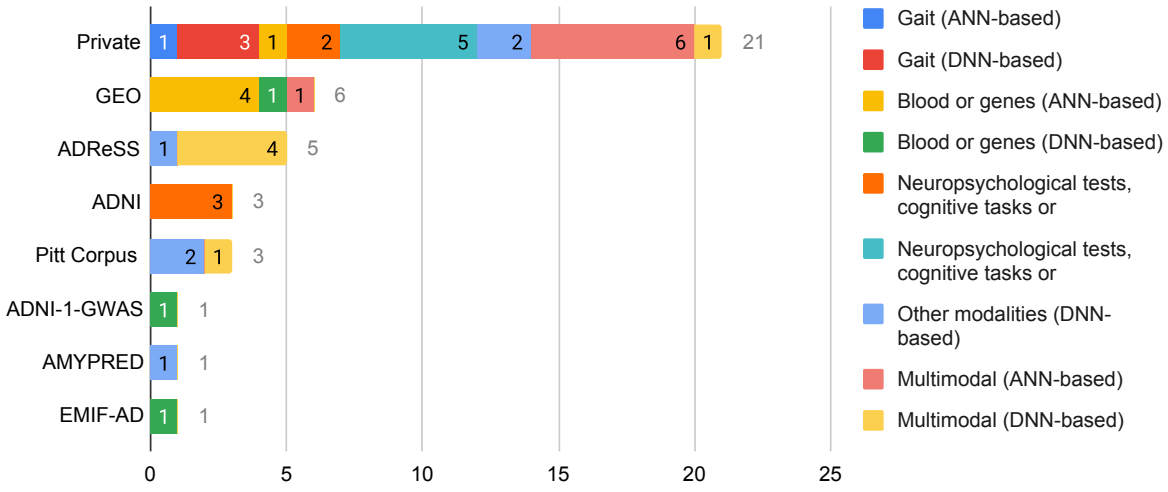


Figure 1.6: Quantity of times that a database was used in the selected literature for each combination of modality and neural family. Source: Cabrera-León et al. [2024a]

drastically increased in the last two years, Figure 1.4.

The quantity of works using shallow ANNs was similar to those using DNNs with all modalities but those labeled as "Others" (mainly related with speech and eye). Only DNNs, principally Recurrent Neural Networks (RNNs), were used in longitudinal studies. Transformers were common due to the usage of textual information and Natural Language Processing. CNNs were not as popular as with neuroimaging, subsubsection 1.5.2.2, yet they were used with gait and neuropsychological tests, or as modular approaches, combined with others.

Transfer learning, as a way to reduce training time by pretraining with a dataset different to the wanted one, was scarcely used, possibly as CNNs were not so frequent.

Many different clinical criteria were used, including novel ones: human gait, speech and blood biomarkers. Using more than one modalities was common in cross-sectional studies. All studies in [Cabrera-León et al., 2024a] but two longitudinal ones used neuropsychological scales.

Binary classification tasks were more prevalent, especially CN-AD, CN-MCI and nonAD-AD, than multiclass ones (principally CN-MCI-AD).

More than half of the cross-sectional studies used private databases, whereas a similar ratio with ADNI in longitudinal studies. Private databases were more prevalent when DNNs were used, whereas ADNI with neuropsychological scales and shallow ANNs.

Most researches worked with highly imbalanced datasets, and only undersampling techniques were used in a few of them to correct this.

Unlike in [Cabrera-León et al., 2025], data augmentation techniques were not used to increase the input data by fabricating new samples.

Imputation methods were rarely mentioned in the selected works in [Cabrera-León et al., 2024a], whether because it was non existent or the incomplete samples were just discarded. When used, they were based on changing the missing value with a meaningful value for the class of that incomplete sample.

Different performance metrics were used and almost all research works report values for more than one. In terms of performance metrics, DNNs performed as good as shallow ANNs.

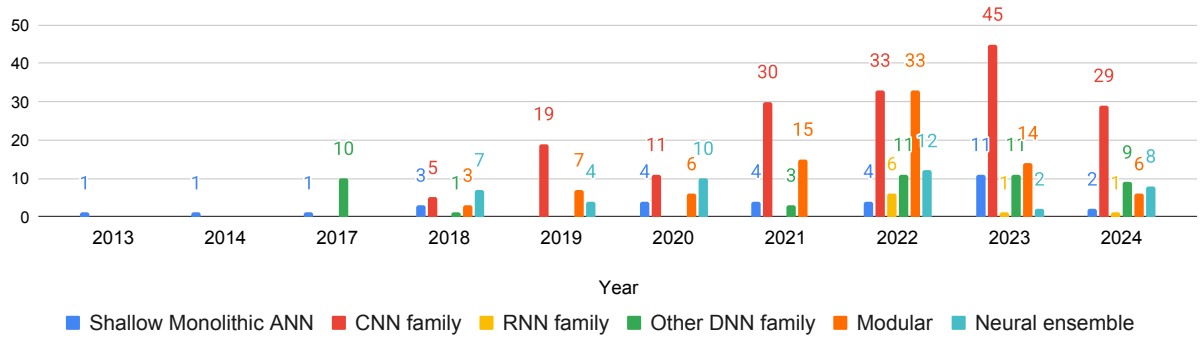


Figure 1.7: Quantity of cross-sectional studies that used neuroimaging data per year, grouped by neural family. Source: Cabrera-León et al. [2025]

1.5.2.2 State of the art of works based on neuroimaging techniques

In [Cabrera-León et al., 2025], 224 works based on neuroimaging techniques were reviewed. They were extracted from the PubMed database and needed to comply with a selection criteria.

In Figure 1.7 how the number of cross-sectional studies on this topic that fulfill the selection criteria has increased yearly throughout the last decade is shown this increasing trend can be expected for the whole year. The number of articles based on shallow ANNs and neuroimaging has stagnated, probably due to the superb performance of DNNs with this type of input data, whose high popularity can also be seen in this figure. Modular approaches became more common than CNNs during 2023. In 2024, modular approaches, neural ensembles and those based on other DNNs have become similarly popular.

Figure 1.8 shows how common the different classification tasks have been in the selected cross-sectional studies. Almost 54% correspond to the CN-AD and CN-MCI classification tasks. The CNN family has been the most frequently used neural method, followed by modular approaches, in all but one classification task.

Databases are sometimes not used alone but in combination with other private or non-private databases, mostly to check generalizability of the results or to cover each other shortcomings. In Figure 1.9 in the selected cross-sectional literature ADNI was the most popular database as it was used in almost 65.7% of the times that a database was utilized, followed by far by the heterogeneous group of private databases (above 11.5%) and OASIS (almost 7%). In the last two years, the Kaggle dataset has increased its popularity from the 11th to the 4th position. The reasons for these results might be deduced from the aforementioned characteristics of public and private databases. On the other hand, almost 89.66% of the prognosis or longitudinal studies used data from ADNI.

A lot less studies made use of shallow ANNs with neuroimaging data than with deeper ones, which can be explained by the superior results of the latter seen in the last decade. Shallow monolithic ANNs appeared in less than 1 out of 10 studies, being the most common ones Multilayer Perceptrons (MLPs) and Extreme Learning Machines (ELMs). A variant of a Wavelet Neural Network (WNN) yielded superb results in several classification tasks, sometimes outperforming the previous ones in some tasks. Use of MRI alone was higher than for the other modalities, and multimodality was more uncommon than with DL models. The ADNI database, being the most popular, allowed working on any of the tasks studied, which also occurred with the DNNs. As it currently lacks EEG data, other databases were utilized

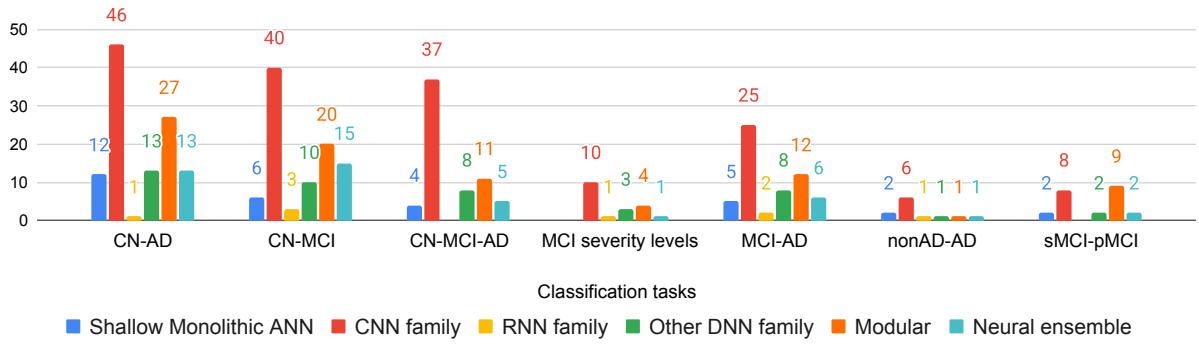


Figure 1.8: Quantity of cross-sectional studies that used neuroimaging data per classification task, grouped by neural family. Source: Cabrera-León et al. [2025]

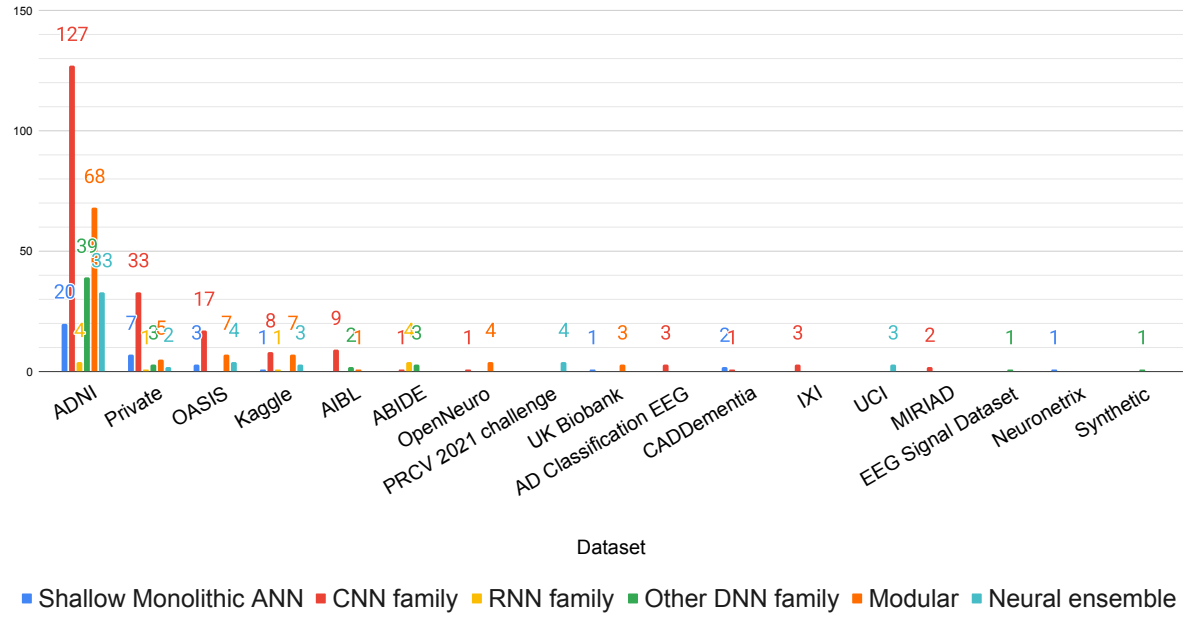


Figure 1.9: Quantity of times that a database was used in the selected literature for each neural family. Source: Cabrera-León et al. [2025]

when using this modality. The number of patients when EEG data, and private databases in particular, were used is several orders of magnitude lower than in any other case.

More than 4 of 10 cross-sectional studies used a monolithic model of the CNN family. More powerful CNN variants have started being used, some of which can overshadow the popular "vanilla CNN" ones such as Residual Network (ResNet) and Visual Geometry Group (VGG). Examples of models, whether pretrained via transfer learning or trained from scratch, were found, the latter being the most common, although authors that tested both almost always recommended the former. sMRI was preferred for any classification task, but several combinations of modalities were also tested. In one of the studies using EEG data, a CNN underperformed some simpler ML models, something deemed atypical considering the good results CNNs provide. This could be explained by the low number of samples used for training, as DNNs require lots of training samples and much more than the ML models. In the last two years, the use of Graph Convolutional Networks (GCNs) in this field has increased, including in modular approaches, as its ability to work with graph-structured data has proven useful for functional neuroimaging.

Among the other DNNs not belonging to the CNN family, which were scarcer, no significant differences were found between the performance of the models in this heterogeneous group. In the last two years, the usage of transformers has increased, probably due to their good performance with non-neuroimaging and the development of Vision Transformer (ViT) that is capable of using neuroimaging. Compared to CNNs, modalities did differ within the non-CNN family because those unrelated to sMRI were slightly more frequently used. Usage of fMRI with Long Short-Term Memory (LSTM) was found in several research and tackling different tasks. Similarly with multimodal data where fMRI was included.

The group of modular methods is the most heterogeneous of those tested, in both models and features. Generally, one of the modules is a CNN that works directly with the raw neuroimaging data, followed by another model, including non-neural ones such as Support Vector Machine (SVM) and Random Forest (RF), for classification purposes. When a module is located before the CNN, it is used for preprocessing the data or for data augmentation, such as in all the examples related to GANs. Several of these works were aimed to analyze the data augmentation power of such networks, including some cross-modality examples (that is, fabricating data for a modality different to the one the input data provided to the GAN belong to). The performance of modular approaches is often high and on par to that from monolithic CNNs.

Neural ensembles were mostly built with CNNs and used sMRI data. As it happened with modular approaches regarding CNNs, they yielded similarly good results despite the increased model complexity and training time required. A fusion of several SVMs outperformed an ensemble of CNNs in the deemed most difficult classification task.

Regarding longitudinal or prognosis studies, half of them used a monolithic model from the CNN family, also achieving the best results in any of the tasks dealt with. None was based on shallow ANNs. ADNI was the most popular database, especially in the tabulated state-of-the-art studies. Multimodality was more common in the prognosis or conversion prediction than in the classification tasks that made use of longitudinal data, which might be explained by its increased difficulty. MRI was the most frequently used and was able to tackle any of the classification tasks.

Barely any of the studies included in this review mentioned using techniques to handle missing data, in spite of being quite common in clinical settings. Of the three approaches commonly used to tackle this issue, filling empty values with plausible yet fabricated ones

obtained with imputation methods is generally considered the most optimal one as it did not decrease the quantity of data to work with. However, as the new values are not real, special caution should be taken when inferences based on these fabricated data are made.

Similarly, works in this review rarely used class balancing techniques, even though datasets with evident unbalanced classes were being utilized. Among the reasons for this scarcity, even though they were not clearly stated, might be the probable usage of the following methods to deal with this issue: using robust models, setting class weights, and changing the loss function. Another possibility is that the ample use of DNNs working with raw data in a way akin to brute force frequently brings along simple or no preprocessing steps, one of them being the balancing of the classes. Fabricating samples of the minority classes via oversampling methods has the same inconvenience as when using imputation techniques to handle missing data. Conversely, deleting samples of the majority classes via undersampling methods diminishes the number of data available to train with, similar to what occurs when discarding samples with missing values when tackling this problem.

Regarding datasets, in cross-sectional studies ADNI was used almost 6 times more frequently than the one in the second position, the private one. Kaggle being the fourth more used dataset might be explained by the popularity of Kaggle's competitions, and the free access and easier availability of this dataset as no data use agreement is needed. However, it should be noted that, unlike ADNI and OASIS, we consider that the Kaggle dataset should be considered inadequate for publication as it lacks information on data origin, who and how the labeling of the images was done, demographics, or the data preprocessing that was used. In prognosis and longitudinal studies almost all researchers made use of ADNI due to the high number of visits with multimodal data available.

CN-AD was the most common classification task and, due to being the easiest one, also the one where the best performance values were obtained. This happened with all the model families in the cross-sectional studies discussed. On the other hand, in a few cases the number of studies where a particular neural network family was used to tackle certain classification task was scarce. For example, a single study was found for nonAD-AD with shallow ANNs [Ruiz-Gómez et al., 2018], MCI severity levels with neural ensembles [Kiran et al., 2024], nonAD-AD with neural ensembles [Battineni et al., 2020], and CN-MCI with CNNs and using longitudinal data [Pena et al., 2019]. Low performance results might explain this unpopularity in all but the last two cases, where values above 0.97 and 0.84 were yielded, respectively. The latter two cases might be explained by the low complexity of those classification tasks when, respectively, an overpowered neural method or data from many visits are utilized. This can be inferred because both tasks have been efficiently dealt with by using less complex techniques and cross-sectional data.

Finally, it is interesting to note that, although in recent studies, essentially at a clinical level, the latest advances point towards a treatment of an AD continuum [Jack Jr. et al., 2024], in computational studies and specifically in all those analyzed in this review, the use of the conceptual construct MCI and its different levels, such as EMCI and LMCI, continues to dominate. The fact that most accessible data are built within the latter framework might explain its ample usage, since the idea of the AD continuum still needs stabilization and more generation of data.

Chapter 2

Neural Computing and Theoretical Foundations

In this chapter we will describe the computational basis of this PhD thesis, neural computation methods. Later, we will explain some theoretical foundations on these preprocessing steps: feature ranking, data scaling, and data projection. Finally, we will expose some metrics that can be used to evaluate a model.

2.1 Neural computation

2.1.1 Artificial Neural Networks

Several complex problems are extremely difficult or even impossible to be formulated as algorithms, often because they involve many subtle factors [Kriesel, 2007]. Humans tend to solve such problems by learning and estimating outcomes. For this reason, as computers cannot do that on their own, they require methods that have the capability to learn, without needing to explicitly program the problem. One of such methods is the neural computation family. Any system where the information is processed by networks of neurons — Artificial Neural Networks (ANNs) — belongs to the family of neural computation methods. It is a huge family[1] of Machine Learning (ML) techniques. ANNs are inspired by biological neural networks, principally the human brain, the cells that form it (neurons have been the most studied cells), and the interconnections between neurons [Fiesler and Beale, 1997; Kriesel, 2007; Hagan et al., 2014; Shalev-Shwartz and Ben-David, 2014].

As it happens with other ML approaches, ANNs can work as individual classifiers or the decisions of several of them or together with non-neural methods can be combined. Depending on the way they are combined and the classifiers that are used, many terms have been coined to differentiate them: "modular networks", "mix of classifiers", "ensembles", "committees", etc [Hastie et al., 2009; Polikar, 2006].

2.1.1.1 Biological inspiration

Neural computation is a discipline inspired by the Biological Neural Networks (BNNs) of the Central Nervous System (CNS) of animals and, particularly, of the human brain [Kohonen,

[1] http://www.asimovinstitute.org/neural-network-zoo/

1988; Basogain Olabe, 1998; Gabrielsson and Gabrielsson, 2006]. Mainly, researchers of several fields (including but not limited to engineering, philosophy, physiology and psychology) have focused on the analysis of the composition and behavior of the BNNs, the fundamental units that constitute them (the biological neuron), and the interconnection between those units (synapse). For his researches on these topics, the Spanish scientist and Nobel prize winner Santiago Ramón y Cajal is considered the father of modern neuroscience [Santiago Ramón y Cajal, 1917].

ANNs has been defined as networks of simple, normally adaptive, elements, which are massively connected in parallel and hierarchically organized, that try to interact with the real world in the same way as the biological CNS does [Kohonen, 1988]. They exhibit characteristics common to the brain [Jain et al., 1996; Basogain Olabe, 1998]:

- Learn: acquire knowledge based on experience and practice. ANNs are able to adapt their behaviors based on the environment, learning by adapting their synaptic weights [Rumelhart et al., 1986].

- Generalize: abstract what is common and essential to many things in order to form a general concept that encompass all of them. ANNs are partially tolerant to faults (due to redundancy), and to noises, distortions and incompleteness in the inputs.

- Abstraction: separate characteristics of an object by means of an intellectual operation in order to analyze it in isolation. Some ANNs are able to extract information that is implicit in the data.

A typical biological neuron, a.k.a. nerve cell[2], can be principally divided in the next three parts [Kriesel, 2007], Figure 2.1:

- Dendrites: extensions that transmits impulses received from adjacent neurons to the soma. Their quantity is variable and they are structured heterogeneously depending on the type of neuron. They can be seen as the inputs in the artificial neurons.

- Soma: a.k.a. cell body, perikaryon, or neurocyton; contains the cell nucleus surrounded by the cytoplasm. It synthesizes neurotransmitters and other molecules. It acts as the threshold and the sum of the inputs found in the artificial neurons.

- Axon: extensions where the electrical impulses or action potentials from the cell body to the next neuron. It can be seen as the output of artificial neurons.

In Figure 2.1 there is a depiction of the anatomy of a biological neuron with its main parts, whereas the basic structure of a chemical synapse between two neurons is in Figure 2.2. Connections between biological neurons can be electrical or chemical [Kriesel, 2007; Cabrera León, 2015]. The electrical synapse is bidirectional, less complex, faster, strong, direct and non-adjustable, whereas the chemical one is more diverse thanks to the variety of neurotransmitters that exist. Electrical activity in the axon of the presynaptic neuron produces the release of neurotransmitters, which were loaded in secretory vesicles, into the synaptic cleft. Then, the neurotransmitters arrive to receptors in the dendrite of the postsynaptic neuron, where they may start an electrical response or a secondary messenger

[2]Different types of biological neurons exist, and there are several ways of classifying them too. Similarly, many different cell types can be found in the nervous system other than neurons such as glias.

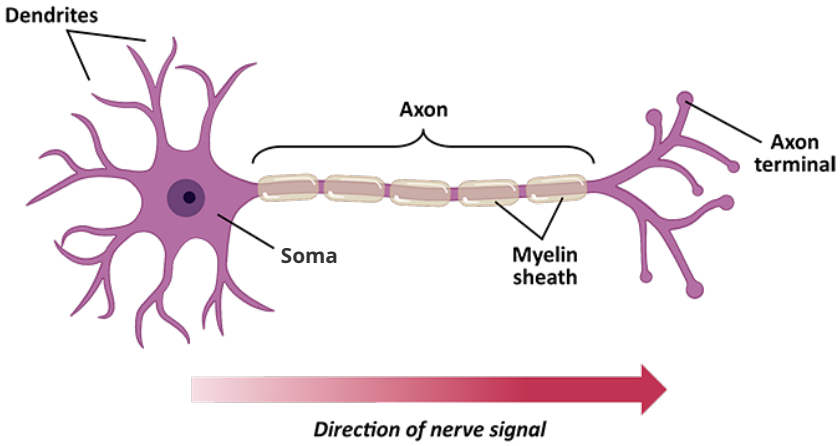


Figure 2.1: Anatomy of a biological neuron. Based on National Institutes of Health [2025]

pathway that may either excite or inhibit the postsynaptic neuron and are later transported to its nucleus. Chemical synapses are very adjustable due to the, generally, short live of the neurotransmitters and their heterogeneity, hence requiring specific receptors, Figure 2.2. Apart from its lower speed, connections being one-way is another disadvantage of the chemical synapse.

In general, the artificial neurons in ANNs can be schematically represented by Figure 2.3, where its counterparts in the biological neurons are also indicated.

A comparison between the characteristics of ANNs and BNNs has been made [Jain et al., 1996; Kriesel, 2007; Gabrielsson and Gabrielsson, 2006; Eluyode and Akomolafe, 2013; Cabrera León, 2015], Table 2.1. Furthermore, a comparison between the properties of the original Von Neumann computer, which is the base of all contemporary computers, and the BNNs has been made too [Jain et al., 1996; Cabrera León, 2015], Table 2.2. It can be seen that all current supercomputers made use of characteristics that are common in BNNs such as parallel an distributed computing, distributed and integrated memory (cache memory), and the high number of processors (although several orders of magnitude lower than the quantity of neurons in the brain).

2.1.1.2 Characteristics

A general framework that can be used to model or characterize an ANN is made up of these eight major aspects [Hecht-Nielsen, 1990; Rumelhart et al., 1986; Cabrera León, 2015]:

- A set of processing units and their local memories. They can be hidden or output units[3].
- A state of activation of each processing unit in a particular moment. It can be continuous or discrete, and limited or not to a certain range.
- An output function for each processing unit. It is frequently a threshold function.
- A pattern of connectivity among the processing units, where each connection has a synaptic weight that indicates its strength.

[3]Although input units are often conveniently depicted in the bibliography, they are not considered neurons *per se* as they do not process the input nor have bias.

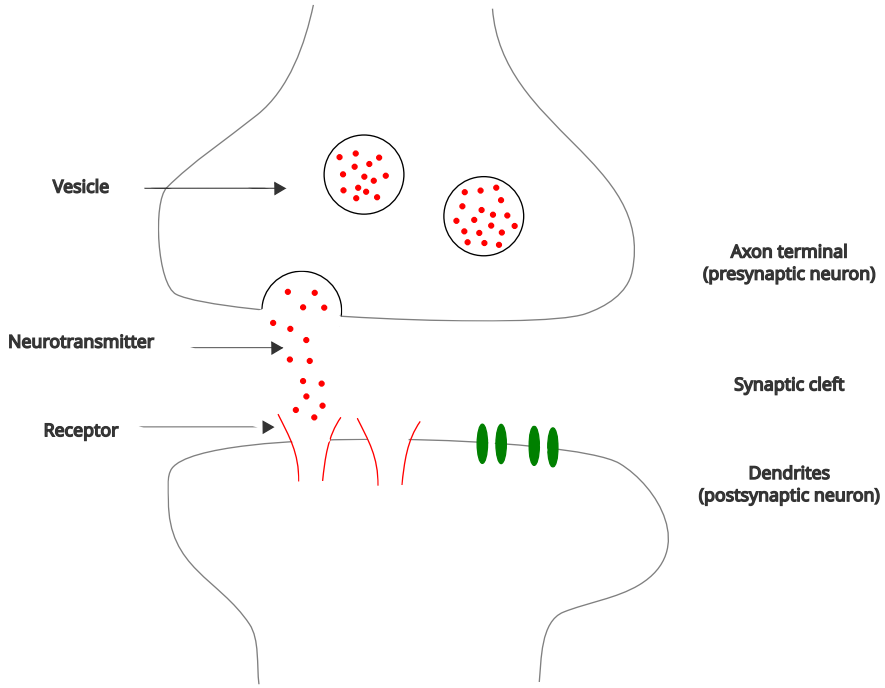


Figure 2.2: Basic structure of a typical chemical synapse between two biological neurons.

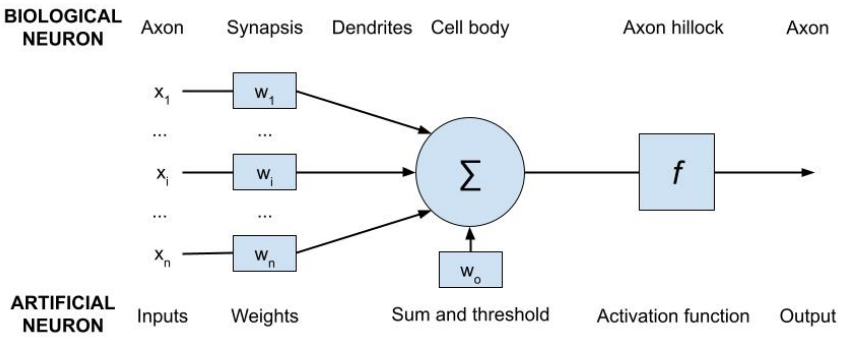


Figure 2.3: Scheme of an artificial neuron with its counterparts in the biological neuron.

Table 2.1: Comparison of Biological Neural Networks and Artificial Neural Networks.

Characteristics	Biological Neural Networks	Artificial Neural Networks
Connections	Excitatory and inhibitory	Excitatory and inhibitory
	Electrical and chemical	Only electrical
	Random and great quantity (each neuron may have $10^3 - 10^4$)	Precisely specified (several orders of magnitude lower)
Processing	Highly parallelized	Highly parallelized
	10^{14} synapses	10^8 transistors
Network size	10^{11} neurons	$10^2 - 10^4$ neurons
Processing speed	Several milliseconds + delays for chemical stabilization (refractory period)	Several nanoseconds
Information storage	In the synapses	In the weights matrix
Parts	Dendrites, cell body and axon	Input, node and output
Signals	Frequency modulated (pulses)	Amplitude modulated (numerical)
Learning	Based on past events, adjusting synaptic connections	Based on past events, adjusting weights
Computation style	Parallel & distributed	Parallel & distributed

Table 2.2: Comparison of Biological Neural Networks and the original Von Neumann computer.

Characteristics	Biological Neural Networks	Original Von Neumann computer
Processor	Simple	Complex
	Low speed	High speed
	High number (10^{11})	One or a few
Memory	Integrated	Separated from the processor
	Distributed	Localized
	Addressable by content	Not addressable by content
Computation	Distributed	Centralized
	Parallel	Sequential
	Self-learning	Stored programs
Reliability	Robust	Very vulnerable
Relevant attributes	Perceptual problems	Numerical and symbolic manipulations
	Based on knowledge	Based on memory
Working environment	Poorly defined	Well-defined
	No restrictions	Well-delimited

- A propagation rule or network function, that is used to distribute the activity patterns through the network of connectivities.
- An activation rule for combining the inputs to a processing unit with the current state of the unit, producing a new level of activation for this unit.
- A learning rule that can be used to change the patterns of connectivity based on experience. Such changes may be produced by the creation, modification (through the synaptic weights) or loss of synaptic connections.
- A representation of the environment where the system must operate. Both a local and global information environment exist.

Following the previous framework, any ANN can be characterized by these levels [Jain et al., 1996; García Báez, 2005; Cabrera León, 2015]:

Topology : a.k.a. the interconnection model, it indicates the structure of the network, how the neurons are interconnected [Miikkulainen, 2010]. This can be formally described by an oriented graph, where its nodes corresponds to the neurons or processing units, whereas the edges, to the communication channels between neurons. A floating point value is associated to each edge and it indicates the weight of the connection. The communication scheme between neurons can be hierarchical (connections only exist between consecutive layers) or non-hierarchical (connections may exist between non-consecutive layers). Most ANNs have a layered structure, ranging from 1 (monolayer) to multiple layers (multilayer), where each layer normally has a different use:

- Input layer: usually not considered when counting the number of layers as neurons in this one do not compute or process the information they receive. However, there are some architectures where this does not hold and the information is indeed processed by them.
- Hidden layers: ranging from 1 to more than one. In the latter case, these ANNs belong to the Deep Learning (DL) family, hence called Deep Neural Networks (DNNs). These layers are packed between the rest of type of layers, connecting both and sometimes between neurons in the same layer.
- Output layer: gives the network output, may have connection with any other layer, including itself.

The flow of information through these layers may be: feedforward (from input to output layers), recurrent (backward connections exist), or lateral connections (between neurons in the same layer).

The connections between layers may be: fully connected (all neurons are connected to the neurons of other layer), partially connected (some neurons are not fully connected), and one-to-one (in layers with the same shape and size, one neuron in a layer is connected to only one in the other layer).

Neurodynamics : indicates the way the information is processed locally, by the neurons. It is mathematically expressed by the propagation function, the activation function, and the output function.

These functions may be time-continuous or discrete, and, in the latter case, changes to the neurons' output values can be made asynchronously or synchronously, and between all or part of the neurons. The most common asynchronous method is called "random", where a random unit is chosen and its output is calculated. Another popular method is based on following the topological order, where all the calculations are made synchronously layer after layer, from the input to the output.

Learning : is the process whereby the network improves his performance in a certain task over time by the use of the input data. This concept is related to the "plasticity-stability" dilemma: the system should be sufficiently adaptable in order to learn new knowledge (plasticity) and, at the same time, it should not forget what was already learned (stability) [Gabrielsson and Gabrielsson, 2006]. In ANNs the learning process can be divided into two phases or operating modes: the loading mode, when the learning is done due to weights of the connections being updated based on the information extracted from the input data, and recovery mode, when answers for presumably unknown input samples are obtained based on the stored knowledge.

According to how the learning process is guided, several paradigms can be found:

- Supervised learning, when each sample of input pattern comes along with the label or expected output. Several groups exist:
  - By error correction: weights are updated so that the error (which can be local to the neuron or global to the network) is reduced. Perceptron, Learning Vector Quantization (LVQ) and Support Vector Machine (SVM) are some of the most popular examples.
  - Reinforcement: instead of the correct output, an evaluation (reward or punishment) of the network output is provided to the network when it guesses or fails, respectively.
  - Stochastic: random changes to the weights, based on some stochastic activation function or stochastic weights, and evaluating their effects according to the desired output and some probability distribution (through the use of energy functions as representatives of the stability of the network). One of the most popular examples of stochastic learning is the Boltzmann machines.
- Semi-supervised learning, when the network is trained with not all but a small quantity of expected outputs, so most input data have none [Chapelle et al., 2006; Zhu, 2005, 2011].
- Unsupervised learning, when no labels at all are presented to the network during the training phase. The learner is frequently meant to find some hidden patterns in the data. They can be divided into:
  - Hebbian: based on the observations made by the Canadian psychologist Donald O. Hebb, who said: "Let us assume that the persistence or repetition of a reverberatory activity (or "trace") tends to induce lasting cellular changes that add to its stability. [...] When an axon of cell A is near enough to excite a cell B and repeatedly or persistently takes part in firing it, some growth process or metabolic change takes place in one or both cells such that A's efficiency, as one of the cells firing B, is increased" [Hebb, 1949]. This theory is also called "Hebb's rule", "Hebb's postulate", or "cell assembly theory",

and it can be summarized by the sentence from US neurobiologist Carla J. Shatz "cells that fire together, wire together" [Schatz, 1992].

– Competitive: all neurons in the network compete to get activated. Two approaches are commonly found: "winner-takes-all" (a.k.a. hard competitive learning), where only one (or also a few around it) neuron, which is called the Best Matching Unit (BMU), activates for a specific input sample, and "winner-takes-most" (a.k.a. soft competitive learning), where the synaptic weights adapt regardless of the topological structure of the neurons in the network, but depending on the relative distances of the neurons in the input space [Martinetz and Schulten, 1991]. The most famous representative of the "winner-takes-all" family is the Self-Organizing Map (SOM) [Haykin, 1999; Kohonen, 2001; Haykin, 2009]. On the other hand, some "winner-takes-most" methods are the Neural Gas (NG) [Martinetz and Schulten, 1991] and its variants, such as the Growing Neural Gas (GNG) [Fritzke, 1995], which are much more flexible than the SOM. The "winner-take-all" approaches may bring along two problems [Fritzke, 1997a]: the appearance of "dead neurons" (units that are never BMU for any input, so their positions are never updated), and that different random initializations may give very different results. A possible solution to the latter is "winner-takes-most", reducing the risk of the system not being able to get out of a poor starting local minimum due to suboptimal initialization values [Fritzke, 1997a].

In general, training ANNs, which is the process in which network learning happens, is not straightforward. Also, it has some issues [Hastie et al., 2009]: suboptimal weights initialization, overfitting (which can be solved by early stopping the training or other regularization methods), scaling of the inputs is often needed, wrong number of hidden layers and neurons (more of them allow the network to adapt better to the input data), and multiple local error minima (might require several runs with random configurations of parameters).

Additionally, considering the many software solutions for training ANNs and DNNs, researchers need to carefully select and analyze them in terms of characteristics and quality [Hastie et al., 2009].

2.1.2 Some models

We need to define some common notation useful to describe some models which are similar to those developed in this PhD, which can be found below.

2.1.2.1 Common notation

Based on [Fritzke, 1997a; Hagan et al., 2014; Shalev-Shwartz and Ben-David, 2014] the following common notation can be defined:

- Domain set: a.k.a. input space; it is the set of samples that we want to label or cluster. It is usually dependent on the problem that is assessed, belonging to a specific problem domain. It may follow some probabilistic distribution or, according to [Fritzke, 1995], input signals may obey some unknown probability density function. It will be represented by $\mathcal{X} := \{\mathbf{x}_i \mid \forall i \in \{1, \ldots, l\},\ \mathbf{x}_i \in \mathbb{R}^n,\ \ n \in \mathbb{N}\}$.
- Labels set: only used in supervised learning, it is the set of expected outputs for a given domain set. In general, it can be represented by $\mathcal{Y} := \{\mathbf{y}_i \mid \forall i \in \{1, \ldots, l\},\ \mathbf{y}_i \in$

$\mathbb{R}^m,\ m \in \mathbb{N} \cup \{0\}, |\mathcal{Y}| \geq 0\}$. The number of classes m in a learning problem is given by the quantity of unique elements in this set: 2 for binary tasks, whereas values of at least 3 for multiclass tasks. In this work, the labels set will be represented by a set of values from $\mathbb{N}$ and $m = 2$ or $m = 3$.

- Training data: in supervised learning, it is a, normally finite, sequence of pairs $S := ((\mathbf{x}_i, \mathbf{y}_i) \mid \forall i \in \{1, \ldots, k\},\ k \leq l,\ \mathbf{x}_i \in \mathcal{X},\ \mathbf{y}_i \in \mathcal{Y})$, whereas only input data but no labels are provided in the unsupervised case. Together with validation and testing ones, they form the input data that a learner works with.
- Learner: a.k.a. classifier or predictor; refers to any method that requires to output a prediction rule $p : \mathcal{X} \to \mathcal{Y}$ and can be used to predict the label or cluster of new samples. ANNs are a type of learner, inspired by BNNs, that are composed of units and connections between them.
- Unit: a.k.a. vertex, neuron or node; it is the basic component of an ANN and is inspired by the biological neuron. Each of the N units in an ANN $u \in \mathcal{U} \mid \mathcal{U} := \{u_1, \ldots, u_N\}$, has an associated reference vector $\boldsymbol{\omega} \in \mathbb{R}^n$. This $\boldsymbol{\omega}$, also called weight or prototype vector, represents, biologically speaking, the strength of the synapse. Fritzke [1995] indicates that "reference vectors can be regarded as positions in input space of the corresponding units". However, in other ANNs these weights are associated not to the units but to the connections between the input layer and those units, and they indicate the strength of the connections, which is deemed more plausible biologically speaking. As different types of ANNs exist, units may also have associated other parameters such as error and label.
- Connection: a.k.a. edge, link or synapse; represents that there is a direct path between two units, that is, with no other neurons in between. Units may have connections with themselves or, more frequently, to other neurons in the same or different layers. It can be represented by the tuple $l := (u, v) \in \mathcal{C}$, where u, v are the pair of units that are being connected; and $\mathcal{C}$, the set of all connections. As different types of ANNs exist, connections may have other parameters too such as weight and activation status.

2.1.2.2 Simple competitive algorithm

Neurons in the simple competitive learning algorithm compete for the right to respond to certain input data. Over time each neuron will specialize in a set of patterns. Unlike the more complex competitive approaches that will be shown later, this one lacks the concept of "neighborhood", where close neurons respond to similar input samples.

Its main basic parts are: all neurons have the same characteristics and only differ in their – often random – weights, they respond differently to input patterns; the "strength" of each neuron is limited; the competitive mechanism only allow that one neuron (or a small group of them) responds, the so-called BMU.

The competition is generally defined by calculating a similarity measure (*e.g.* Euclidean distance) between an input sample $\mathbf{x}$ and the weights $\boldsymbol{\omega}$ of each neuron. The BMU is the most similar one, so its output is activated, whereas the outputs of the rest are deactivated.

2.1.2.3 Self-Organizing Map

Teuvo Kohonen's SOM, a.k.a. Self-Organizing Feature Map (SOFM) [Kohonen, 2001], is a paradigmatic unsupervised and competitive ANN. SOM quantifies the input space in different

regions represented by a specific number of output neurons, a.k.a. detectors. SOMs might be used as a visualization tool of high-dimensional data by projections over lower-dimensional maps [Rojas, 1996], most frequently a 2D grid. During this projection process, Kohonen maps try to extract the features of the input space preserving its topological properties.

SOM structure is made of an input layer fully interconnected, by excitatory connections, with the neurons in the Kohonen layer. A neighborhood relationship between these neurons exists, which is normally defined by a rectangular or hexagonal lattice, and is based on the distance between the neurons on the grid. Lattice shape enforces neighborhood topology, that is, the number of neighbors a neuron has [Westerlund, 2005; Cabrera León, 2015]: at distance 1, 6 for hexagonal, and 4 or 8 for square. Their neurodynamics is simplified by computing the least distance between the inputs $\boldsymbol{x}$ and a model [Kohonen, 2001]. This model is a parametric real vector that can be seen as the weight vector $\boldsymbol{w}$ in this neural architecture. The winning neuron, a.k.a. BMU, is the one with the minimum distance value to a specific input sample.

The learning process belongs to a winner-take-all, unsupervised and competitive training paradigm. The main variations are seen in the modification of the synaptic weights, which not only affects the winning neuron but also, to a lesser extent, the set of neurons in the winners' neighborhood N, Equation 2.1. During the training period, the neighborhood relationship between nodes decreases both in time and distance (commonly following a Gaussian function), and the learning rate α decreases with time.

$$\Delta_{li} = \begin{cases} \alpha(x_i - w_{li}) & \text{if } i \in N(\underset{k}{argmin}\ \{net_k(\boldsymbol{x})\}) \\ 0 & \text{otherwise} \end{cases} \tag{2.1}$$

The total number of neurons and the connections (unweighted) between each neuron and its neighbors do not change throughout the learning process, unlike the weight vectors of the neurons. Optimal initial and final values of the time-dependent parameters are key for a good convergence of the SOM.

Kohonen's SOM has been used as a basis for more complex neural architectures such as the Counterpropagation Network (CPN) [Hecht-Nielsen, 1987; Freeman and Skapura, 1991], and has inspired the development of many other ANNs, some of them described below.

2.1.2.4 Neural Gas and Competitive Hebbian Learning

Introduced by Martinetz and Schulten a decade after the appearance of the Kohonen's SOM [Martinetz and Schulten, 1991], the name of the NG was coined due to its neurons occupying the whole input space, akin to how gases behave inside a container.

The NG follows a straightforward algorithm [Fritzke, 1997a]. In each iteration t a new input sample $\boldsymbol{x}$ is selected and the distances between $\boldsymbol{x}$ and all N neurons are calculated and sorted in increasing order. The feature vectors $\boldsymbol{\omega}_i$ of these neurons are updated according to Equation 2.2, where ϵ is the learning rate, λ is the neighborhood range, i = 1, ..., N, and k = 0, ..., N-1.

$$\omega_{i_k}^{t+1} = \omega_{i_k}^{t} + \epsilon \cdot e^{-k/\lambda} \cdot (\boldsymbol{x} - \omega_{i_k}^{t}) \tag{2.2}$$

Both the number of neurons to be adapted λ and the the strength of the adaptation ϵ decrease after each iteration so that the algorithm converges before the maximum number of iterations t_{max} is reached. Initial and final values of these time-dependent parameters

$\lambda_i, \epsilon_i, \lambda_f, \epsilon_f$ require to be selected appropriately for a fast and optimal convergence after enough iterations [Fritzke, 1997a].

Compared to the SOM, in the NG there is no underlying graph so all points can move freely without the bonds that bind them together. Compared to the GNG, subsubsection 2.1.2.5, the NG does not change its size. That is, not only no new neurons are created as all exists from the beginning, but also no neurons can be deleted. Additionally, the concept of "connections between neurons" does not exist.

On the other hand, Competitive Hebbian Learning (CHL) [Martinetz, 1993] is rarely used on its own but in conjunction with other methods such as the NG [Fritzke, 1997a]. Fritzke described the CHL as having a zero learning rate because the feature vectors are not modified *i.e.* the neurons have fixed positions, so the network does not adapt. Instead, it creates neighborhood edges between neurons, and Martinetz proved that the generated graph preserves the topology in an optimal way. The CHL algorithm is also easy: given N disconnected neurons with random feature vectors associated $\boldsymbol{\omega}$ and an empty set of connections $\mathcal{C}$, for each input sample $\boldsymbol{x}$ it finds the two closest neurons c and c_2 according to a given distance metric, and creates a connection between them if it did not exist (also adds it to $\mathcal{C}$). This process is repeated until the maximum number of input samples have been used.

Finally, the combination of NG and CHL was developed in order to make a more powerful method [Fritzke, 1997a]. Each connection between neurons has the *age* property so that unnecessary ones are detected and marked for removal. Initially, all N neurons are randomly positioned in the input space and are completely disconnected. As in the NG, in each iteration the distances between it and every neuron are sorted, and their weights are updated in the same way. A connection between the two closest neurons c and c_2 is created if it did not exist, and, in any case, its *age* is set to 0. All connections emanating from c have their age incremented. If the age of a connection becomes larger than the maximal age $T(t) = T_i(T_f/T_i)^{t/t_{max}}$, it is removed. The whole learning algorithm iterates for the maximum number of adaptation steps t_{max}.

Similar to in NG, initial and final values of the time-dependent parameters $\lambda_i, \epsilon_i, T_i, \lambda_f, \epsilon_f, T_f$ require to be selected appropriately for a fast and optimal convergence after enough iterations [Fritzke, 1997a].

2.1.2.5 Growing Neural Gas

A GNG is an ANN that follows unsupervised learning [Fritzke, 1995, 1997a,b]. Fritzke [1997b] described the GNG both as an incremental variant of NG [Martinetz and Schulten, 1991] and CHL [Martinetz, 1993], and as a variant of the Growing Cell Structures (GCS) [Fritzke, 1994] without the strict topological constraints.

Fiesler and Beale [1997] defined an "ontogenic neural network" as an ANN where not only their interconnection strengths change according to a predetermined learning rule but also the ANN automatically adapts its topology (*i.e.* the number of layers and number of neurons per layer) to the problem. Henceforth, the GNG and most of its variants can be considered "ontogenic neural networks". Due to these properties, ontogenetic ANNs are good for clustering, vector quantization and data visualization [Fritzke, 1997b; Cabrera-León et al., 2023, 2024b].

In Figure 2.4 the topology of the original GNG is shown, where red neurons indicate dead neurons that are going to be deleted; a green neuron, a new one that is inserted in

regions where the error was the highest; dashed edges, connections that are eliminated both when there is a dead neuron or when a new neuron is added; and, dotted edges, connections that are created during the insertion of a new neuron. As an example of a wrong behavior there is the red-blue unit located in the middle, which represents a neuron that belongs to the light blue cluster but was erroneously assigned to the dark blue one.

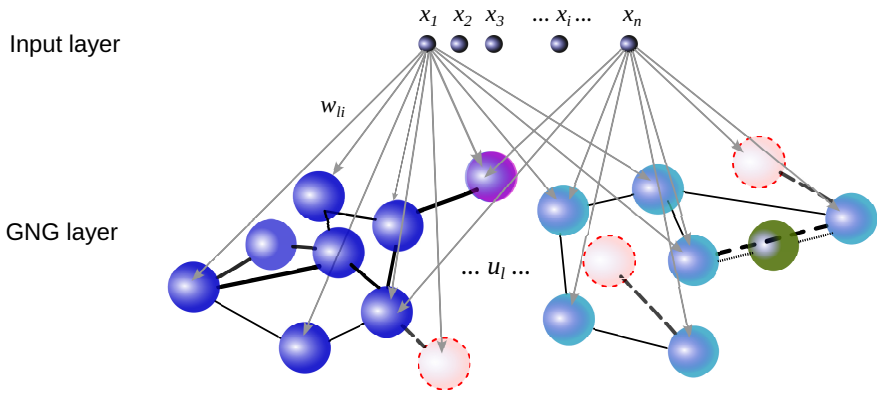


Figure 2.4: Topology of the original Growing Neural Gas.

A GNG g can be formally described by the tuple $g := (\mathcal{U}, \mathcal{C}, \theta)$, where $\mathcal{U}$ is the set of units (a.k.a. vertices, nodes or neurons); $\mathcal{C}$, the set of connections (a.k.a. edges or links); and θ, the set of hyperparameters [Fritzke, 1995; Kerdels, 2016]. If we ignore θ, a GNG, according to graph theory, can be considered a simple, normally disconnected, graph: an undirected graph without loops nor multiple edges [Clark and Allan Holton, 1991; Rosen, 2004]. That is, a GNG may be seen as a set of disjoint subgraphs or connected components, Equation 3.4, whose connections do not have orientations. Therefore, a connected graph consists of a single connected component. As each subgraph is a graph, it has a similar formal description. Based on [Clark and Allan Holton, 1991], $s_g := (\mathcal{U}_{s_g}, \mathcal{C}_{s_g}, \theta)$ is a subgraph of the graph g if $\mathcal{U}_{s_g} \subseteq \mathcal{U}$ and $\mathcal{C}_{s_g} \subseteq \mathcal{C}$.

In the GNG, a neuron $u \in \mathcal{U}$ has associated a reference vector $\boldsymbol{\omega}$, which relates inputs $\mathbf{x}$ with u, and an accumulated error $e \in \mathbb{R}$. For the sake of convenience, for unit u we will define $\boldsymbol{\omega}_u$ and e_u as synonyms of the previous elements, respectively.

In the GNG, a connection $l \in \mathcal{C}$ can be depicted by the tuple $l := (u, v)$, where $u, v \in \mathcal{U}$ are the pair of units that are being connected. As we can see from this first formula, connections in a GNG are unweighted, unlike in other ANNs, and are used to define the topological structure of the network. Each connection has an age associated $age_{(u,v)} \in \mathbb{N} \cup \{0\}$.

Unlike in a NG and a SOM, the parameters in a GNG are constant over time. This set of hyperparameters is comprised of $\theta := \{\varepsilon_c, \varepsilon_n, d, \beta, \lambda, a_{max}, \varsigma, epochs\}$, where:

- ε_c is the step or learning rate for the winner neuron, a.k.a. BMU, c.
- ε_n is the step or learning rate for the neighbors of the winner neuron, Equation 2.7.
- d is the decaying parameter that ensures that the errors of all neurons will be reduced over time.
- β is another decaying parameter but this one is only applied to the two largest accumulated errors, after the insertion of a new neuron.
- λ indicates the number of iterations or training steps before a new neuron is created.

- a_{max}, is the maximum age of a connection, which will imply deleting it when reached.
- ς is the largest size of the network, *i.e.* its maximum number of units[4].
- *epochs* [5] are the number of times the model has trained with the full input set. For example, training for one epoch implies that each sample in the training dataset has been used once by the model.

In the next paragraphs we are going to describe the GNG, its neuron model, its network topology (including both the network framework and the interconnection structure), and its learning algorithm. It should be noted that, for the learning algorithm, we will be following the step order indicated in [Fritzke, 1995, 1997a; Kerdels, 2016], which is different to the one found in [Fritzke, 1997b]. However, after comparing them manually, this alteration apparently does not produce substantial changes in the final results.

Neuron model:

- Propagation function: either based on distance measures — usually the Euclidean distance, Equation 2.3 — or based on the dot product, Equation 2.4. Hence, in such cases the network input of neuron u, net_u, can be defined as:

$$net_u(\mathbf{x}) := f_{prop}(\dots x_i \dots \ , \ \dots \omega_u \dots)$$
$$net_u(\mathbf{x}) := \|\mathbf{x} - \boldsymbol{\omega}_u\|$$

$$net_u(\mathbf{x}) := \sqrt{\sum_i (x_i - \omega_{iu})^2} \tag{2.3}$$

$$net_u(\mathbf{x}) := \sum_i x_i \cdot \omega_{iu} \tag{2.4}$$

  where $\mathbf{x} = (x_1, x_2, \dots, x_n) \in \mathbb{R}^n$ are the components of the input vector; and $\boldsymbol{\omega}_u = (\omega_{1u}, \omega_{2u}, \dots, \omega_{nu}) \in \mathbb{R}^n$, the associated reference vector of neuron u.

- Activation function: it corresponds to a competitive-like one, in this case the minimum, so the activation function of neuron u, a_u, is given by:

$$a_u := f_{act}\left(net_u\left(\mathbf{x}\right)\right) := \begin{cases} 1 & \text{if } u = \underset{k \in \mathcal{U}}{\arg\min}\left(net_k\left(\mathbf{x}\right)\right) \\ 0 & \text{otherwise} \end{cases} \tag{2.5}$$

- Output function: identity/linear function.

$$f_{out}\left(a_u\right) := a_u \tag{2.6}$$

Network topology: unlike with the SOM but similar to the NG, there is no grid structure in the GNG.

[4]This parameter is optional because another stopping criterion can be used in the learning algorithm.

[5]Other formal definitions of GNG and variants did not mention *epochs*, despite having used it in the algorithm [Fritzke, 1995; Kerdels, 2016].

- A single layer of neurons[6], starting from 2 neurons and growing to the maximum indicated size[7].
- All neurons in the input layer are connected to the neurons in the main layer.
- The neurons in the main layer are partially interconnected.
- Neighborhood, unlike in other ANNs, is not related to the distance between neurons but to the existence of a connection between them. Hence, this "topological neighborhood" of neuron u, $\mathcal{N}(u)$, is related to the "lateral connections update", Equation 2.12, and is defined as:

$$\mathcal{N}(u) := \{k \in \mathcal{U} \mid \exists (u,k) \in \mathcal{C}\} \tag{2.7}$$

- A list of nearest neurons to an input $\mathbf{x}$ is calculated in each training step, albeit only the first and second positions, which refers to the BMU c and second BMU c_2, are necessary in GNG:

$$c := \underset{u \in \mathcal{U}}{\arg\min} \left(net_u\left(\mathbf{x}\right)\right) \tag{2.8}$$

$$c_2 := \underset{u \in \mathcal{U}-\{c\}}{\arg\min} \left(net_u\left(\mathbf{x}\right)\right) \tag{2.9}$$

- Connection senescence: only the connections emanating from the current winner neuron c age after every training step, although the age of the connection between c and its nearest neighbor is refreshed (*i.e.*, age reset to 0).

$$age_{(u,l)} := age_{(l,u)} := \begin{cases} age_{(u,l)} + 1 & \text{if } u = c,\ l \in \mathcal{N}(u) - \{c_2\} \\ 0 & \text{if } u = c,\ l = c_2 \\ age_{(u,l)} & \text{otherwise} \end{cases} \tag{2.10}$$

- Neural apoptosis: the death of a neuron happens when there are no connections emanating from it (that is, it has no neighbors), which happens due to the aforementioned "connection senescence", Equation 2.10.

$$\mathcal{U} := \mathcal{U} - \{r\},\ \text{ if } \mathcal{N}(r) = \emptyset,\ \ \forall r \mid r \in \mathcal{U} \tag{2.11}$$

- Lateral connections update is an unsupervised process based on Hebbian competitive learning. A connection is deleted when its age reaches the value indicated by the hyperparameter a_{max}.

$$\begin{gathered} \mathcal{C} := \mathcal{C} \cup \{(c, c2)\} \\ \mathcal{C} := \mathcal{C} - \{(u,l)\},\ \text{ if } age_{(u,l)} > a_{max},\ \ \forall (u,l) \in \mathcal{C} \end{gathered} \tag{2.12}$$

where a_{max} is the maximum age of a connection before deletion; $age_{(u,l)}$, the current age of the connection between neurons u and l; c, the BMU; and c_2, the second BMU[8].

[6] We are not counting the input layer, which also exists, as it commonly occurs in the majority of ANNs.
[7] See footnote 4.
[8] In the original article, [Fritzke, 1995], c and c_2 are called s_1 and s_2, respectively.

- Weight update, due to being a competitive ANN, is different in the winner neuron c and its direct topological neighbors in the graph, $\mathcal{N}(c)$. The weight update $\Delta\boldsymbol{\omega}_u$ is defined as:

$$\Delta\boldsymbol{\omega}_u := \begin{cases} \varepsilon_c \cdot (\mathbf{x} - \boldsymbol{\omega}_u) & \text{if } u = c \\ \varepsilon_n \cdot (\mathbf{x} - \boldsymbol{\omega}_u) & \text{if } u \in \mathcal{N}(c) \\ 0 & \text{otherwise} \end{cases} \tag{2.13}$$

  where ε_c is the learning rate of the winner neuron and ε_n is the learning rate of its direct topological neighbors, being recommended that $\varepsilon_n < \varepsilon_c$ in order to avoid slow and erratic training.

- Error update: each neuron has a local error variable related to the distance between the input $\mathbf{x}$ and the winner neuron c. During the adaptation, accumulating the squared distances to this error helps to identify neurons situated in areas of the input space where the error in the mapping is high. In such areas new neurons are inserted in order to reduce the error.

$$\Delta\, e_c := net_c^2(\mathbf{x}) \tag{2.14}$$

  Considering the decaying hyperparameter $d \in (0, 1)$, the error for neuron u, e_u, is updated at the end of each iteration according to the next formula, a decaying process that affects every neuron $u \in \mathcal{U}$:

$$\Delta\, e_u := d \cdot e_u \tag{2.15}$$

- Neurogenesis: a new neuron is created when the current training step is a multiple of the hyperparameter λ. Given the existing neurons q (which has the largest accumulated error of all units) and f (which has the largest accumulated error of all units among the neighbors of q), the new neuron r will be situated between them, connected to them, the old connections deleted, the error of both q and f reduced by the decaying hyperparameter $\beta \in (0, 1)$ (typically $\beta < d$), and r will hold the new error of q.

$$\begin{gathered} \mathcal{U} := \mathcal{U} \cup \{r\}, \;\; \text{if } \exists q \mid q := \underset{k \in \mathcal{U}}{\arg\max}\, (e_k) \;\wedge\; \exists f \mid f := \underset{k \in \mathcal{N}(q)}{\arg\max}\, (e_k) \\ \omega_r := 0.5 \cdot (\omega_q + \omega_f) \\ \mathcal{C} := \mathcal{C} \cup \{(r, q), (r, f)\} - \{(q, f)\} \\ age_{(r,q)} := 0; \quad age_{(r,f)} := 0 \\ e_q := \beta \cdot e_q \quad ; \quad e_f := \beta \cdot e_f \quad ; \quad e_r := e_q \end{gathered} \tag{2.16}$$

  where e_k returns the accumulated error e of the unit k; and $age_{(u,\, l)}$ returns the age of the connection between neurons u and l .

Learning algorithm: a GNG is initialized with two units with prototype vectors ω, usually randomly, chosen from the training data, and their accumulated error variables e are set to 0. The set of connections $\mathcal{C}$ is initially empty[9]. Then, the GNG grows by

[9] A few implementations have been found where there is a connection between both initial units from Step 1. The behavior of these implementation is identical to the one described here.

processing samples extracted from the training data[10]. The learning algorithm of the GNG comprises the next 11 steps:

Step 1: Start with two units, $U := \{a, b\} \mid e_a := 0, e_b := 0$, and ω_a and ω_b are reference vectors usually randomly chosen from the input data. $\mathcal{C} := \emptyset$

Step 2: Iterate through the training data S.

Step 3: Find the nearest unit c and the second nearest unit c_2, Equation 2.8 and Equation 2.9, respectively. They are also called BMU and second BMU.

Step 4: Increment the age of all edges emanating from c, Equation 2.10.

Step 5: Add the squared distance between the observation and the associated reference vector of c to the accumulated error of c, Equation 2.14.

Step 6: Move c and its direct topological neighbors towards the observation by the fractions ε_c and ε_n, respectively, of the total distance, Equation 2.13.

Step 7: If there is a connection between c and c_2, set the age of this connection to zero, Equation 2.10. If this connection does not exist, create it, Equation 2.12.

Step 8: Remove connections with an age larger than a_{max}, Equation 2.12. If this results in units having no emanating connections, remove them as well, Equation 2.11.

Step 9: If the number of steps so far is an integer multiple of parameter λ, insert a new unit, Equation 2.16:

a) Determine the unit q with the maximum accumulated error.

b) Insert a new unit r halfway between q and its neighbor f with the largest error variable.

c) Insert edges connecting the new unit r with q and f, and remove the original edge between q and f.

d) Decrease the error variables of q and f by multiplying them with the constant β. Initialize the error variable of r with the new value of the error variable of q.

Step 10: Decrease the error variables of all neurons by multiplying them with a constant d, Equation 2.15.

Step 11: Go to Step 2 if a stopping criterion (*e.g.*: network size ς, performance metric, etc) is not yet fulfilled.

The GNG will gradually approximate the structure of the input data manifold [Fiesler and Beale, 1997], after repeating this learning algorithm enough times with all the available training data (the so-called "epochs").

Some advantages of the GNG compared to the NG and CHL combination of Martinetz and Schulten are [Fritzke, 1997b]: the network size and the total number of adaptation steps do not need to be predefined, all hyperparameters are constant, and the growth process can be stopped when some performance criterion has been met. Also, GNG is capable of continuous learning *i.e.* there is no need to train the network from the beginning if new input

[10]In [Fritzke, 1995, 1997a,b; Kerdels, 2016] $p(\xi)$ represents a continuous probability density function that generates the input samples; and ξ, an input sample. Fritzke [1997b] also indicates that such input samples may also be obtained from a finite training dataset $\mathcal{D} := \{\xi_1, \ldots, \xi_M\}, \xi_i \in \mathbb{R}^n$, which is similar to the nomenclature we preferred to use.

data are available. All intermediate stages of the network properly describe the underlying distribution due to the fractal-like growth of the network, and the resolution of these stages depends on the current number of neurons in the network [Fritzke, 1997b].

2.1.3 About Deep Learning

DL is a huge family of ANNs where many hidden layers exist within a network. They used to have high computation cost and training time, and many techniques have been developed to diminish them [Cabrera-León et al., 2024a, 2025].

Some examples of DNNs are: Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), Autoencoders (AEs) and transformers. Only the first one will be described below as some models of this family were used in this thesis for comparisons purposes.

The first series of CNN architectures was developed in 1998 by Yann LeCun, one of the most influential scientists in the field of artificial intelligence [LeCun et al., 2015]. Years later, the emergence of the massive ImageNet dataset enabled the enormous advances in computer vision of recent decades, with the development of new architectures such as ResNet, which we used in our study.

The operation of the CNN architecture covers three key concepts: local receptive fields, shared weights and biases, and activation and pooling. First, in a typical ANN, each neuron in the input layer is connected to a neuron in the hidden layer. However, in a CNN, only a small region of input layer neurons connects to hidden layer neurons. This region is known as the local receptive fields. These fields are translated through an image using convolution to create a feature map from the input layer to the hidden layer, hence the name. Second, in a CNN, the weight and bias values are the same for all hidden neurons in a given layer, meaning that all hidden neurons are detecting the same feature in different regions of the image. Third, activation applies a transformation to the output of each neuron using activation functions (i.e., ReLU), taking the value and mapping it to the highest positive value, except when the output is negative, which maps it to 0. Finally, by applying pooling, it is possible to further transform the output of the activation step. This involves reducing the dimensionality of the feature map by condensing the output of small regions of neurons into a single output, which in turn reduces the number of parameters the model needs to learn. In short, the structure of a CNN can be summarized as an input layer, hidden layers (convolutional and pooling layers), and finally, the classification layer (fully connected layers, or FCLs, and softMax). These networks are very robust to translations, rotations, and scaling of the input data [Cabrera-León et al., 2024a], making them very useful for image and video recognition.

Many variants of CNN exist nowadays [LeCun et al., 2015], each with its own number of layers, quantity of parameters and properties: LeNet, GoogLeNet, DenseNet, Residual Network (ResNet), Visual Geometry Group (VGG)...

2.1.3.1 DeepInsight and ParallelNet

The purpose of the DeepInsight methodology is to transform a non-image sample into image format (element preparation) and then introduce it into the CNN architecture for feature extraction (training) and finally carry out the prediction and classification process [Sharma et al., 2019].

Non-linear dimensionality reduction techniques, such as t-distributed stochastic neighbor embedding (t-SNE) or Kernel Principal Component Analysis (PCA), are applied to the

training set to obtain a 2D plane. Once the feature positions on the Cartesian plane (2D plane) have been defined, the convex hull algorithm is used to find the smallest rectangle that contains all the points. Next, a rotation is performed to fit the rectangle to the network input (90^o with the horizontal and vertical) and the Cartesian coordinates are transformed into pixels. Finally, once all the features are located in the pixels, the feature values are mapped based on their location. If more than one matches, an average is calculated and they are placed in the same location. This means that the size of the pixel grid must be adjusted to the number of features to avoid resolution issues. In this way, a single image (or feature vector) is generated for each sample, which will be provided to the CNN architecture for model training and predictions.

It is important to note that the feature values must be normalized before applying the transformation to the images. To achieve this, a normalization process called "Norm 1" was carried out. With this method, each feature is normalized with its minimum and maximum value (independent features), resulting in features with values between 0 and 1. If any value of the different features is less than 0 or greater than 1 after normalization, it will be grouped between 0 and 1 to maintain consistency.

Although the DeepInsight pipeline admits the use of different CNNs pre-trained with ImageNet, ParallelNet was the one used in [Sharma et al., 2019]. In this parallel CNN architecture different filter (kernel) sizes can be used effectively to train the model. Each of its 4 parallel layers consists of: a 2D convolution layer, a batch normalization layer (to avoid overfitting during training), a ReLU activation layer, and a max pooling layer. The output of the 4 convolution layers is combined and fed into a fully connected layer. Finally, a softmax layer provides the output as class labels.

2.2 Feature ranking

Datasets from real life experiments are almost always characterized by numerous features (a.k.a. variables or predictors) that are barely of interest for most classification tasks. Therefore, methods for dimensionality reduction have been implemented, which involve feature selection and feature projection. The latter is described in section 2.4.

Feature ranking, a.k.a. feature selection, involves choosing a subset of features considered relevant for the desired task. This concept assumes that any data contain irrelevant or redundant features that can be discarded with reduced or no loss of information. Although its not originally intended, there are feature extraction methods, which involve creating new features from functions of the original ones, where the fabricated features are preferred to the original ones as they allow using less features and these might be more relevant or non-redundant.

There are three main groups of feature selection algorithms:

- Wrapper methods, which use a predictive model to rank feature subsets. They imply training a model and testing it with a hold-out subset, so they are extremely slow but provides the best feature set for that specific model. Ranking of the feature set is given by the error rate of the model in that test set.

- Filter methods use an indirect metric to rank the feature set. This metric must be fast to calculate and able to rank conveniently the feature set, such as inter-class and intra-class distances and mutual information. They are a lot faster than wrappers and

more easily expose the relationship between features, but the ranking is not specific for the model that will be used.

- Embedded methods perform feature selection while building the model. In terms of computation complexity, they lie between wrappers and filters.

Although most learning algorithms need the feature selection to be performed externally, several of them, such as Decision Trees (DTs), Random Forests (RFs) and Extreme Gradient Boosting (XGBoost), are able to do this on their own, internally, as part of their normal functioning.

In the experiments with real data in this PhD thesis two feature ranking techniques were evaluated:

- Fast Correlation-Based Filter (FCBF) is a hybrid filter and wrap feature selection method developed by Yu and Liu [Yu and Liu, 2003]. It uses symmetric uncertainty to determine the correlation between features and categories as well as to highlight redundancy between the features. FCBF obtains an initial ranking of relevant features that is then cleaned of those that are inter-correlated and do not provide more or different information from that given by the other features initially considered relevant. As a result, calculation efficiency, and therefore speed, is enhanced, which improves the recognition rate. Furthermore, these authors concluded that FCBF demonstrated promising results, achieving the highest level of dimensionality reduction and, hence, a great ability to identify redundant characteristics to address complex and multi-category situations [Yu and Liu, 2003].

- XGBoost [Chen and Guestrin, 2016] is a scalable tree ensemble method mostly used for classification that, as a byproduct, also generates a ranking of features. Unlike FCBF, redundant features are never discarded internally, which results in rankings including more than one feature providing similar information. Hence, the feature ranking provided by XGBoost was considered of lower quality, so the one done with FCBF was preferred.

2.3 Data scaling

Also called data normalization, the goal of data scaling methods is to normalize the range of the features of the data as they tend to vary widely, which may impair the performance of some ML algorithms (*e.g. those that used Euclidean distance*). It is recommended to use scaling methods so that coefficients are penalized correctly when regularization is used as part of the loss function. It is also expected that gradient descent converges faster with scaled data.

Many scaling methods exist but only those that were analyzed in the experiments in chapter 4 are briefly explained [Pedregosa et al., 2011][11]:

- "Standard", a.k.a. Z-score, probably the most popular one, consists of removing the mean and scaling to unit variance.

[11]The names of the scaling methods used by the tool that was used for data scaling, "scikit-learn", have been preserved for convenience, and other names were included when possible.

- "Robust", similar to the "Standard" one but using the median and the quantile range (such as the Interquartile Range) instead. Both the median and the Interquartile Range are robust to outliers, unlike the mean and the variance.
- "MaxAbs", where each feature is scaled and translated individually so the maximal absolute value of each one will be 1.0. Sparsity is kept as the data are not centered or shifted.
- "MinMax", a.k.a. rescalling, where each feature is scaled and translated individually so it lies between the given range, such as between 0 and 1 or -1 and 1.
- "Normalizer", where each sample with one or more non-zero component is rescaled independently so that its L1 or L2 norm equals 1.
- "Yeo-Johnson PowerTransformer", power transformers are a family of parametric, monotonic transformations that are applied featurewise to make data more Gaussian-like. Other power transforms exist too: Box-Cox and uniform output.

2.4 Data projection

The main purpose of using data projection techniques is to reduce the dimensionality of the original data while retaining the maximum quantity of information.

Many projection methods exist but only those that were found optimal in any of the experiments in chapter 4 are briefly explained:

- PCA Jolliffe and Cadima [2016] is an unsupervised dimensionality reduction method, which allows to reduce the size of a dataset by projecting the data to a low-dimensional space that contains most of the original variance. This method produces a series of principal components from a multivariate random data by computing the eigenvectors of its covariance matrix corresponding to the largest eigenvalues, and the projection of the data over the eigenvectors. Sosa-Marrero et al. [2021]. PCA has been widely employed for data processing in many fields García Báez et al. [2007]; Jolliffe and Cadima [2016].
- Kernel PCA [Schölkopf et al., 1998a] is an extension of PCAs where the linear kernel used in the latter is substituted by a non-linear one, allowing to compute principal components in high-dimensional feature spaces. This process is similar to the usage of non-linear kernels in the originally linear SVM [Vapnik, 1999].
- Neighborhood Component Analysis (NCA) is a supervised method aimed to find the best input data projection or linear transformation for a stochastic nearest neighbors rule to yield the best classification accuracy in the transformed space, without assuming that the data have a parametric structure in the low dimensional representation Goldberger et al. [2004].
- t-SNE is a statistical method and non-linear dimension reduction method for high-dimensional data visualization [van der Maaten and Hinton, 2008]. By giving each data point a location in a 2D or 3D map, it allows visualizing high-dimensional data. Distant data points in the original space are modeled by distant ones in the low-dimensional space with high probability.

2.5 Model evaluation metrics

Several ways of evaluating and comparing models have been developed in the last decades [Sokolova et al., 2006; Cabrera León, 2015]. They can be grouped according to its purpose: performance measures, clustering quality metrics, time measurements, and computational costs and resources requirements. A subset of the most convenient ones was selected for the experiments performed in this thesis.

2.5.1 Performance measures

This huge set of metrics measures in different ways how well a model behaves in a specific problem, frequently by assessing different characteristics of the model [Swets, 1988; Bradley, 1997; Sokolova et al., 2006; Akobeng, 2007c,b,a; Powers, 2011]. For simplicity reasons, only performance metrics for binary classifiers will be exposed below, albeit multiclass versions of them may exist too:

- True Positives (TP) are the number of samples that present the condition that were correctly classified as having the condition.
- True Negatives (TN) are the quantity of samples that do not present the condition that were correctly classified as not having the condition.
- False Positives (FP), a.k.a. Type I error, are the number of samples that do not have the condition that were incorrectly classified as having the condition.
- False Negatives (FN), a.k.a. Type II error, are the quantity of samples that have the condition that were incorrectly classified as not having the condition.
- Confusion matrix, a.k.a. contingency table, is a tabulated representation of the performance of a model, where rows represent the predicted or diagnosed classes, and columns, the real classes or conditions [12]. It shows TP, TN, FP and FN in a more meaningful way. The confusion matrix of a model with the best performance must simultaneously fulfill the next two conditions: maximum main diagonal and minimum secondary diagonal. This can be written with the following equations, where N is the number of patterns or input vectors: $TP + TN = N$ and $FP + FN = 0$.

		Condition		
		Positive	Negative	Total
Prediction	Positive	TP	FP	$TP + FP$
	Negative	FN	TN	$FN + TN$
	Total	$TP + FN$	$FP + TN$	N

- Precision, a.k.a. Positive Predictive Value (PPV), is the proportion of samples that truly have the condition from among those that the model indicated that have the condition, Equation 2.17.

$$\text{Precision} = \frac{TP}{TP + FP} \tag{2.17}$$

[12] Some authors use a transposed version of this matrix. Both can be used interchangeably as they are equally valid.

- Sensitivity, a.k.a. recall or True Positive Rate (TPR), indicates how well the model performs with samples that have the condition. In other words, it is the ratio of samples that the model indicated that have the condition from among those that truly have it, Equation 2.18. A high sensitivity brings along a low ratio of FN.

$$\text{Sensitivity} = \frac{TP}{TP + FN} \tag{2.18}$$

- Specificity, a.k.a. True Negative Rate (TNR), indicates how well the model performs with samples that do not have the condition. That is, it is the ratio of samples that the model indicated that do not have the condition from among those that truly do not have it, Equation 2.19. A high specificity brings along a low ratio of FP.

$$\text{Specificity} = \frac{TN}{TN + FP} \tag{2.19}$$

- Accuracy indicates the proportion of samples that were correctly identified from among all, Equation 2.20. Metz [1978]; Provost et al. [1997] do not recommend using accuracy, especially if the class distribution and error costs are unknown, preferring Receiver Operating Characteristic (ROC) curves in these cases.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{2.20}$$

- Balanced accuracy was developed to deal with imbalanced datasets. It is defined as the arithmetic mean of the recall obtained on each class, which can be simplified in binary classification as the half of the sum of sensitivity and specificity, Equation 2.21.

$$\text{Balanced accuracy} = \frac{\frac{TP}{TP + FN} + \frac{TN}{TN + FP}}{2} \tag{2.21}$$

- Error rate is the proportion of samples that were incorrectly identified from among all that were used, Equation 2.22.

$$\text{Error rate} = \frac{FP + FN}{TP + TN + FP + FN} \tag{2.22}$$

- F-score, a.k.a. F1 score, F-measure, Sørensen–Dice coefficient or Dice similarity coefficient, indicates how accurate a classification was done by tanking into account both precision and recall. The nearest to 1 is the value of this metric, the better the model has performed. Powers [2011] recommends only using it in conjunction with other metrics as this one ignores the TN, Equation 2.23.

$$\text{F-score} = 2 \cdot \frac{\text{Precision} \cdot \text{Sensitivity}}{\text{Precision} + \text{Sensitivity}} = \frac{2 \cdot TP}{2 \cdot TP + FP + FN} \tag{2.23}$$

- ROC curves indicate graphically the performance of a classifier, where the X axis is the False Positive Rate (FPR) or 1 - specificity, and the Y axis is the sensitivity [Metz, 1978; Bradley, 1997; Duda et al., 2001; Forcada, 2003; Fan et al., 2006; Brown and Davis, 2006; Akobeng, 2007a; Slaby, 2007b,a; Lobo et al., 2008; Airola et al., 2010; López et al., 2012; Kumar and Indrayan, 2011; García Abad, 2012; Navan, 2014; Tape and University of Nebraska Medical Center, 2014; Systat Software, 2017]. A ROC curve is able to encapsulate all the information of a confusion matrix as FN are the complement of TP, and TN are the complement of FP [Swets, 1988]. A single number, Area Under the Curve (AUC), might be used to summarize a ROC curve, although with limitations. AUC is a metric that is strongly related to the mathematical concept called integral, and one of the simplest ways of calculating it is by means of trapezoidal integration [Bradley, 1997]. AUC ranges between 0 and 1, where 1 is the best value and is obtained by the gold standard test. A model is said to be better[13] than other if its ROC curve passes nearer to the position (0, 1), the so-called "perfect classification point". In this point, where the AUC achieves the maximum value, the quantity of both FN and FP is 0, so 100% sensitivity and 100% specificity, respectively [Zweig and Campbell, 1993; Akobeng, 2007a]. Provost et al. [1997] indicate that ROC curves are independent to both the class distribution and error costs.

- Clinical Utility Index (CUI) measures the clinical value of a diagnostic test considering both occurrence and accuracy of the test [Mitchell, 2008; Mitchell et al., 2009; Mitchell, 2009]. As both sensitivity and specificity are deemed crucial in clinical settings, two variants of the CUI exists, respectively for rule-in and rule-out accuracy, Equation 2.24: CUI+, which is calculated as the product of positive predictive value and sensitivity, and CUI-, as the product of specificity and negative predictive value. Adapted from Landis and Koch [1977], Mitchell [2009] proposed an interpretation for the CUI values: 0.0-0.2, minimal; 0.21-0.4, slight; 0.41-0.6, fair; 0.61-0.8, good; 0.81-0.92, excellent; and 0.93-1, near perfect.

$$\text{CUI}+ = \frac{TP}{TP+FN} * \frac{TP}{TP+FP} \qquad \text{CUI}- = \frac{TN}{TN+FP} * \frac{TN}{TN+FN} \tag{2.24}$$

- Matthews Correlation Coefficient (MCC), a.k.a. phi coefficient or mean square contingency coefficient, is considered a balanced metric that can be reliably used when the classes are of very different sizes, unlike accuracy. A MCC value of +1 means a perfect prediction; 0, an average random prediction; and -1, an inverse prediction, Equation 2.25.

$$\text{MCC} = \frac{TP \cdot TN - FP \cdot FN}{\sqrt{(TP+FP) \cdot (TP+FN) \cdot (TN+FP) \cdot (TN+FN)}} \tag{2.25}$$

In Table 2.3 the optimal, worst and interval of values for each of the mentioned performance measurements have been summarized.

[13]This relationship can also be referred as "dominance", *i.e.* "a classifier dominates other [...]" [Slaby, 2007b].

Table 2.3: Interval, worst and optimal values for the described performance metrics.

Performance metric	Interval	Worst value	Optimal value
Precision	[0, 1]	0	1
Sensitivity	[0, 1]	0	1
Specificity	[0, 1]	0	1
Accuracy	[0, 1]	0	1
Balanced accuracy	[0, 1]	0	1
Error rate	[0, 1]	1	0
F-score	[0, 1]	0	1
Area Under the Curve	[0, 1]	0	1
Clinical Utility Index	[0, 1]	0	1
Matthews Correlation Coefficient	[-1, 1]	-1 or 0	1

2.5.2 Clustering quality metrics

Clustering methods divide an input dataset into a number of groups, the so-called "clusters" [Rosenberg and Hirschberg, 2007]. The main goal of clustering algorithms is to simultaneously achieve high intra-cluster similarity (all samples in the same cluster as similar) and low inter-cluster similarity (samples from distinct clusters are different). Almost always this process is carried out in an unsupervised way, that is, the algorithm ignores the objective labels. For this reason, the performance measures exposed in subsection 2.5.1 are not useful. Unlike them, clustering quality measurements do not require the desired output labels to be calculated or even to exist at all. However, for testing purposes some target groups may be defined, which are commonly called "labels" due to its similarity with those found in classification problems.

Several clustering quality measures have been proposed [Rosenberg and Hirschberg, 2007]:

- Purity is the percentage of the total number of samples that were classified correctly. It ranges between 0 and 1. Purity is calculated by assigning the most frequent class to each cluster and summing the number of correct class labels in each cluster and dividing it by the quantity of samples, as indicated in Equation 2.26, where N is the number of samples; k, the number of clusters; c_i, a cluster in the set C of all clusters; and t_j, the classification which has the maximum count for cluster c_i. High purity is straightforward to obtain when the number of clusters is large, as it will be the maximum value, 1, if each sample gets its own cluster [Rosenberg and Hirschberg, 2007]. For this reason, better use it together with other metrics.

$$\text{Purity} = \frac{1}{N}\sum_{i=1}^{k} max_j \mid c_i \cap t_j \mid \tag{2.26}$$

- Completeness criteria is achieved when all members of a given class are assigned to the same cluster. It is symmetrical to homogeneity [Rosenberg and Hirschberg, 2007].

- Homogeneity criteria is satisfied when a clustering method is able to make that each cluster contains only members of a single class.

- V-measure, or validity-measure, is the harmonic mean of homogeneity and completeness [Rosenberg and Hirschberg, 2007]. It is safer to use an adjusted index if the sample size is small or the number of clusters is big.
- Adjusted Rand Index, as it can be inferred from the name, is the Rand Index adjusted for chance. The Rand Index calculates the similarity of two clusterings by counting all pairs of samples that are assigned in the same or different clusters in the predicted and true clusterings. A score of 0 indicates random labeling, whereas -0.5 is used for especially discordant clusterings.
- Normalized Mutual Information, as indicated by its denomination, is a normalization of the Mutual Information score in order to be between 0 and 1, respectively no mutual information and perfect correlation. This measure is not adjusted for chance.
- Silhouette is the mean of the Silhouette Coefficient for each sample. The silhouette for each sample is calculated combining the mean distance between a sample and all other points in the same class, and the mean distance between a sample and all other points in the next nearest cluster. Silhouette ranges between -1 and 1, respectively incorrect clustering and highly dense clustering. A score near 0 indicates overlapping clusters. Knowing the ground truth labels is not required.
- Davies-Bouldin Index is calculated by averaging the "similarity" between clusters, where similarity stands for the comparison of the distance between clusters with their size. Knowing the ground truth labels is not needed. Davies Bouldin Index can be used to determine the number of optimal clusters.
- Fowlkes-Mallows Index is the geometric mean of the pairwise precision and sensitivity.
- Caliński-Harabasz Index, a.k.a. Variance Ratio Criterion, is the ratio of the sum of between-clusters dispersion and of inter-cluster dispersion for all clusters. Dispersion is defined as the sum of distances squared. Knowing the ground truth labels is not required.

An important property that some clustering quality measurements have is "symmetry": changing one of the arguments does not modify the obtained value for that metric. This way, this symmetric metric can be used to measure the consensus between two independent labeling strategies over the same dataset.

In Table 2.4 the optimal, worst and interval of values and the symmetry property compliance for each of the indicated clustering quality measurements have been gathered.

It should be noted that various clustering quality metrics are affected by the number of clusters and the number of samples. Non-adjusted measures show a dependency between both terms, sometimes greatly increasing the closer the number of clusters is to the quantity of samples. Therefore, given a certain value of overlapping subsamples of the dataset, only adjusted measures can be used confidently as a consensus index to check the average stability of clustering techniques.

2.5.3 Time measurements

Training and running times are the most commonly calculated time-related metrics. Training time indicates the time used by the model during the training phase. It is generally used as

Table 2.4: Interval, worst and optimal values for the described clustering quality metrics. Also, their compliance with the "symmetry" property.

Performance metric	Interval	Worst value	Optimal value	Is symmetric?
Purity	[0, 1]	0	1	
Completeness	[0, 1]	0	1	✗
Homogeneity	[0, 1]	0	1	✗
V-measure	[0, 1]	0	1	✓
Adjusted Rand Index	[-0.5, 1]	-0.5	1	✓
Normalized Mutual Information	[0, 1]	0	1	✓
Silhouette	[-1, 1]	-1	1	
Davies Bouldin Index	$[0, \infty)$	Higher	0	
Fowlkes-Mallows Index	[0, 1]	0	1	
Calinski-Harabasz Index	None	Lower	Higher	

a discriminant metric between models whose values of certain performance and clustering quality metrics are similar for the same task. Models of very different nature may have very distinct training times. For example, lazy learners normally train almost immediately; some shallow ANNs may take minutes, while some DL techniques spend even weeks.

Running time indicates the time spent by the model when using unseen data, during the testing of the model. It is almost never reported, as it will be our case, as more often than not it is in the few seconds range or even less. However, it may be minutes in the case of lazy learners, several times larger than their training times.

Several approaches have been followed to reduce these times, especially the training one, subsection 2.5.4. Another one quite popular with DNNs is by using "transfer learning", as opposed to the commonly used "training from scratch". In "transfer learning" the weights of a DNN, which was previously trained with another dataset, are used as the initial weights of the desired DNN that will work in a different but normally related issue. The main difficulty and risk of "transfer learning" is to mistakenly consider both datasets as similar or related, because this may cause the DNN to learn things that are not relevant to the actual problem to be addressed.

2.5.4 Computational costs and resources requirements

Not as popular as the previous model evaluation metrics, this group mainly comprises Central Processing Unit (CPU) load and characteristics, and disk and memory space.

Most DNNs have extremely high computational costs, especially compared to most ANNs, for three reasons: number of hyperparameters for each neuron and synapse (in the order of millions in many CNN variants), quantity of layers, and number of neurons per layer. Configuring and fine-tuning such high quantity of hyperparameters is quite time-consuming and complex unless the user has experience or a good technical background.

On the other hand, it is generally considered mandatory the usage of multi-threading or multi-core CPUs, multiprocessing systems and even Graphics Processing Units (GPUs) during the training phase in order to reduce the training times to far more acceptable long levels, in some cases from weeks to days or hours. Similarly, the introduction of Field Programmable Gate Arrayss (FPGAs) and other over-specialized hardware as accelerators for

DNNs [Mittal, 2020] may decrease the power consumption and training times even further.

Requirements of memory and disk space have increased in the last decades as DNNs became more prevalent, the ample usage of heavily parallelized High-Performance Computing (HPC) systems, and the popularity of computer vision.

Chapter 3

Description and analysis of two novel ontogenic neural architectures: Modular Hybrid Growing Neural Gas and Supervised Reconfigurable Growing Neural Gas

In this chapter, two novel architectures are described, which are extensions of the ontogenetic and unsupervised Artificial Neural Network (ANN) called Growing Neural Gas (GNG) [Fritzke, 1995], already described in subsubsection 2.1.2.5. On the one hand, we describe the Modular Hybrid Growing Neural Gas (MyGNG), its formal definition, and analyze its main hyperparameters with a synthetic and a real datasets. On the other hand, we describe the Supervised Reconfigurable Growing Neural Gas (SupeRGNG), including its formal definition and biological plausibility. After that, several synthetic datasets will be used to analyze the SupeRGNG, and the recommended values of some of its hyperparameters for datasets with particular characteristics are exposed. In this document the relevance of a hyperparameter is defined by how it influences both the behavior and the classification performance of the ANN, in out case whether the MyGNG or the SupeRGNG.

3.1 Description of the Modular Hybrid Growing Neural Gas

MyGNG is a hybrid ontogenic neural architecture that is an improved yet simpler version of the one first described in [Sosa-Marrero et al., 2021]. The current version was introduced in [Cabrera-León et al., 2024b]. Two modules can be found in the MyGNG, Figure 3.1. The first module is built with an unsupervised, self-organizing and ontogenetic model, Fritzke's GNG [Fritzke, 1995]. The second module was based on the perceptron, a popular supervised neural architecture [Rosenblatt, 1961; Widrow and Lehr, 1990].

MyGNG is hierarchically organized the way it is on purpose: first the data clustering and later the data labeling. Clustering, which other models do not do, is aimed to simplify the input data by projecting it to a space with more dimensions while maintaining its topology to facilitate the classification done later by the labeling module [Cabrera-León

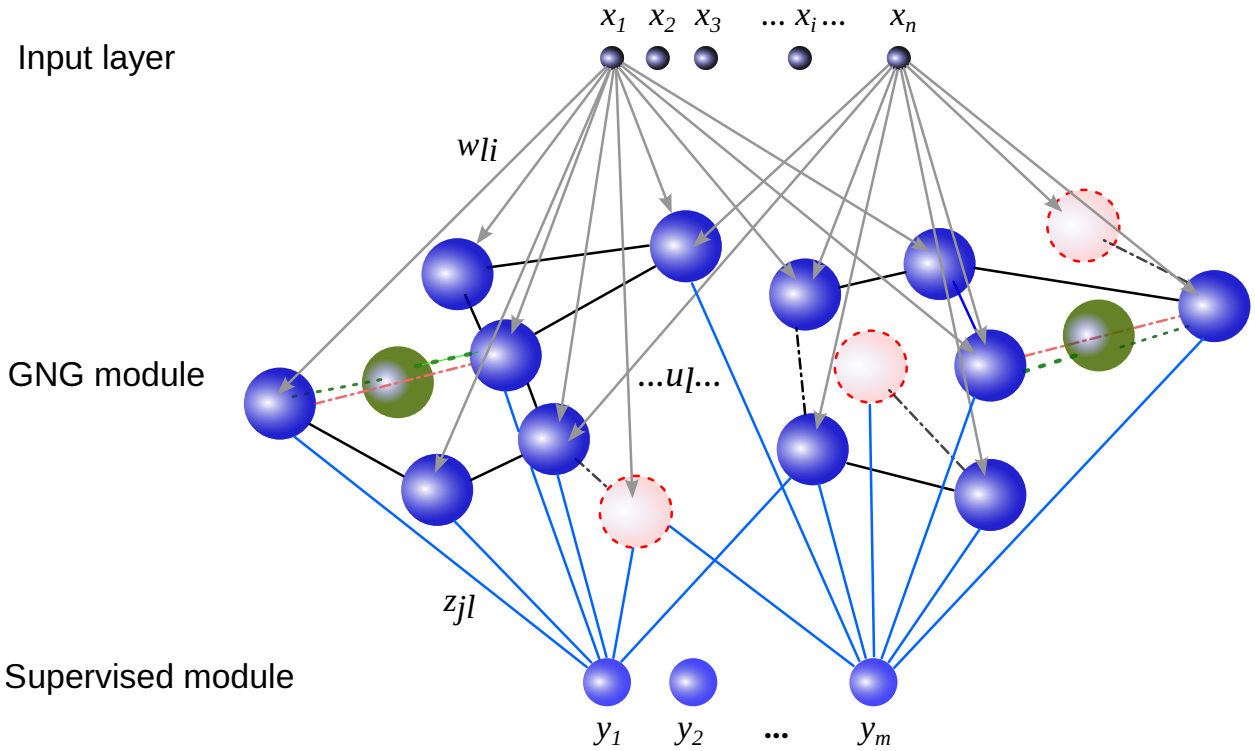


Figure 3.1: Structure of the Modular Hybrid Growing Neural Gas, where x_i is the i-th component of the input vector; w_{li}, the weight of neuron li; and u_l, a neuron/unit of the GNG module.

et al., 2024b]. Apart from the expected increased performance compared to the perceptron, it may also decrease training times, as it also occurs in other hybrid architectures such as the Counterpropagation Network [Hecht-Nielsen, 1987; Cabrera-León et al., 2018a].

These two modules learn sequentially. That is, the training of the MyGNG is done in cascade: the second module (named "Supervised" in Figure 3.1) uses for its training the labels of the data and the output of the first module (named "GNG") after this one has been trained.

In Figure 3.1 the structure of this improved MyGNG is depicted, with the input layer and its two sequential modules. As in Figure 2.4 and subsubsection 2.1.2.5, colors of the neurons indicate where two biologically-related processes have happened (blue is the base one, and it is used for neurons that are adapting to the input data): neurogenesis and neural apoptosis. Green neurons represent new neurons, that is, they have been recently created (that is, neurogenesis) where the GNG algorithm considered more convenient (that is, between the neuron with the greatest error and its neighbor with the greatest error). These green neurons require that the old connections are deleted (red lines) and new connections are created (green dashed lines). Conversely, red neurons are those that have been removed (that is, neural apoptosis), which occur after all connections to them have been deleted.

The main difference between the improved MyGNG presented in this work and the original one in [Sosa-Marrero et al., 2021] is how the "Supervised module" is built, which will be explained later in Equation 3.2. This module has now less complexity as it is based on a perceptron instead of the complex "Supervised module" of the original MyGNG, which made use of neural neighborhoods [Sosa-Marrero et al., 2021]. These neural neighborhoods are unnecessary and, hence, not used in the improved MyGNG introduced in the current work.

Regarding the first module of the MyGNG presented in this work, in [Fritzke, 1995, 1997b] the GNG is described as a self-organization map based on a dynamic graph of connected neurons. Starting from a low number of interconnected neurons, this graph will adapt, shrink and grow, hence producing topological learning that will allow clustering the input space. This generation and continuous update of the graph is made by a competitive learning algorithm [Kamimura, 2010], where the winner neuron[1] s_1 is the one whose weights (ω) are the most similar to the input vector ξ . Equation 3.1), where ε_b and ε_n are the learning rates for the winner and its neighbors, respectively [Fritzke, 1995], indicates that the adjustment of the winner neuron and its direct topological neighbors defines the adaptation process.

$$\begin{aligned} \Delta w_{s_1} &= \varepsilon_b(\xi - w_{s_1}) \\ \Delta w_n &= \varepsilon_n(\xi - w_n) \text{ for all direct neighbor } n \text{ of } s_1 \end{aligned} \tag{3.1}$$

It should be noted that only the GNG equations that have been considered the most relevant are explained in this work. For the full list of equations, we recommend consulting subsubsection 2.1.2.5 and Fritzke's cited works [Fritzke, 1995, 1997b]. These most relevant parts are related to the two processes that make GNG stand out from other ontogenic neural architectures: neurogenesis and neural apoptosis.

A local error variable is calculated for the winner neuron in each iteration, Equation 2.15. This error is related to the neurogenesis (that is, the creation of a neuron) process because it allows identifying regions where the input signals are not sufficiently correctly represented. That is, a new neuron needs to be inserted between the unit q with the maximum error and its neighbor f in the graph that has the highest error, Equation 2.16. An insertion occurs every λ adaptation steps. Error variables of these units q and f are reduced in proportion to the parameter β.

Altering connections modifies the network topology. A new connection is created on each adaptation step between the winner and the second winner neurons. Conversely, a connection is removed when the value of its age property is above the a_{max} parameter. Neural apoptosis (that is, the deletion of a neuron) occurs when it becomes isolated after all the connections to that neuron got erased.

The responsible of the hybrid character of this MyGNG is the addition of a monolayer-perceptron-based output module (supervised learning) after the GNG-based one (unsupervised learning). The learning process of the perceptron is given by the "Perceptron rule" shown in (3.2), which indicates how the weights are updated [Widrow and Lehr, 1990]. In this equation, $x(k)$ is an input; $\omega(k)$, weights; ρ, the learning rate; and $\tilde{e}(k) = d(k) - y(k) = d(k) - \text{sgn}[\omega^T(k) \cdot x(k)]$, being sgn the sign function; $d(k)$, the desired output value of the perceptron for the input $x(k)$; and $y(k)$, the obtained output value of the perceptron for that input.

$$\omega(k+1) = \omega(k) + \rho \cdot \frac{\tilde{e}(k)}{2} \cdot x(k) \tag{3.2}$$

The following algorithm describes how the MyGNG works, where the parts regarding the GNG were loosely based on Fritzke's GNG algorithm [Fritzke, 1995]:

[1] In this work, s_i indicates the neuron with the i-th minimum Euclidean distance between the input vector and the neuron's weights vector.

1. Create two neurons, a and b, with weights ω_a and ω_b, respectively.
2. Extract a sample ξ from the input space or, alternatively, generate an input signal according to the probability density function $P(\xi)$.
3. Find s_1 and s_2, the two neurons that are the nearest to the input sample.
4. All connections to s_1 have their *age* property incremented.
5. Increment the local error variable of s_1. These errors are used later to find where to insert a new neuron.
6. Move s_1 closer to ξ (3.1). Similarly, move all the direct neighbors of s_1 but by a lesser amount.
7. The age of the connection between s_1 and s_2 is reset to 0. Create it if it did not exist.
8. Neural apoptosis: after removing all connections whose *age* is greater than a_{max}, delete all neurons without connections.
9. Neurogenesis: every λ iterations, insert a new neuron between the neuron with the largest error and its direct neighbor with the largest error. The connection between the erroneous neurons is deleted, and two new connections are created: between each of them and the new one. The error of the erroneous neurons is diminished.
10. Decrease all error variables.
11. Go to Step 2 if the stopping criterion (for example, epochs, performance metric, size of the network...) of the GNG is not met yet. In our case, it is the epochs or number of times all the input samples are used for training the GNG.
12. Obtain an output from the GNG and the associated class label (that is, the expected output of the perceptron).
13. Update the weights of the perceptron (3.2).
14. Go to Step 12 if the stopping criterion of the perceptron is still not met. In our case, it is the epochs used for training the perceptron.

3.2 Behavior of the Modular Hybrid Growing Neural Gas with overlapping data

3.2.1 Datasets for the study of Modular Hybrid Growing Neural Gas

Two datasets were used to study the MyGNG. On the first hand, the "blobs" dataset comprises three isotropic Gaussian 2D blobs of samples, each one belonging to a different class and with the same number of samples per class. The "make_blobs" function found at scikit-learn was used to create the blobs [Pedregosa et al., 2011]. The samples were generated with the density function shown in Equation 3.8, where i is the dimension, 1 or 2, which

corresponds to the X and Y axis; μ_i, the mean of the class in the i-th dimension; and σ_i, the standard deviation of the class in the i-th dimension. The purpose of using this dataset was to check the behavior of the network with increasing levels of overlapping. Unlike with the SupeRGNG subsubsection 3.5.1.3, instead of varying the distances by moving the centers μ of one of the clusters and keeping the same standard deviation σ of the three clusters to vary the overlapping, the centers were the same but the values of the standard deviation of the clusters were varied. Solely three variants were used, with values of $\sigma = \{0.5, 1.0, 1.5\}$.

On the other hand, the famous real dataset "iris" [Fisher, 1936; Anderson, 1936] was used to emulate the behavior of the MyGNG with different levels of overlapping between three classes of flowers: "setosa", "versicolor" and "virginica". Fifty samples of each class exists.

Depictions of both datasets are shown in subfigures a) in Figure 3.2, Figure 3.3, Figure 3.4, and Figure 3.5.

3.2.2 Analysis of the behavior of the MyGNG with overlapping data

The hyperparameters of the MyGNG includes those of the networks its modules are built from: GNG and single-layer perceptron. The hyperparameters of the GNG can be grouped according to its functionality: neuron movement ($\varepsilon_b, \varepsilon_n$), error decreasing ($\beta, d$), and network growing ($\lambda, a_{max}$). Conversely, in the perceptron the learning rate ρ is the hyperparameter that can be analyzed[2].

There is no stopping criteria defined for the GNG algorithm by Fritzke [1995], although this author mentioned performance metrics or size of the network as possible. The latter have been used in the MyGNG.

The optimal set of features for "blobs" that was found using grid search was: $max\_nodes = 125, a_{max} = 5, \lambda = 25, \varepsilon_b = 0.65, \varepsilon_n = 0.0125, \beta = 0.7, d = 0.16, epochs = 10, \rho = 0.01, epochs\_perceptron = 25$. In Figure 3.2, Figure 3.3 and Figure 3.4 scatterplots of the original dataset and the correctly classified samples and missclassifications ("OK" and "NOk" in the subfigures b), respectively) in the three different overlapping levels in the "blobs" dataset are depicted, together with the three binary Receiver Operating Characteristic (ROC) curves. The ROC curves show the classification degradation with increasing levels of overlapping. As expected, the greater the overlapping, the lower the performance of the MyGNG. In the most extreme case when σ has the greatest value, the class situated in the middle is barely correctly classified so the Area Under the Curve (AUC) is not good, subfigures b) and d) in Figure 3.4, respectively.

The famous real dataset "iris" was also used to emulate the behavior of the MyGNG with two overlapped classes. The optimal set of features using the same approach was: $max\_nodes = 75, a_{max} = 5, \lambda = 25, \varepsilon_b = 0.75, \varepsilon_n = 0.01, \beta = 0.7, d = 0.16, epochs = 10, \rho = 0.01, epochs\_perceptron = 25$ As shown in Figure 3.5, the "setosa" class is far from the others, so the AUC in such case was 1, whereas with the rest of classes lower yet still good values were yielded. As with "blobs", errors were situated in the border of the overlapped classes, where the most difficult samples to classify are located. In Table 3.1, the performance of MyGNG with "iris" is shown.

[2]There is a parameter used as epochs in the perceptron, *epochs_perceptron*, but it was given a fixed value.

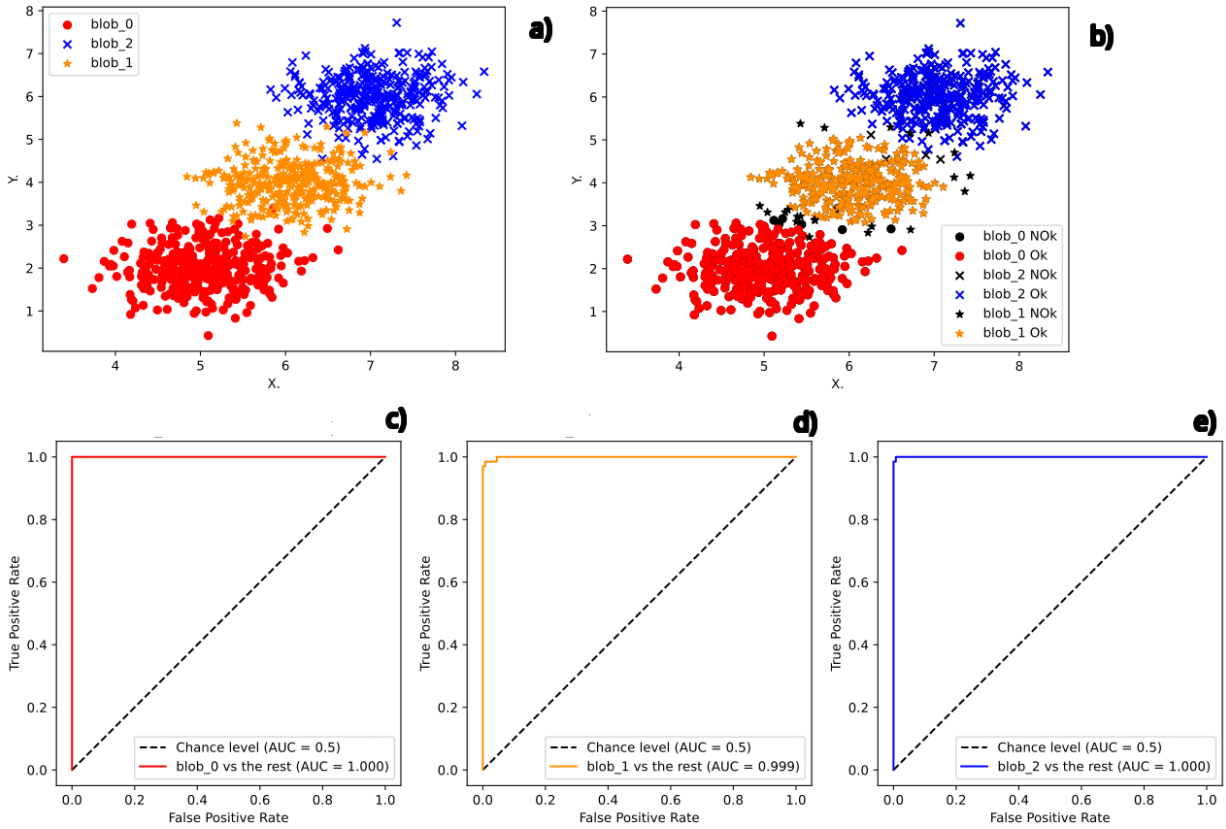


Figure 3.2: Example of the behavior of the MyGNG with the "blobs" dataset with $\sigma = 0.5$. a) Scatterplot of the dataset, b) Scatterplot of the correct and wrong classifications (per class), c) ROC curve and AUC of the first class vs the rest, d) ROC curve and AUC of the second class vs the rest, e) ROC curve and AUC of the third class vs the rest.

Table 3.1: Performance results of the MyGNG with the "iris" dataset.

One vs Rest	Accuracy	Sensitivity	Specificity	Precision	AUC
setosa vs Rest	0.9974	0.9847	0.9874	0.9923	1
versicolor vs Rest	0.9782	0.9651	0.9742	0.9623	0.9955
virginica vs Rest	0.9691	0.9513	0.9454	0.9736	0.9851

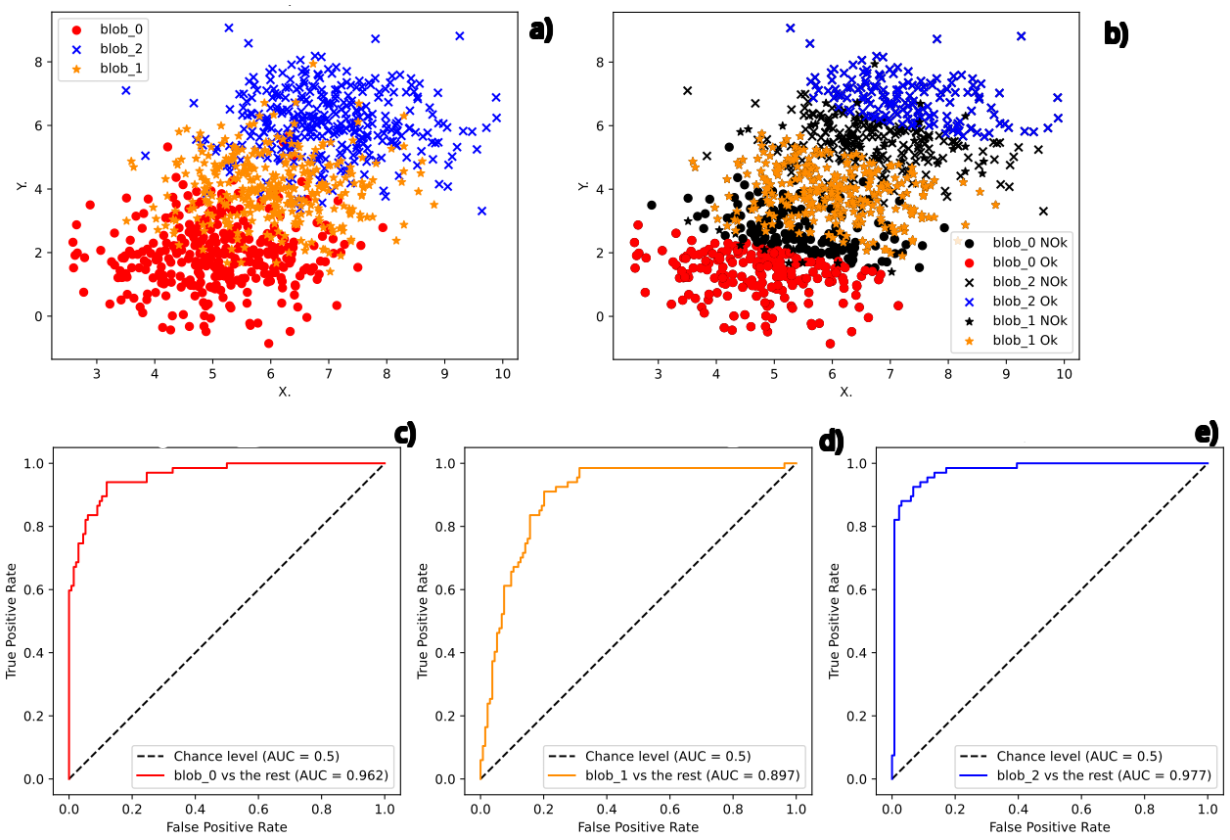


Figure 3.3: Example of the behavior of the MyGNG with the "blobs" dataset with $\sigma = 1.0$. a) Scatterplot of the dataset, b) Scatterplot of the correct and wrong classifications (per class), c) ROC curve and AUC of the first class vs the rest, d) ROC curve and AUC of the second class vs the rest, e) ROC curve and AUC of the third class vs the rest.

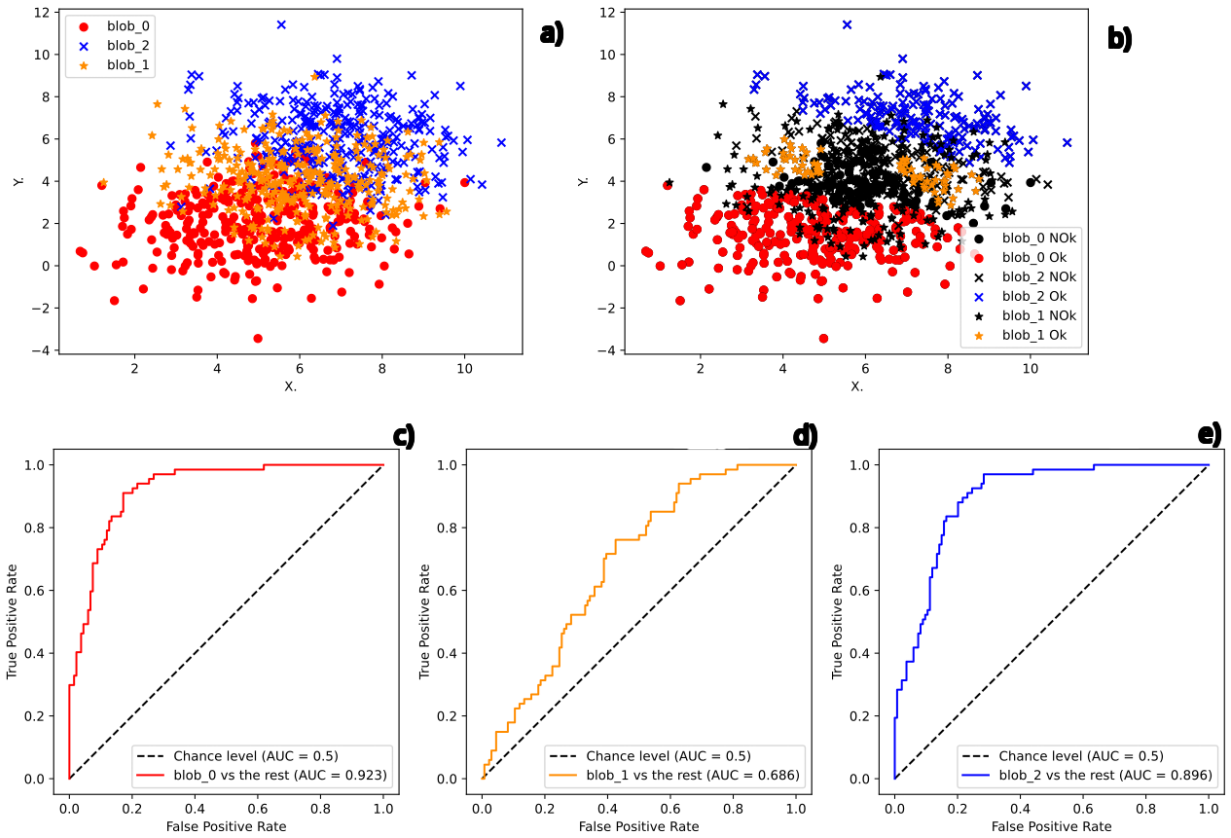


Figure 3.4: Example of the behavior of the MyGNG with the "blobs" dataset with $\sigma = 1.5$. a) Scatterplot of the dataset, b) Scatterplot of the correct and wrong classifications (per class), c) ROC curve and AUC of the first class vs the rest, d) ROC curve and AUC of the second class vs the rest, e) ROC curve and AUC of the third class vs the rest.

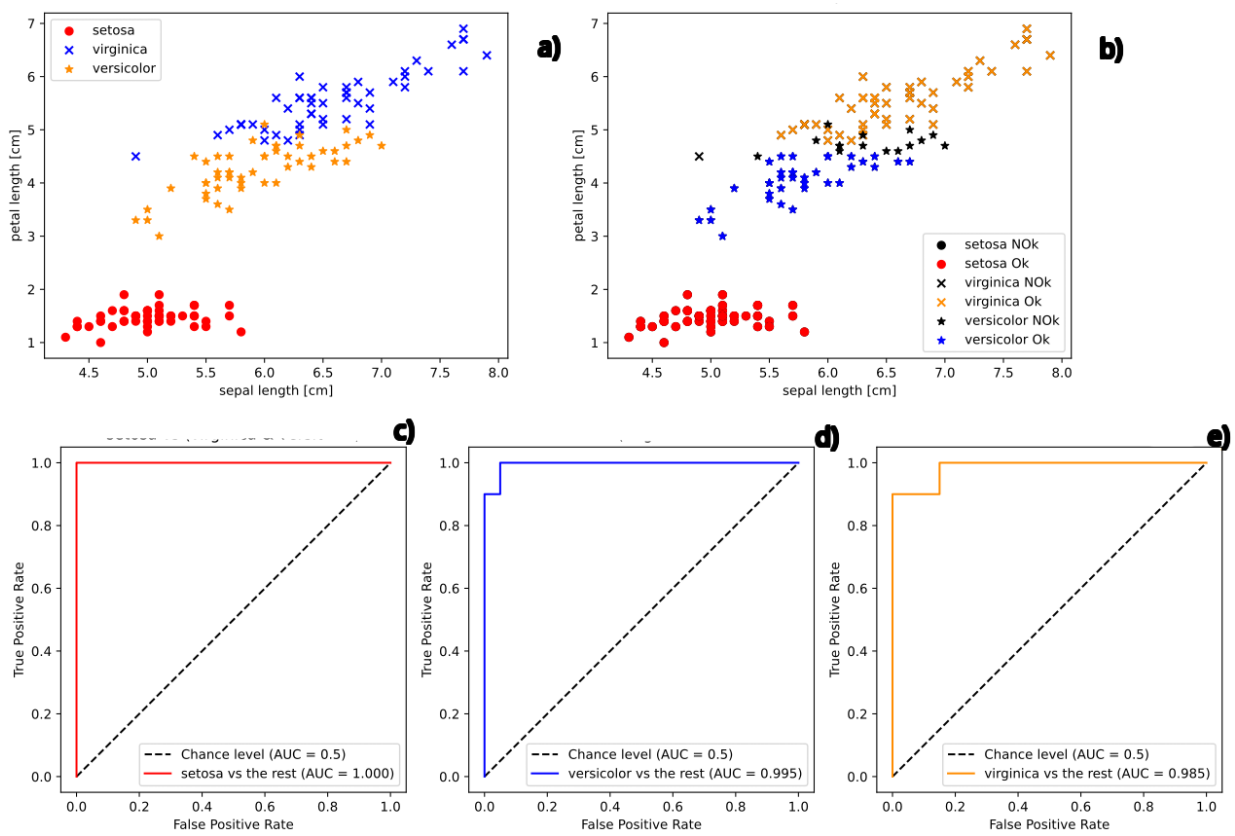


Figure 3.5: Example of the behavior of the MyGNG with the "iris" dataset. a) Scatterplot of the dataset, b) Scatterplot of the correct and wrong classifications (per class), c) ROC curve and AUC of the "setosa" class vs the rest, d) ROC curve and AUC of the "versicolor" class vs the rest, e) ROC curve and AUC of the "virginica" class vs the rest.

3.3 Description of the Supervised Reconfigurable Growing Neural Gas

As deduced from its name, the SupeRGNG is based on the GNG and includes several mechanisms regarding its reconfigurable capability that require the labels associated to the input data (supervised learning). Additionally, it includes early stopping, a supervised mechanism aimed to shorten the training time, which can also be found in other ANNs, which stops training when no further improvements in a performance metric is found for a certain period. As a regularization mechanism, early stopping aims to minimize the generalization error by preventing overfitting, a risk common to supervised approaches.

The reconfigurable characteristic, whose goal is to refine the clusters that were obtained in an unsupervised way by the GNG, can be divided in two processes, which are performed with the same frequency and sequentially in the next order:

1. Disconnection, where neurons with dissimilar labels that were erroneously connected, became disconnected. That is, the interclass boundaries are made.

2. Reconnection, where a new connection between clusters with identical labels that were mistakenly separated is created in order to unite them. That is, the intraclass connections are made.

The previous supervised procedures can be carried out dynamically after any training epoch and with the desired frequency of iterations after that epoch, being both values defined by the user. Conversely, in our previous variant of SupeRGNG, which also had no early stopping, the reconfigurable capability was only done after the unsupervised learning of the original GNG finished [Cabrera-León et al., 2023]. Disconnections and reconnections in both variants are carried out with the same frequency, and one after the other.

In Figure 3.6 the topology of the SupeRGNG is shown, where red neurons indicate dead neurons that are going to be deleted; green neurons, new ones that are inserted in regions where the error was the highest; dotted-dashed edges, connections that are eliminated both when there is a dead neuron or when a new neuron is added; and, dotted edges, connections that are created during the insertion of a new neuron. The main difference with Figure 2.4 is related to what to do with the red-blue unit located in the middle, which represents a neuron that belongs to the light blue cluster but was erroneously assigned to the dark blue one. Whilst in the GNG the topology would remain unchanged, in the SupeRGNG edges connected to that neuron will be corrected, whether through deletion (red, dotted line) or creation (green, dotted line).

3.3.1 Model definition

Both the original GNG and the SupeRGNG have quite similar neuron models, network topologies and learning algorithms, subsubsection 2.1.2.5.

On this basis, a SupeRGNG g can also be formally described by the tuple $g := (\mathcal{U}, \mathcal{C}, \theta)$, where $\mathcal{U}$ is the set of units (a.k.a. vertices, nodes or neurons); $\mathcal{C}$, the set of connections (a.k.a. edges or links); and θ, the set of hyperparameters [Fritzke, 1995; Kerdels, 2016]. It can also be seen as a set of connected components within a simple, normally disconnected, graph: an undirected graph without loops nor multiple edges [Clark and Allan Holton, 1991; Rosen, 2004].

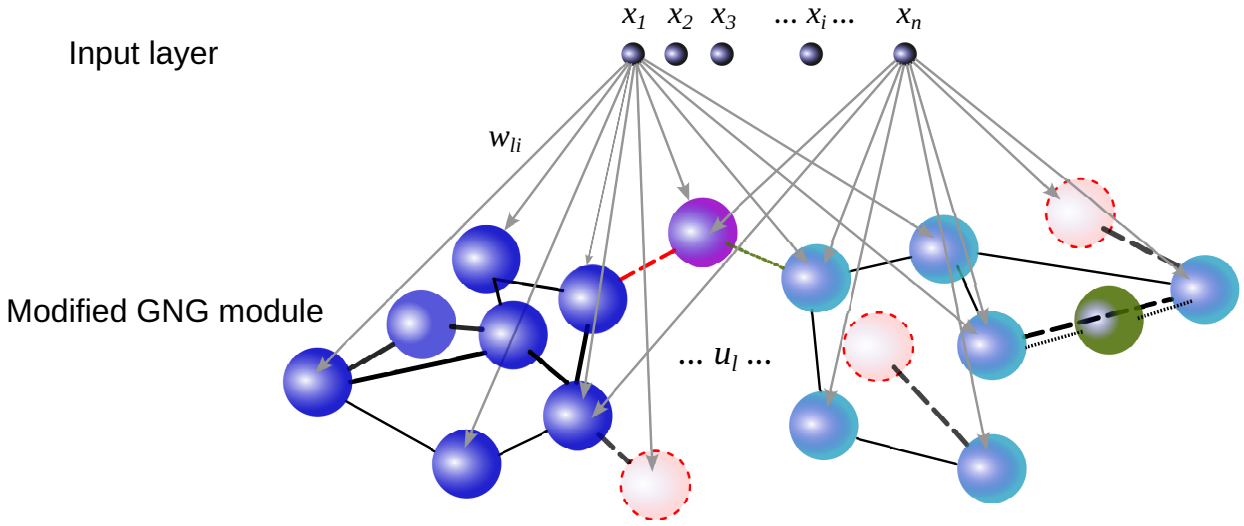


Figure 3.6: Topology of our SupeRGNG, showing examples of dead neurons (red), new neurons (green), unsupervised disconnections (black, dotted lines), supervised disconnections (red, dotted lines), and supervised reconnections (green, dotted lines).

A neuron $u \in \mathcal{U}$ of a SupeRGNG has associated a reference vector $\boldsymbol{\omega}$, which relates inputs $\mathbf{x} \in \mathbb{R}^n$, $n \in \mathbb{N}$ with u, an accumulated error $e \in \mathbb{R}$, and, in contrast to the neuron of the original GNG, subsubsection 2.1.2.5, the label that most frequently was assigned to this unit during training $y \in \mathbb{R}$. For the sake of convenience, for unit u we will define $\boldsymbol{\omega}_u$, e_u and y_u as synonyms of the previous elements, respectively.

A connection $l \in \mathcal{C}$ can be depicted by the tuple $l := (u, v)$, where $u, v \in \mathcal{U}$ are the pair of units that are being connected. As in GNG and unlike in other ANNs, connections in SupeRGNG are unweighted and they are used to define the topological structure of the network. Each connection has an age associated $age_{(u,v)} \in \mathbb{N} \cup \{0\}$.

The set of hyperparameters in the SupeRGNG extends those found in the GNG. The common hyperparameters are also constant over time, and have the same meaning and use subsubsection 2.1.2.5. This extended set is comprised of $\theta := \{\varepsilon_c, \varepsilon_n, d, \beta, \lambda, a_{max}, \varsigma, epochs, \eta, \iota\}$, and the new ones are:

- η, which indicates from which epoch the interclass boundaries (disconnection process) and the intraclass connections (reconnection process) begin to be made.
- ι, which specifies how often (in number of iterations) the disconnections and reconnections will be made after the initial η epochs.

In the next paragraphs we describe the SupeRGNG, its neuron model, its network topology (including both the network framework and the interconnection structure), and its learning algorithm.

Neuron model: as inferred from the definition of a unit above, the neuron model in the SupeRGNG only differs from that in the GNG in that there is a label saved in each neuron. This added property does not modify the formal definitions of the propagation, activation and output functions.

Network topology: there is no grid structure in this supervised version of GNG.

- Identical number of layers and layer connections as in the GNG:
  - A single layer of neurons[3], starting from 2 neurons and growing to the maximum indicated size[4].
  - All neurons in the input layer are connected to the neurons in the main layer.
  - The neurons in the main layer are partially interconnected.
- The topological neighborhood is the same as in the regular GNG, Equation 2.7.
- The positions of the first and second neurons are identically calculated, Equation 2.8 and Equation 2.9.
- Connection senescence: as in the GNG, connections emanating from the current winner neuron c age after every training step, although the age of the connection between c and its nearest neighbor is refreshed (*i.e.*, age reset to 0), Equation 2.10.
- Neural apoptosis: the death of a neuron happens when there are no connections emanating from it (that is, it has no neighbors), which happens due to the aforementioned "connection senescence", Equation 2.10.
- Lateral connections update follows the same Hebbian competitive learning as the GNG, Equation 2.12. A connection is deleted when its age reaches the value indicated by the hyperparameter a_{max}. Additionally, connections in SupeRGNG can also be deleted without taking into account its value of a_{max} or created during the disconnection and reconnection processes, respectively.
- Weight update, due to being a competitive ANN, is different in the winner neuron c and its direct topological neighbors in the graph, $\mathcal{N}(c)$, Equation 2.13.
- Error update: each neuron has a local error variable related to the distance between the input $\mathbf{x}$ and the winner neuron c, Equation 2.14. During the adaptation, accumulating the squared distances to this error helps to identify neurons situated in areas of the input space where the error in the mapping is high. In such areas new neurons are inserted in order to reduce the error. Considering the decaying hyperparameter $d \in (0, 1)$, the error for neuron u, e_u, is updated at the end of each iteration according to Equation 2.15, a decaying process that affects every neuron $u \in \mathcal{U}$.
- Neurogenesis: a new neuron is created when the current training step is a multiple of the hyperparameter λ, Equation 2.16.
- Path: as this family of ANNs can be seen as simple graphs and not multigraphs (*i.e.* without multiple connections between the same pair of endpoints) in graph theory [Clark and Allan Holton, 1991; Rosen, 2004], the formal definition of a path $\mathcal{P}(s, t)$ can be simplified as the sequence of vertices (that is, units) that connects two of them $s, t \in \mathcal{U}$:

$$\begin{gathered}\mathcal{P}(s,t) := (v_0, \ldots, v_k) \mid \exists (v_{i-1}, v_i) \in \mathcal{C}, i = 1, 2, \ldots, n, \\ v_i \in \mathcal{U},\ v_0 = s,\ v_k = t\end{gathered} \tag{3.3}$$

- Connected component: a subgraph in which each pair of units is connected with each other via a path, and no path exists that connects any unit of this subgraph

[3]We are not counting the input layer, which also exists, as it commonly occurs in the majority of ANNs.
[4]See footnote 4 in page 59.

to any unit in the rest of the graph. In our case, a connected component can be considered a cluster. In graph theory, the connected components within a graph are the set of disjoint connected subgraphs [Clark and Allan Holton, 1991; Rosen, 2004]. Based on [Clark and Allan Holton, 1991], we can define the connected component s_g of the graph g as follows:

$$s_g := (\mathcal{U}_{s_g}, \mathcal{C}_{s_g}, \theta),\ g := (\mathcal{U}, \mathcal{C}, \theta) \mid \mathcal{U}_{s_g} \subseteq \mathcal{U},\ \mathcal{C}_{s_g} \subseteq \mathcal{C} \tag{3.4}$$

where $\mathcal{U}$ is the set of units; $\mathcal{C}$, the set of connections; and θ, the set of hyperparameters.

On the other hand, each connected component has associated the most common, according to the Simple Majority Voting (SMV) method, label associated with the units that form it. This label y_{s_g} can be calculated as the mode, the most frequent value.

- Tuning inter-class boundaries: includes disconnections and reconnections, which are performed after η epochs has passed and, after that, with a frequency of ι iterations.

  a) Assign a label to each cluster: As each cluster will correspond to a connected component, this assignation is based on the SMV of the labels y_u of each neuron u within the connected component.

  b) Find the inter-class boundaries: there are connections between neurons (or group of them, hence they can be called clusters) with different labels. That is, check $foralll \in CC_1,\ m \in CC_2 \mid y_l \neq y_m \wedge \exists\ (l, m) \in \mathcal{C}$, where $CC_1, CC_2 \subset g$ are different connected components in the network graph g; $y_l, y_m \in \mathcal{Y}$, the labels associated with units l and m, respectively; and $\mathcal{C}$, the set of connections.

  c) Separating clusters with different labels: after finding the inter-class boundaries, disconnecting the neurons or clusters with different labels is straightforward.

$$\forall (l, m) \in \mathcal{C}, \mathcal{C} := \mathcal{C} - \{(l, m)\}, \text{if } y_l \neq y_m \wedge l \in s_{g_1}, m \in s_{g_2} \tag{3.5}$$

  where $s_{g_1}, s_{g_2} \subset g$ are different connected components in the network graph g; $y_l, y_m \in \mathcal{Y}$, the labels associated with units l and m, respectively; (l, m), a connection between these units; and $\mathcal{C}$, the set of connections.

  d) Connecting clusters with the same labels: although random units could have been chosen as connection points, the nearest neurons in each cluster were finally selected because they were deemed more convenient. The nearest neurons are those that had the minimum distance according to the definition of distance that is used, such as Equation 2.3 or Equation 2.4.

$$\begin{gathered} \forall (l, m) \notin \mathcal{C},\ \mathcal{C} := \mathcal{C} \cup \{(l, m)\},\ age_{(l,\, m)} := 0, \\ \text{if } y_l = y_m \ \wedge\ l \in s_{g_1}, m \in s_{g_2} \ \wedge \underset{l \in s_{g_1},\, m \in s_{g_2}}{\arg\min}\ (d(l, m)) \end{gathered} \tag{3.6}$$

  where $s_{g_1}, s_{g_2} \subset g$ are different connected components in the network graph g; $y_l, y_m \in \mathcal{Y}$, the labels associated with units l and m, respectively; (l, m), a connection between these units; $age_{(l,m)}$, the age of the connection between these units, as in Equation 2.16; $d(l, m)$, the distance between them, such as Equation 2.3 or Equation 2.4; and $\mathcal{C}$, the set of connections.

Learning algorithm:

- Early stopping[5]: if the disconnection or reconnection processes were active, the early stopping process is launched. In other words, after the "warm-up period", where the model trains for η epochs without the early stopping process getting activated, the training stops when no improvements have occurred in the current value *monitor* of the metric being monitored (default, AUC[6]) for the last *patience* consultationswith the validation set. Assuming that *monitor* is a metric where higher values mean better, and it is checked after the "warm-up period" of η epochs, an improvement better than the desired minimum value $delta_{min}$ has occurred in the epoch *ep* in the window of consultations of size *patience* is explained by Equation 3.7. If there have not been any improvement during that window, the state and performance results of the network from this "best epoch" are restored.

$$(monitor_{ep} - delta_{min}) > monitor_{best\_epoch} \tag{3.7}$$

  Early stopping is the most popular regularization method in Deep Learning (DL) because it is both simple and effective [Goodfellow et al., 2016], although its use is not limited to Deep Neural Networks (DNNs) as it was already being used in shallow ANNs before deep ones became popular [Zhang and Yu, 2005]. Early stopping allows the network to not continue training when the monitored metric has stopped improving for several epochs instead of training during the whole number of epochs. Hence, its goal is to shorten the training time of the network without impairing its classification performance. Early stopping is typically used on a validation set, and is designed to make the algorithm halt as soon as overfitting starts to happen [Zhang and Yu, 2005; Goodfellow et al., 2016]. Generally, early stopping implies that every time lower validation set errors are obtained (which it is expected to imply lower test set errors too), a copy of the model parameters needs to be stored. So, when the training stops, these parameters are returned instead of the latest ones [Goodfellow et al., 2016]. This extra space cost needed for the copy is considered negligible, whereas the periodic evaluation of the validation set during training might incur a significant cost [Goodfellow et al., 2016].

- Algorithm steps: a SupeRGNG is initialized with two units with prototype vectors ω, usually randomly, chosen from the input data, and their accumulated error variables e are set to 0. The set of connections $\mathcal{C}$ is initially empty. Then, the SupeRGNG grows by processing samples extracted from the training data[7]. The learning algorithm of the SupeRGNG comprises the next 12 steps, where Steps 1 to 10 and Step 13 are identical to those of the GNG:

  Step 1: Start with two units, $U = a, b \mid e_a := 0, e_b := 0$, and ω_a and ω_b are reference vectors usually randomly chosen from the input data. $\mathcal{C} := \emptyset$

[5]The early stopping in other models and implementations is not identical to the one presented here, as, generally, the hyperparameters related to the "warm-up period" and the number of times without improvements are both measured in epochs. Conversely, in SupeRGNG the former is in epochs as it is related to η, whereas the latter is in number of consultations and it is related to ι, which refers to iterations.

[6]As the greater the value of AUC is, the better it is considered, in Equation 3.7 there is a ">" symbol when monitoring the metric. For other metrics, such as loss, the opposite happens, so a "<" is more appropriate.

[7]See footnote 10 in page 62.

Step 2: Iterate through the training data S.

Step 3: Find the nearest unit c and the second nearest unit c_2, Equation 2.8 and Equation 2.9, respectively. They are also called Best Matching Unit (BMU) and second BMU.

Step 4: Increment the age of all edges emanating from unit c, Equation 2.10.

Step 5: Add the squared distance between the observation and the associated reference vector of c to the accumulated error of c, Equation 2.14.

Step 6: Move c and its direct topological neighbors towards the observation by the fractions ε_c and ε_n, respectively, of the total distance, Equation 2.13.

Step 7: If there is a connection between c and c_2, set the age of this connection to zero, Equation 2.10. If this connection does not exist, create it, Equation 2.12.

Step 8: Remove connections with an age larger than a_{max}, Equation 2.12. If this results in units having no emanating connections, remove them as well, Equation 2.11.

Step 9: If the number of steps so far is an integer multiple of parameter λ, insert a new unit, Equation 2.16:

a) Determine the unit q with the maximum accumulated error.

b) Insert a new unit r halfway between q and its neighbor f with the largest error variable. The label of r is the same as that of the unit q.

c) Insert edges connecting the new unit r with q and f, and remove the original edge between q and f.

d) Decrease the error variables of q and f by multiplying them with the constant β. Initialize the error variable of r with the new value of the error variable of q.

Step 10: Decrease the error variables of all neurons by multiplying them with a constant d, Equation 2.15.

Step 11: If the number of epochs is at least η and the current number of iterations so far is a multiple of ι, tune the inter-class boundaries.

a) Assign a label to each cluster.

b) Find the inter-class boundaries.

c) If labels of the neurons or clusters of neurons are different, delete the edge even if the neurons involved in this connection become isolated after the disconnection, as in the next substep they will be correctly reconnected to the proper clusters, Equation 3.5.

d) Connect all clusters whose labels are the same and that were whether erroneously separated or became disconnected in the previous substep, Equation 3.6. The closest pair of neurons in these clusters are chosen to take part in this new connection.

Step 12: Early stopping: if the network have been training for at least η epochs and the current number of iterations so far is a multiple of ι (as with the previous disconnect and reconnect processes), the need to do the early stopping process is checked.

(a) Calculate the current value of the *monitor* metric with the validation set.

(b) Increase the number of waiting consultations made.

(c) If the current value of *monitor* has improved to such an extent that it is better than the one considered best so far (*i.e.* the difference between both values is not less than $delta_{min}$), update this best metric value and save the current network if necessary. Continue with the training process without doing the following steps of the early stopping procedure.

(d) If the number of consultations of the *monitor* metric that has been made is not less than *patience* (no relevant improvements of *monitor* have occurred in that period either), the training is marked to stop.

(e) If the early stopping needs to happen, finish training the network and restore the performance values and network state from that so-called "best epoch".

Step 13: Go to Step 2 if a stopping criterion (in this case *epochs*, although other possibilities exist, such as the network size ς, and certain performance metric) is not yet fulfilled.

3.4 Biological plausibility

One of the greatest challenges in the design and modeling of ANNs has been developing neural architectures, and their associated learning mechanisms, that best reflect and represent those found in existing biological systems [García Báez, 2005]. Most current ANN models or architectures attempt to reflect a certain biological flavor, but the incorporation of the main biological mechanisms responsible for the functioning of the biological neuron, and therefore of the biological neural network, is highly complex and difficult to find. In this sense, one of the most successful are the Adaptive Resonance Theory (ART) family of architectures [Carpenter and Grossberg, 2003] and Hybrid Unsupervised Modular Adaptive Neural Network (HUMANN) [García Báez, 2005; García Báez et al., 2010, 2012]. Among the biologically implausible characteristics frequently found in ANNs, the most common ones are simplifications, assumptions and limitations, which are introduced for better problem solving and greater computational performance, which bring along a reduced biological inspiration of those ANNs [da Silva and Rosa, 2013].

Structural and functional are the two main levels for defining a biological or artificial system, and can be used to express its biological plausibility [García Báez, 2005].

GNG, MyGNG and SupeRGNG are characterized by several bio-inspired properties, as stated by the names chosen for the functionalities in subsubsection 2.1.2.5 and subsection 3.3.1:

- Ontogeny, as these ANNs modify their interconnection strengths according to a predetermined learning rule, and adapt their topology to the problem [Fiesler and Beale, 1997].

- Connections aging and age resetting are similar to how telomerase works: cells can divide (and, hence, live) for a given period of time (*i.e.* they reach the Hayflick limit, which is approximately 50-70 cell divisions) after which they become senescent and no more cell divisions happen [Siegel, 2008]. However, thanks to the enzyme telomerase, the short bits of DNA known as telomeres that get shortened after each division are

restored, so the cell can again divide. Telomerase activation can produce cancerous growth, which imply that cells never die and they frequently divide. This can be replicated in GNG and SupeRGNG by choosing some combinations of parameters, mainly big values of a_{max} and low values of λ.

- Neural apoptosis, as neurons can die. In words of [Mazarakis et al., 1997], apoptosis "is a well conserved and highly regulated mechanism of cell death for the removal of unnecessary, surplus, aged or damaged cells". It happens in any kind of biological cells, not only neural ones. Although the final consequence is the same, the cell death, apoptosis should not be confused with "necrosis" because important differences exist [Bredesen, 1995].

- Neurogenesis, as new neurons can be created. As it occurs in mammals, this process happens not only during embryonic and perinatal stages (in the first iterations or epochs in our case) but also throughout life (during the network training in our case) [Ming and Song, 2011].

- Topographic maps, which are also present in the Self-Organizing Map (SOM) and, functionally, in the brain too [Kohonen, 2001], where similar patterns share near spacial locations.

- Positive tropism, as the network moves toward any available input sample, trying to cover all of them in order to better represent the input space. Among others, this behavior is akin to that of *Physarum polycephalum*, a slime mold that, while in the plasmodial stage, is able to move, by extending and developing pseudopods in a pulsatile and rhythmic way[8], towards attractants such as nutrients [Patino-Ramirez et al., 2019].

On the other hand, SupeRGNG has some specific bio-inspired characteristics:

- Boundary is defined in [Fagotto, 2014] as a "physical frontier that prevents mixing between two cell populations [and that] generally corresponds to a sharp and smooth delimitation". Boundaries are necessary for tissue separation. During boundary formation, cells typically rearrange, producing a straight boundary.

- Cell sorting refers to the process during which cells exchange neighbors in order to increase the number of homotypic contacts while, at the same time, decrease the quantity of heterotypic contacts [Fagotto, 2014]. Homotypic and heterotypic interactions occur between cells of the same and different type, respectively [Fagotto, 2014].

- Negative tropism, such as the one found in the myxomycete *Physarum polycephalum*, as this mold avoids moving toward repellents, that is, dangerous substances or stimulus such as salts and light (a.k.a. aphototropism), respectively [Patino-Ramirez et al., 2019].

- Shortest path problem, which have been solved by many organisms such as the *Physarum polycephalum*, with its ability to find the minimum-length solution between two and more points with nutrients in a labyrinth [Nakagaki et al., 2000]. In SupeRGNG the shortest path needs to be found when the supervised reconnection of neurons process occurs.

[8] *Physarum polycephalum* is able to move 1mm/s approximately, even up to few centimeters [Patino-Ramirez et al., 2019], and to change directions every 100 seconds.

- Autophagy, a.k.a. autophagocytosis, is a mechanism of the cell whereby defective or superfluous components are removed. The main difference with apoptosis is that the cell still exists after the autophagy mechanism because only some of the cell's non-useful organelles (parts of a cell) or unused proteins are eradicated [Mizushima and Komatsu, 2011]. In SupeRGNG a process akin to autophagy happens when unnecessary edges between neighboring neurons are deleted.

- Adherens junctions is a type of cell junction where cadherins, a type of cell adhesion molecule, from two proximal cells forms homophilic bindings between themselves when their conformation changes from flexible to rigid after Ca^{2+} ions get into the calcium-sensitive extracellular region. This results in cells of a similar type sticking together, which can further lead to multiple cells assembling into tissues. Such adhesives forces may even prevent tissues from dissociating into cells under external stress while easing cell movement [Harris and Tepass, 2010]. In SupeRGNG a similar process occurs when neurons with identical label connect if a connection between them does not exist.

3.5 Analysis and parametric study of the Supervised Reconfigurable Growing Neural Gas

In order to check both the stability and performance of the SupeRGNG in different types of problems, in the next subsections its behavior is described and analyzed by means of different datasets. Three synthetic datasets have been used to study the capabilities of the SupeRGNG. Hence, the purpose of each of them differs and is related to its main characteristic, which will be used to title the next subsections. All but the first synthetic dataset, which is binary, are multiclass.

The behavior of the SupeRGNG with these datasets has been studied according to the values of the most important hyperparameters of the network, analyzing its influence in the final classification results.

3.5.1 Datasets for the study of the SupeRGNG

In order to study the behavior and adaptation capability to different data of SupeRGNG several synthetic data sets with dissimilar characteristics have been used, which will be described below.

3.5.1.1 Non-Gaussian and non-linearly separable data

This dataset is made up of 1000 samples not generated with a Gaussian function and that belong to two different classes, shaped as concentric squares that cannot be linearly separated, so it is also known as "2_concentric_squares", Figure 3.7. Its Gaussian version is called "bullseye". The internal region is a square whose side measures 0.3, whereas the external one is a band that is built by subtracting from a square with a side of 1.3 a concentric square with a side of 1. As both classes have almost identical number of samples, the internal square is much denser.

This dataset was used in order to check the capability of the SupeRGNG with non-Gaussian data that cannot be separated with a linear hyperplane, where classes also have

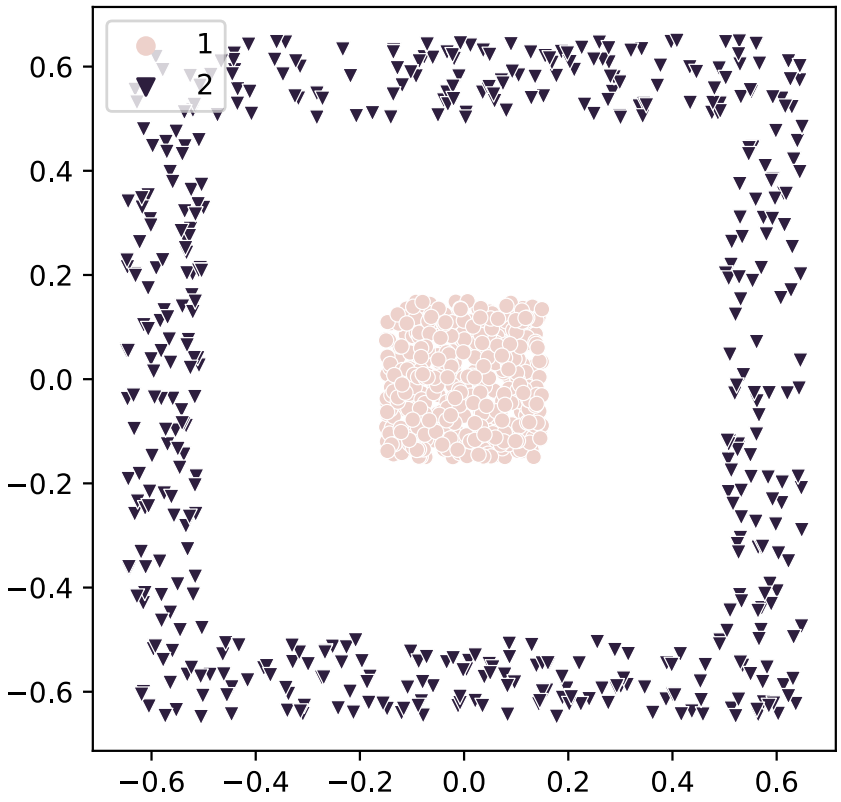


Figure 3.7: Non-Gaussian and non-linearly separable data, a.k.a. "pgb_squares".

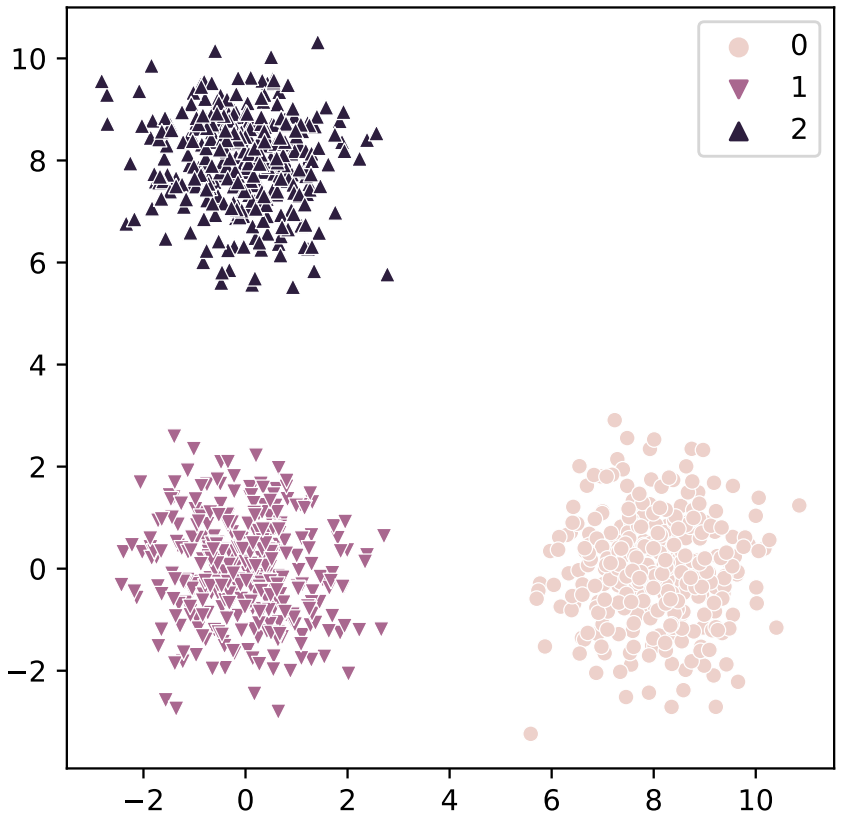
different density and shape. These characteristics have proven to be problematic for certain classification and clustering models, which make assumptions about some of these characteristics, especially the last two, in the classes to be found. It has been used for a similar purpose with HUMANN in [García Báez, 2005], and this author indicated that it was based on a problem found at a previous edition of [Koutroumbas and Theodoridis, 2008][9].

3.5.1.2 Imbalanced data

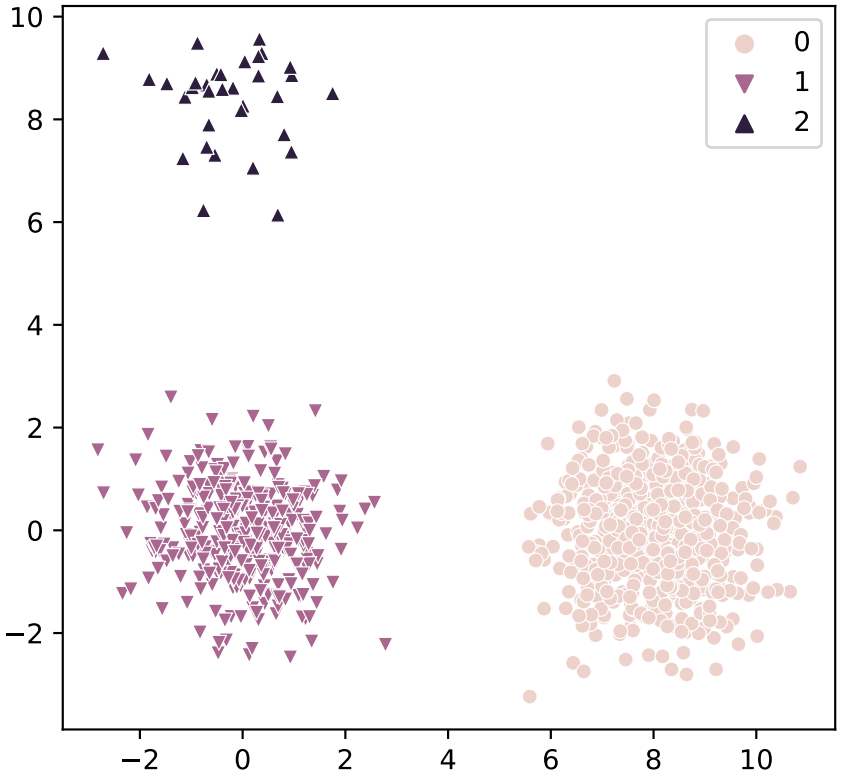
This dataset consists of three isotropic Gaussian 2D blobs (*i.e.* round clouds) of samples, each one belonging to a different class, and ensuring that they are linearly separable. The "make_blobs" function found at scikit-learn was used to create the blobs [Pedregosa et al., 2011].

Unlike the second synthetic dataset used in [García Báez, 2005], where both the shape and the density of the patterns change, only the latter varies in ours and this occurred for two of the classes instead of just one. Also, each blob is isotropic and not elliptical. Conversely, in both cases the quantity of classes, the number of dimensions and centers of the blobs are the same, Table 3.2. Similarly occurred with the density function, Equation 3.8, where i is the dimension, 1 or 2, which corresponds to the X and Y axis; μ_i, the mean of the class in the i-th dimension; and σ_i, the standard deviation of the class in the i-th dimension.

[9] In that reference, which is the fourth edition, it is the problem 12.5.



a) Class ratio ±0%.



b) Class ratio ±90%.

Figure 3.8: Imbalanced data, also named “unbalanced_blobs”. Extreme cases.

$$f(x) = \frac{1}{\sqrt{2\pi\sigma_i^2}} \; e^{-\frac{(x_i-\mu_i)^2}{2\sigma_i^2}} \tag{3.8}$$

Table 3.2: Values of the parameters that define the classes in the imbalanced dataset, showing the six different variants.

Class	μ_1	μ_2	σ_1	σ_2	Number of samples
0	8	0	1	1	{ 333, 499, 532, 566, 599, 632 }
1	0	0	1	1	333
2	0	8	1	1	{ 333, 167, 134, 100, 67, 34 }

A total of six variants of this dataset have been analyzed, which were created by varying the class ratio in each of the variants (*i.e.* the number of samples of two of the blobs are different, one increasing in the same percentage as the other decreases), Table 3.2: $\pm 0\%, \pm 50\%, \pm 60\%, \pm 70\%, \pm 80\%, \pm 90\%$. The plots of the extreme cases can be seen in Figure 3.8. Therefore, the purpose of this dataset is to analyze how well the SupeRGNG works with different levels of class ratios.

3.5.1.3 Overlapping data

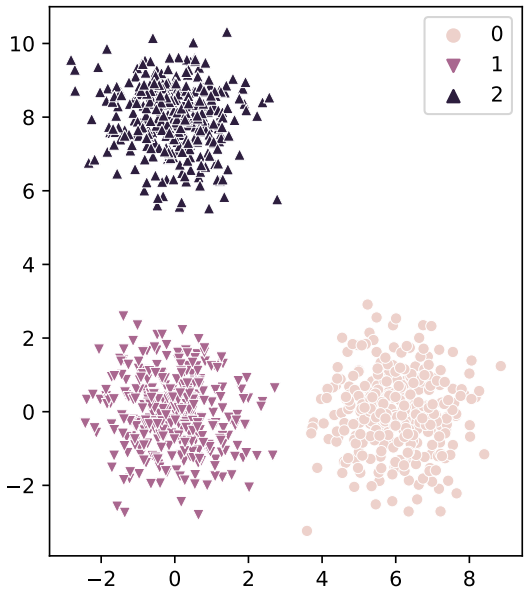
This dataset comprises three isotropic Gaussian 2D blobs of samples, each one belonging to a different class and with the same number of samples per class. The "make_blobs" function found at scikit-learn was used to create the blobs [Pedregosa et al., 2011]. The samples were generated with the density function shown in Equation 3.8, where i is the dimension, 1 or 2, which corresponds to the X and Y axis; μ_i, the mean of the class in the i-th dimension; and σ_i, the standard deviation of the class in the i-th dimension.

As inferred from Table 3.3, a total of six variants of this dataset have been analyzed, which were created by varying the value of the mean of one of the blobs so that the distance also changes and, consequently, the overlapping coefficient (OVL) of the samples of two of the three classes too. The six overlapping percentages[10] chosen to be studied are: 0%, 10%, 21%, 45%, 62%, 80%. The plots of the extreme cases and their areas of the Gaussian curves intersections can be seen in Figure 3.9. Henceforth, the purpose of this dataset is to analyze the performance of the SupeRGNG with dissimilar levels of class overlapping.

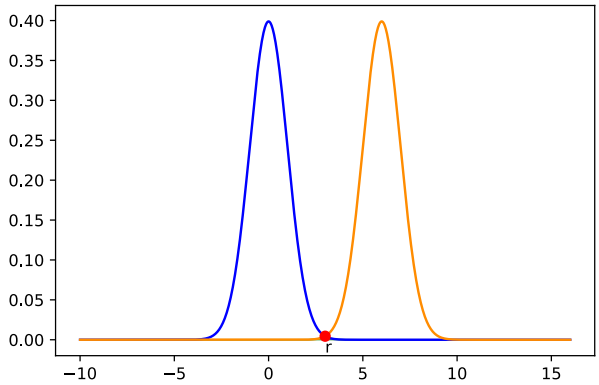
Table 3.3: Values of the parameters that define the classes in the overlapping dataset, showing the six different variants.

Class	μ_1	μ_2	σ_1	σ_2	Number of samples
0	{ 6, 3.25, 2.5, 1.5, 1, 0.5 }	0	1	1	333
1	0	0	1	1	333
2	0	8	1	1	333

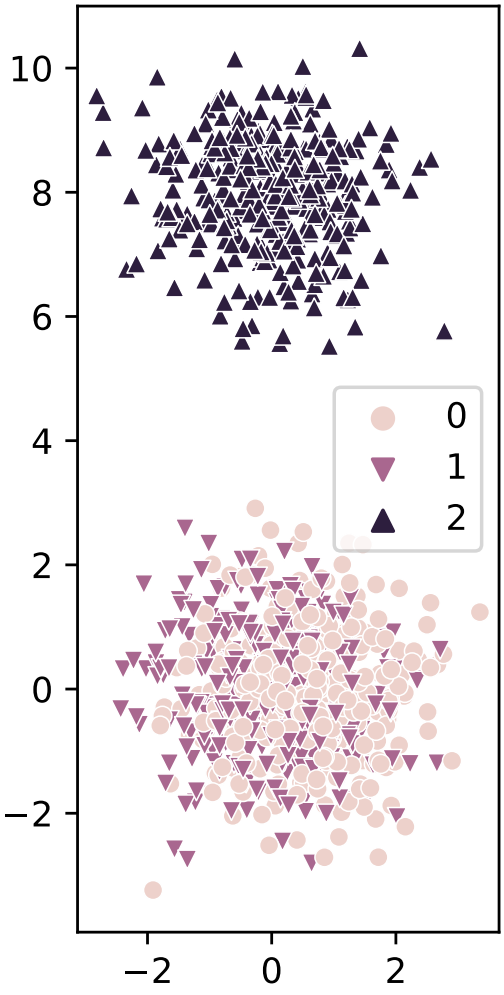
[10]This OVL — or area between the two normal distributions that refers to each blob that overlaps — was calculated with the "statistics" Python module, which references the methods presented in [Inman and Bradley Jr, 1989].



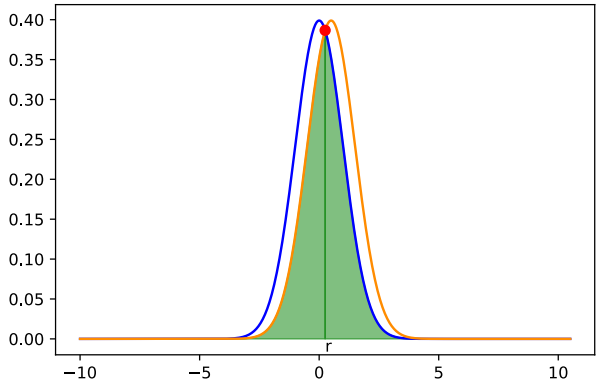
a) Overlapping 0%, maximum distance.



b) Area of the Gaussian curves intersection when overlapping is 0%.



c) Overlapping 80%, minimum distance.



d) Area of the Gaussian curves intersection when overlapping is 80%.

Figure 3.9: Overlapping data, also named "blobs". Extreme cases and their areas of the intersections of the Gaussian curves.

3.5.2 Analysis of the hyperparameters of the SupeRGNG

The hyperparameters of the SupeRGNG includes those of the GNG and two related to the disconnection and reconnection mechanisms. They can be grouped according to its functionality: neuron movement ($\varepsilon_b, \varepsilon_n$), error decreasing ($\beta, d$), network growing ($\lambda, a_{max}$), and reconfigurability (η, ι).

Using the synthetic datasets described in the previous subsection, the most relevant hyperparameters of the SupeRGNG are studied, which were identified by means of both initial experiments or searching the literature for similar studies on the GNG, Appendix B. From this search we also obtained some recommended values, which were considered whether as initial (for relevant hyperparameters) or as fixed values (for lesser ones, Table 3.4). A value of 1000 was selected for *epochs* because it was considered high enough for any network to grow and learn taking into account the expected values of the other hyperparameters and the size of the input data. For those related to early stopping, the metric being monitored was AUC, a value of 0 was given to $delta_{min}$ (which means that any beneficial change is considered an improvement), and a *patience* of 10 was considered adequate.

Table 3.4: Fixed values given to the hyperparameters of the SupeRGNG.

Hyperparameter	β	d	ε_c	ε_n	*epochs*
Fixed value	0.5	0.99	0.2	0.06	1000

From these initial values for the relevant hyperparameters, the search for optimal values was carried out by launching several tests with different sets or ranges of values. As each synthetic dataset has different characteristics, these tests were not necessarily the same in all cases, although, for convenience, the initial ones were identical. Conversely, some hyperparameters had fixed values in every test, Table 3.4, which were mostly based on other GNG-related researchers' works, Appendix B.

All these tests were launched on DigitalOcean Paperspace, a High-Performance Computing (HPC) cloud platform that can be used both as a Machine Learning (ML) environment or to build virtual servers with different operating systems (OSes), disk spaces, Central Processing Units (CPUs), Graphics Processing Units (GPUs) and multiGPUs in several geographical locations[11].

Filled contour maps are 2D visualization methods that were used to compare the behavior of the hyperparameters. They provide the same information as a 3D surface plot but the do nor require the third dimension nor modify elevation and azimuth angles to avoid hiding parts of the graph. Similar to contour lines, they both represent points with the same value, whether with a line (contour lines) or with the same color (filled contour maps). In geography they are used to represent points of equal elevation with the same color.

Before we start the analysis, the concept of "Mix" should be defined. When the analysis is not related to changes between different versions of the dataset, such as the class overlapping and the class ratios, "Mix" means a set of configurations where all their hyperparameters but two have fixed values. Otherwise, all hyperparameter but one (the other variable to analyze is the class overlapping or the class ratios) are fixed.

Two types of metrics were collated from the tests conducted: one for performance (AUC, how well it classifies) and four for dynamics (the ratio of created neurons, which means how

[11]More up-to-date information on the characteristics of DigitalOcean Paperspace can be found at `https://docs.digitalocean.com/products/paperspace/`.

Table 3.5: Set of optimal configurations used in the non-Gaussian and non-linearly separable dataset. The number of configurations in each of these tests was 750 in both.

Hyperparameter	Values
λ	[10, 20, 30, 40, 50]
a_{max}	[10, 20, 30, 40, 50]
max_nodes	[150, 450, 750]
η	[10, 20, 30, 40, 50] [100, 200, 300, 400, 500]
ι	[400, 800]

fast it grows; the ratio of deleted neurons, how fast it shrinks; the ratio of deleted edges in a supervised way, the frequency of the disconnection process; and the ratio of reconnections, the frequency of the reconnection process).

The hyperparameters that were considered relevant and, therefore, their influence in the classification performance and behavior of the SupeRGNG was studied, were: ι, max_nodes, λ, a_{max}, and η.

3.5.2.1 Hyperparameter ι

The first hyperparameter that was studied was ι for two reasons: it is unique to the SupeRGNG, and its impact in the performance of the network was estimated to be high because, together with η, they control de frequency of the disconnection and reconnection mechanisms.

Two values of ι were studied, 400 and 800. The latter was chosen based on the size of the training data: the total number of samples was 999 or 1000 in the synthetic datasets. It must have this value for the SupeRGNG to be able to apply the disconnection and reconnection mechanism just after all the training samples have been used by the network, *i.e.* precisely at the end of an epoch.

The values shown in Table 3.5 refers to the first two sets of values that were tested in the three synthetic datasets, and configurations belonging to these two sets will be used across page 97. It was found that the 1500 configurations achieved a value of AUC of 1 with both the non-Gaussian and non-linearly separable dataset, and with the imbalanced dataset. The AUC values of a subset of these configurations is depicted in Figure 3.10, where $\lambda = 10, a_{max} = 10, max\_nodes = 150$.

From these two tests, apart from finding two sets of perfect configurations for both the non-Gaussian and non-linearly separable dataset and all the different class ratios in the imbalanced dataset, it was concluded that, in terms of performance, $\iota = 400$ does not provide an advantage in those datasets, so in following tests only $\iota = 800$ was used. However, the best configuration with 80% overlapping used $\iota = 400$ and belonged to the second set: $\lambda = 50, a_{max} = 10, max\_nodes = 750, \eta = 400$. So, for such type of datasets, 400 might be better.

3.5.2.2 Hyperparameter max_nodes

Three values of max_nodes were studied in all the tests used in the three synthetic datasets, which can be labeled as “low”, “medium” and “high” considering the size of the training data: 150, 450 and 750.

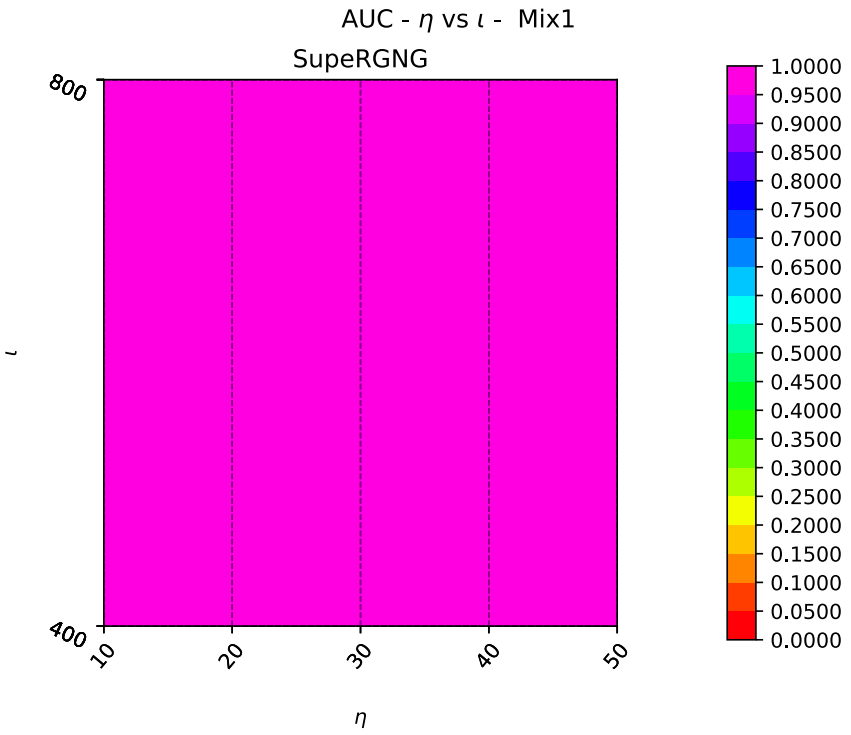


Figure 3.10: Example of several configurations with AUC=1 with the non-Gaussian and non-linearly separable dataset.

Although with the non-Gaussian and non-linearly separable dataset no big differences in performance or behavior were found, with the class overlapping dataset it was clearly the opposite. In fact, the greater the class overlap, the greater the negative impact of using smaller network sizes, Figure 3.11. Figure 3.12 also shows that bigger networks are better with maximum ovelapping.

Additionally, a prospective study with a value way higher than the size of the training dataset was studied with the 80% OVL data. It does not only confirmed that a value of *max_nodes* close to the number of training samples is beneficial for datasets with heavily overlapped classes but also showed that using more neurons than the size of the training data, and even bigger than the input data, improved the classification performance close to 0.91 AUC, Figure 3.13.

3.5.2.3 Hyperparameter λ

Hyperparameter λ governs the network growing rate, how frequent a new neuron is created. As in the GNG, λ heavily affects the topology of the network, several sets of values were tested. A longer range of values for λ were used in Table 3.6. Larger values of $\lambda > 1700$ damaged the performance of the network, Figure 3.14.

3.5.2.4 Hyperparameter a_{max}

Hyperparameter a_{max} defines the apoptosis of connections and, as a byproduct, of neurons too. Hyperparameters a_max and λ were now given larger values and ranges too, Table 3.7.

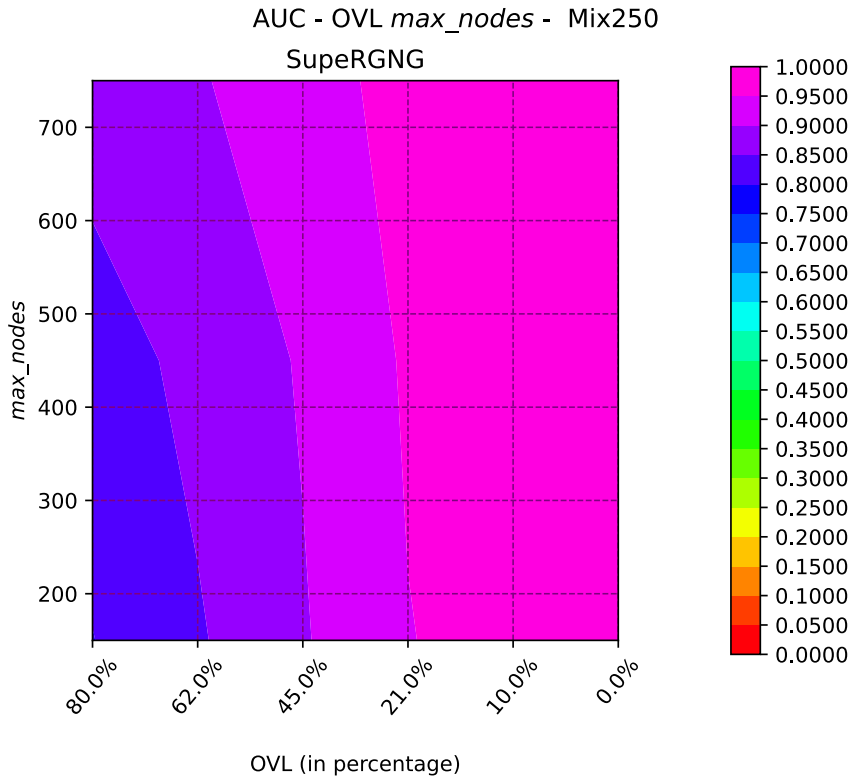


Figure 3.11: Bigger networks are better with overlapping 80%.

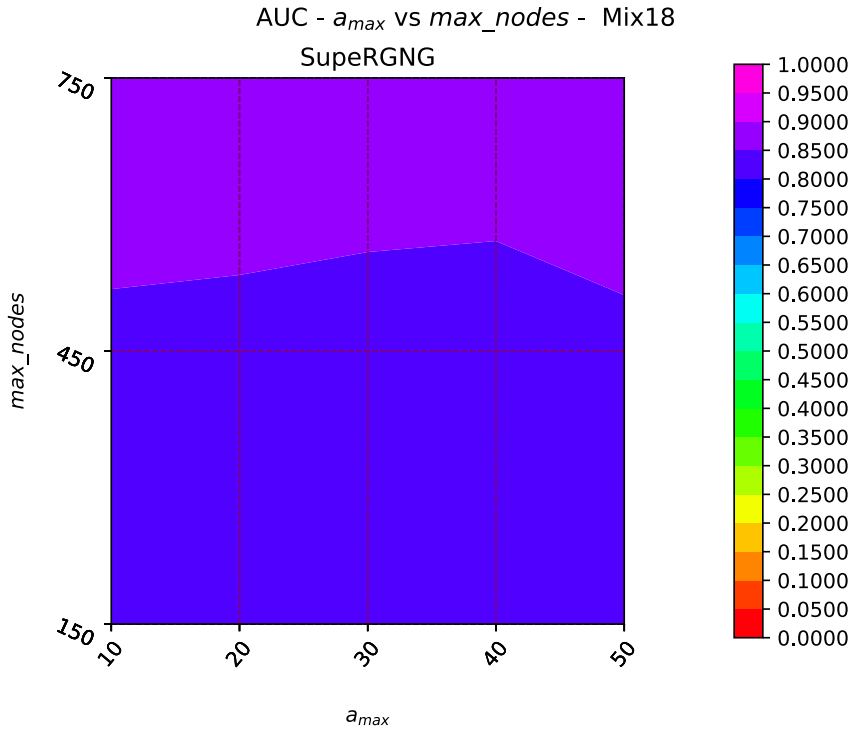


Figure 3.12: Bigger networks are better with overlapping 80%.

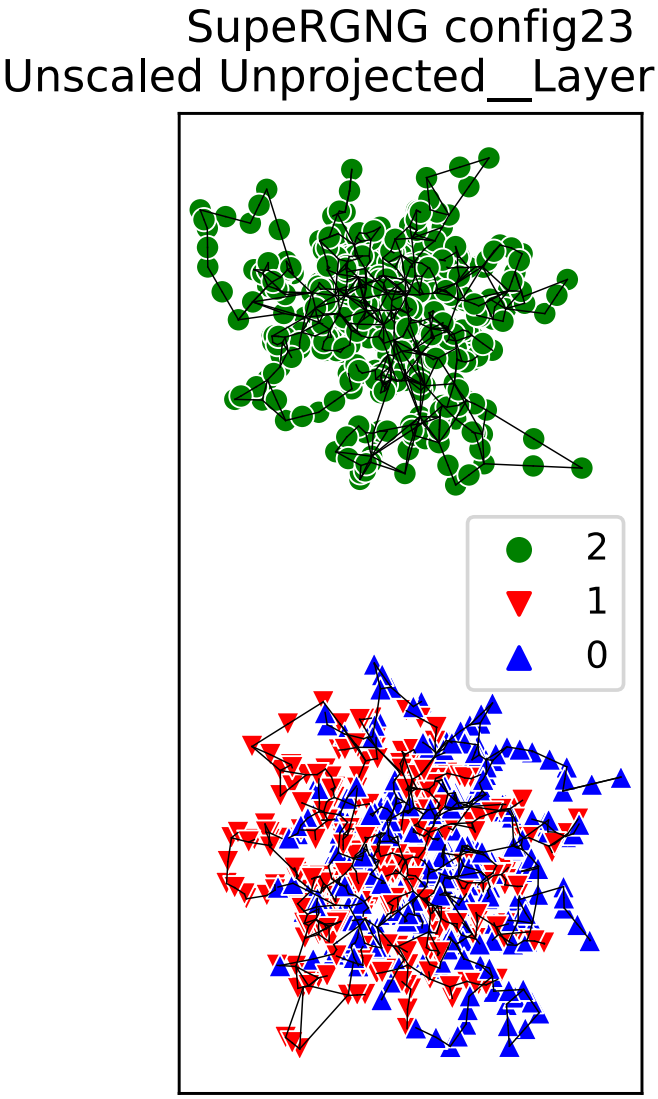


Figure 3.13: Example of a SupeRGNG network with more nodes than the training data (80% overlapping).

Table 3.6: Configurations used in the non-Gaussian and non-linearly separable dataset. The number of configurations in each of these tests was 750 in both.

Hyperparameter	Values
λ	[100, 300, 500, 700, 900, 1100, 1300, 1500, 1700, 1900]
a_{max}	[10, 20, 30, 40, 50]
max_nodes	[150, 450, 750]
η	[10, 20, 30, 40, 50]
ι	[800]

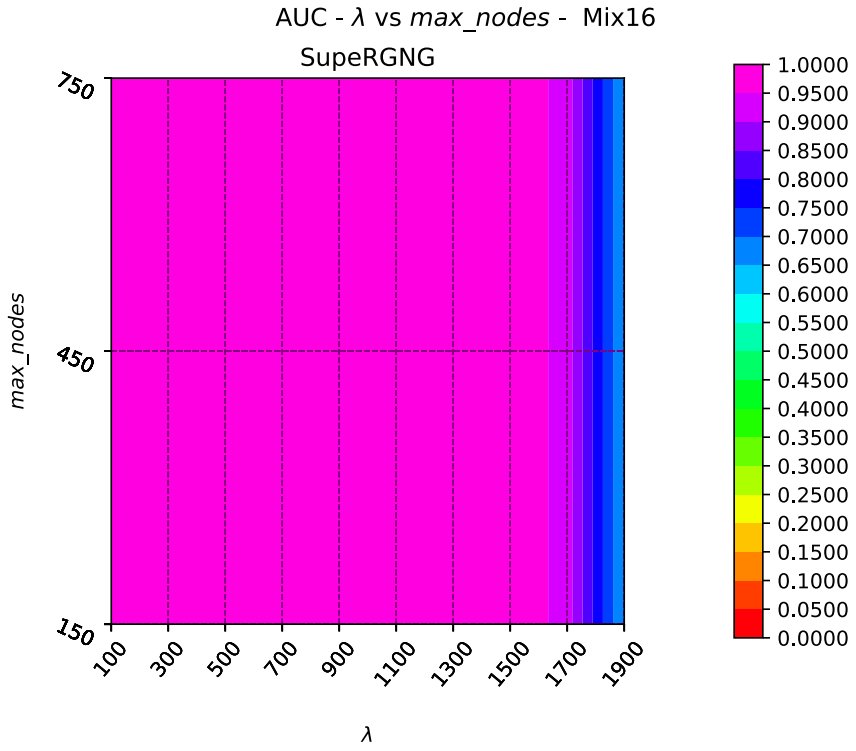


Figure 3.14: High λ might be suboptimal with the non-Gaussian and non-linearly separable dataset.

Table 3.7: Configurations used in the non-Gaussian and non-linearly separable dataset. The number of configurations in each of these tests were 1500 and 1500.

Hyperparameter	Values
λ	[2100, 2300, 2500, 2700, 2900, 3100, 3300, 3500, 3700, 3900]
a_{max}	[2100, 2300, 2500, 2700, 2900, 3100, 3300, 3500, 3700, 3900]
max_nodes	[150, 450, 750]
η	[10, 20, 30, 40, 50] [100, 200, 300, 400, 500]
ι	[800]

In Figure 3.15 it is depicted that combining a low $\eta = 10$ with any high value of a_{max} studied in Table 3.7 gives quite low AUC values, around 0.40.

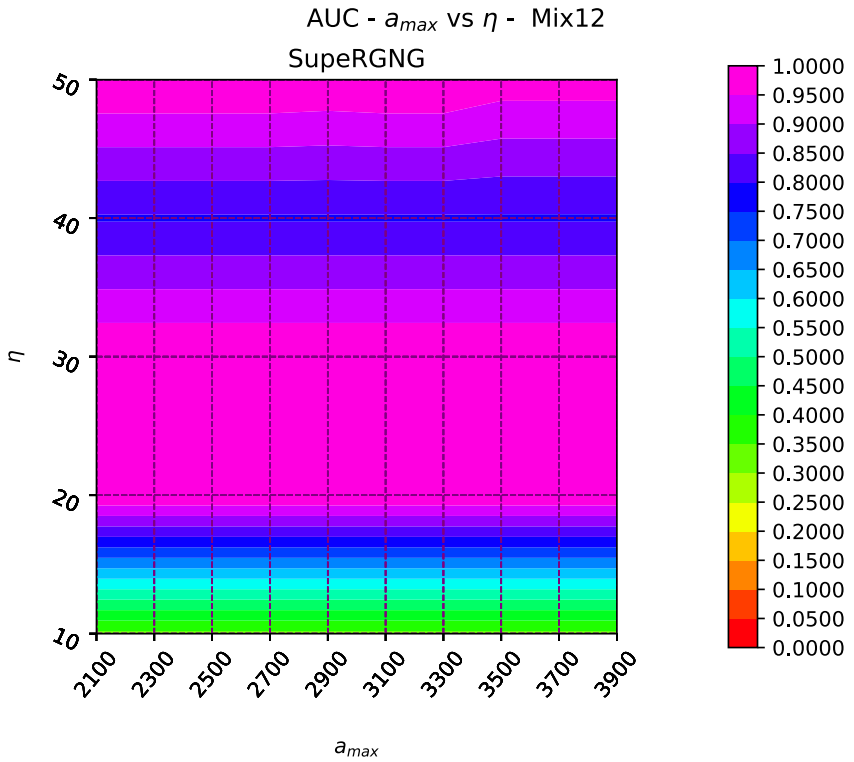


Figure 3.15: Low η might be suboptimal with the non-Gaussian and non-linearly separable dataset.

In Figure 3.16 and related to the imbalanced dataset, it is seen that a_{max} affects the ratio of reconnections in the expected way: the greater it is, the less reconnections are needed to be made as the connections were more durable.

3.5.2.5 Hyperparameter η

Hyperparameter η controls the starting point of the disconnection and reconnection mechanisms. Values from the second set of configurations in Table 3.7 demonstrated that medium values of $\eta = \{100, 200, 300\}$ negatively impacted performance, Figure 3.17.

When conducting a study on the effects of overlapping in the hyperparameters, the most common behavior looks like Figure 3.18: performance lowers if overlapping grows.

3.5.2.6 Hyperparameter η combined with others

During previous tests it was noticed that the behavior of the SupeRGNG is impacted by λ, a_{max} and η independently, so it was decided to study them combined.

With the non-Gaussian dataset, a set of configurations showed low performance when a too low value of $\eta = 10$ together with a very big $\lambda \in \{1700, 1900\}$ were used Figure 3.19. This behavior repeats for the three values of *max_nodes* that were used – 150, 450, 750 – when $a_{max} \in \{30, 40\}$. Therefore, such ranges of values for these hyperparameters are discouraged.

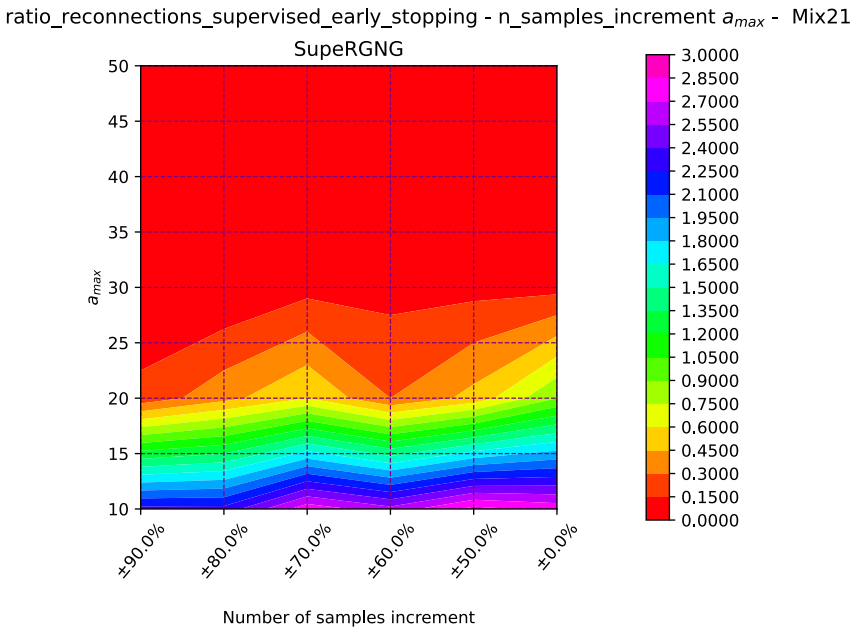


Figure 3.16: Ratio of reconnections changes with a_{max} in the imbalanced dataset.

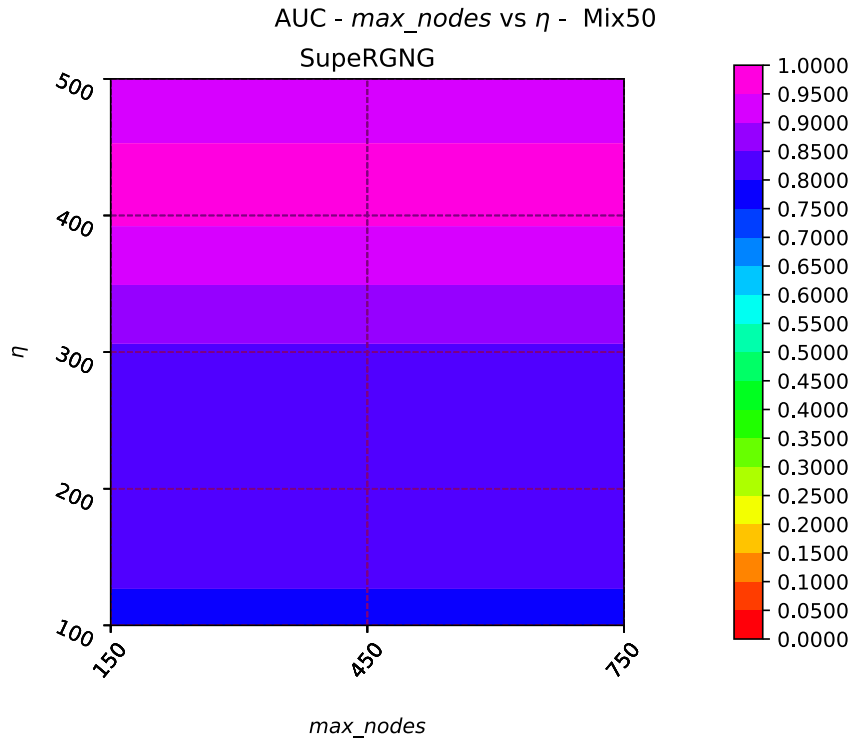


Figure 3.17: Medium η might be suboptimal with the non-Gaussian and non-linearly separable dataset.

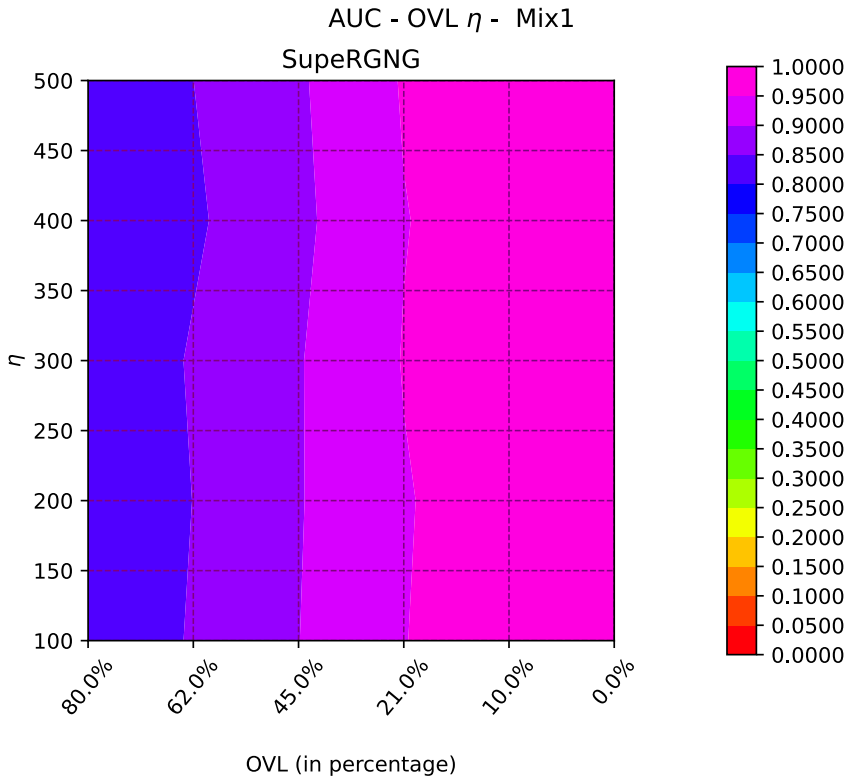


Figure 3.18: Low η might be suboptimal with the non-Gaussian and non-linearly separable dataset.

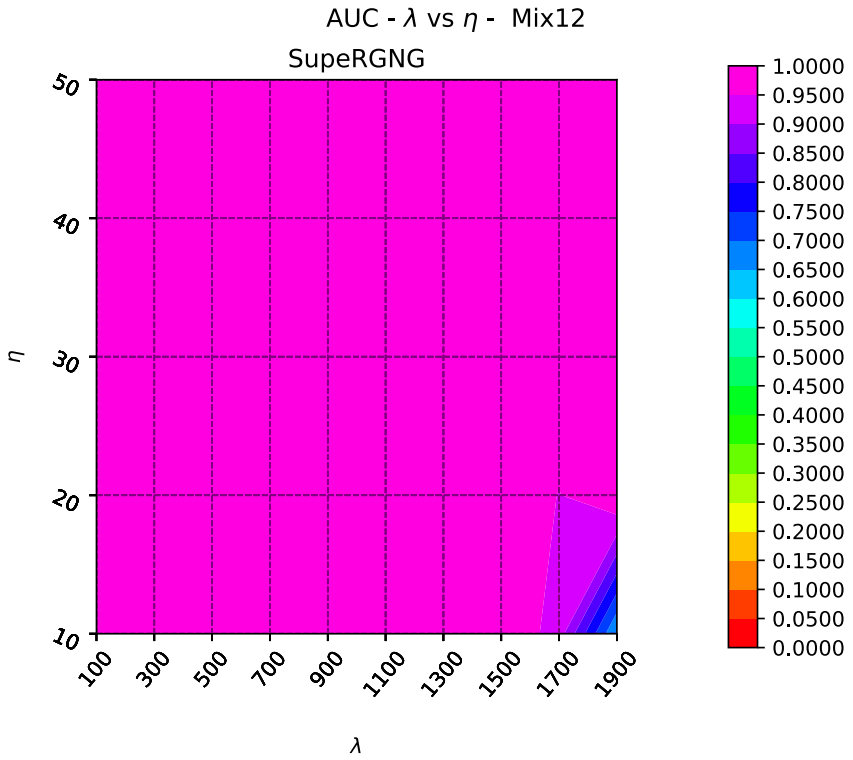


Figure 3.19: Low η might be suboptimal with the non-Gaussian and non-linearly separable dataset.

Table 3.8: Recommended values for the hyperparameters of the SupeRGNG considering the characteristics of the dataset used.

Characteristic of the dataset	Value of the characteristic	Recommendations	Examples
Non-Gaussian and non-linear separability		Do not combine small η AND (high λ OR high a_{max})	$\eta < 30$ AND ($\lambda >= 1000$ OR $a_{max} >= 1000$)
		Low-Medium values of η are discouraged	10-300
Class imbalance	Low High		
Overlapping	Low		
	High	High max_nodes	For training sets of size 800, 750-790. 1100 was even better

3.5.2.7 Summary of the influence of the hyperparameters

A summary has been built after the analyses carried out in the previous subsubsections, where adequate values for the hyperparameters of the SupeRGNG according to the characteristics of the dataset used are exposed, Table 3.8.

It has been shown that the SupeRGNG performs extremely good with non-Gaussian and non-linearly separable datasets and is extraordinarily robust to class imbalance (up to 90%), achieving perfect values of AUC. Furthermore, the SupeRGNG is very good working with different levels of overlapping classes, being able to tackle up to 80% yet yielding 0.8986 AUC.

Chapter 4

Applications of the Modular Hybrid Growing Neural Gas and the Supervised Reconfigurable Growing Neural Gas in the detection of Alzheimer's Disease and Mild Cognitive Impairment

In this chapter, on the one hand, applications of the Modular Hybrid Growing Neural Gas (MyGNG) with real data, unlike with the Supervised Reconfigurable Growing Neural Gas (SupeRGNG), were limited to one binary classification problem and a multiclass one: distinguishing the two levels of neurodegeneration (MCI vs AD) and the possible three levels in which usually the Alzheimer's Disease (AD) continuum has been summarized (CN vs MCI vs AD).

On the other hand, we will describe applications of the SupeRGNG in a real environment (*i.e.* with non-synthetic datasets): in the detection of a family of degenerative nerve diseases such as AD and Mild Cognitive Impairment (MCI). As this is a wide task, it has been divided into several classification problems, as seen in the following sections. Both binary and multiclass tasks were tackled, which, combined, cover all the AD continuum, from Cognitively Normal (CN) subjects that do not have the disease (but probably present normal levels of deterioration due to aging), to patients with increasing cognitive impairment, which can be labeled whether as MCI or AD.

Furthermore, the chosen classification problems have been widely used by other researchers [Cabrera-León et al., 2024a], so qualitative comparisons with those works were carried out. Similarly, quantitative comparisons of both systems with other Machine Learning (ML) classifiers have been conducted.

In each of these applications the intelligent system comprised of several stages, some of which could be divided into steps. These stages follow the typical hybrid pipeline, where non-neural methods are used for preprocessing whereas the processing is carried out by a neural model, as it can be seen in Figure 4.1:

1. Input data, where we describe the input data that were used: original database and

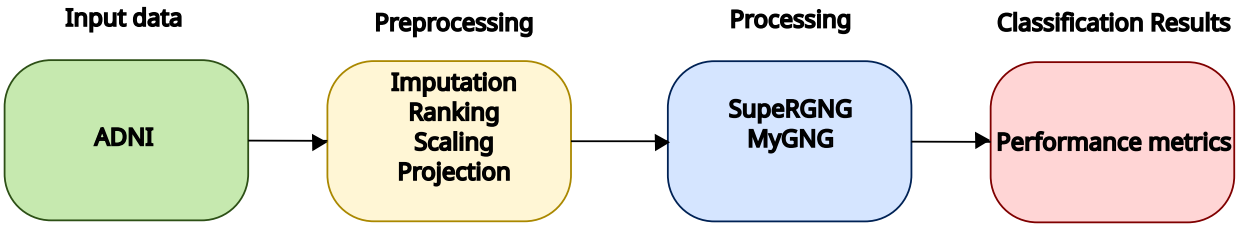


Figure 4.1: General diagram with the stages followed by the intelligent system in each application.

statistical analysis of the features.

2. Preprocessing, which includes the imputation feature ranking, feature scaling, and data projection that were applied.

3. Processing, where different configurations of the neural models are trained, in this case the SupeRGNG and the MyGNG, in order to find the most appropriate combinations of hyperparameters for each application. It includes the data partitioning techniques used.

4. Results, where the classification performance is analyzed according to the metrics used and discussed. Also, comparisons with other methodologies are exposed.

It should be noted that a common methodology was followed in all the applications. It involved analyzing and finding the optimal combinations of scaling techniques (including the unscaled case), projection methods (including the unprojected case), number of projected components, and values of the hyperparameters of the networks implemented. The combinations were considered optimal when all their values of performance metric were better, prioritizing the Area Under the Curve (AUC) value for its robustness to imbalanced classes but only if it ensured that the other performance metrics were also good, followed by the other types of model evaluation metrics, section 2.5.

In each of the following classification problems the p-values of the features used were calculated, taking into account the number of classes and other characteristics of the variables to choose the most appropriate statistical test, Figure C.1. The next significance levels (a.k.a. α) were selected[1]: 0.0001, 0.001, 0.01, 0.05, 1. They corresponded, respectively, to the next symbols used in several tables found in this chapter: "****", "***", "**", "*", "ns" (not significant). These α cutoff points were chosen for being the most popular in the majority of fields, including the medical one where this work belong to. However, smaller values are preferred in other fields: high-energy physics need p-values lower than 0.003 or even 0.0000003[2] [Gale et al., 2016].

Before expounding each of these applications the common information environment is described. At the end of this chapter, a quantitative comparative study of our systems with some popular ML and Deep Learning (DL) methods and a qualitative comparative study with works by other authors are included, in both cases separated by classification task.

[1]As a reminder, if a p-value is less than α, it can be concluded that there is a statistically significant difference between the two groups. Otherwise, it cannot be concluded that.

[2]A p-value of 0.0000003 equates to about 1 in 3.5 million or saying that the strength of the data supporting what is being checked is 5-sigma [Gale et al., 2016]. The latter value was used during the Higgs boson discovery.

4.1 Information environment

Unlike in previous decades, where AD researchers required to build their own information environment, nowadays there are many that can be accessed whether publicly or after asking for an authorization or accepting their owner's data usage agreement, Table 4.1. Using non-private databases allow researchers other than their owners to access, generally multimodal, data from multiple subjects possibly from different clinical sites [Cabrera-León et al., 2024a].

All the non-synthetic data used in this PhD thesis were collected from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database[3] [University of Southern California, 2004]. Led by Principal Investigator Michael W. Weiner, MD, the ADNI started in 2003 as a public-private partnership, which provides access to a huge database with many medical tests collected over long periods of time from different patients [Alzheimer's Disease Neuroimaging Initiative (ADNI), 2013]. The main objective of ADNI has been to test whether Magnetic Resonance Imaging (MRI), Positron Emission Tomography (PET), other biological markers, and clinical and neuropsychological tests can be combined to measure the progression of MCI and early AD [Cabrera-León et al., 2023]. Also, ADNI has been analyzing AD-related omics and imaging [Yao et al., 2017]. That is, data from different modalities have been included in this database. ADNI is a multi-site initiative, with over 50 clinical sites in the USA and Canada.

ADNI comprises several studies carried out through the last two decades: ADNI1, ADNIGO[4], ADNI2 and ADNI3. In each of these studies different clinical criteria and techniques have been used, discarding some of them while incorporating new ones, especially some related with neuroimaging and genes, as knowledge about AD increases. On November 15, 2022 the ADNI4 study got into the recruitment phase and, since July 24, 2025 its status changed to the active and not recruiting phase[5]. Its estimated completion date will be July 31, 2027. That is, ADNI4 data were not still available when the preliminary literature review and the data acquisition stage were performed. For the ADNI4 study they are expecting to enroll 750 new participants, while 750 will be rollover subjects from previous ADNI studies. These participants will be studied for up to 5 years, and will be enrolled across three cohorts: dementia, MCI and CN. These cohorts are not identical to those used in the previous ADNI studies, which were more focused on predementia and MCI severity levels, including a self-perceived state: CN, Significant Memory Concern (SMC), Early Mild Cognitive Impairment (EMCI), MCI, Late Mild Cognitive Impairment (LMCI), and AD. In Table 4.2 the number of participants for each of the previous cohorts and ADNI studies whose data are currently available is shown, together with their starting date and duration.

A huge file combining multiple diagnostic labels, demographic data and multimodal clinical criteria from several longitudinal visits was built by merging different files from ADNI. This process was akin to the "ADNIMERGE" file already available but with much more clinical criteria. Although in our file data from multiple time visits were collated, which allows longitudinal studies to be performed, only data from the baseline were used in these experiments. After merging the files, the total number of diagnostic criteria was 202.

[3]More up-to-date information can be found at `www.adni-info.org`.

[4]GO stands for "Grand Opportunities".

[5]For the most up-to-date status see `https://classic.clinicaltrials.gov/ct2/show/NCT05617014`.

Table 4.1: Non-private databases with Alzheimer's Disease, Mild Cognitive Impairment and Cognitively Normal subjects.

Database	Access	Number of subjects	Types of biomarkers	Link
ADNI	Auth	1833	D, NT, NI, G, BS, M/S.	http://adni.loni.usc.edu/
ADReSS challenge	Auth	156	D, NT, M/S.	https://dementia.talkbank.org/
AIBL	Auth	1000+	D, NT, NI, BS.	https://aibl.csiro.au/
AlzBiomarker	Auth	N/A	BS.	https://www.alzforum.org/alzbiomarker
CADDementia	Auth	354	D, NI.	https://caddementia.grand-challenge.org/
DementiaBank & Pitt corpus	Auth	473	D, NT, M/S.	https://dementia.talkbank.org/
GEO	Public	N/A	G.	https://www.ncbi.nlm.nih.gov/geo/
Kaggle: Alzheimer's Dataset (4 class of Images)	Public	6400	NI.	https://www.kaggle.com/datasets/tourist55/alzheimers-dataset-4-class-of-images/data
OASIS	Auth	1098	D, NT, NI.	https://www.oasis-brains.org
Rdatasets: apoeapoc	Public	353	D, G.	https://vincentarelbundock.github.io/Rdatasets/
Rdatasets: nep499	Public	499	D, G.	https://vincentarelbundock.github.io/Rdatasets/
MAS	Auth	1037	D, NT, NI, G, BS, M/S.	https://cheba.unsw.edu.au/research-projects/sydney-memory-and-ageing-study
TADPOLE	Auth	1000	D, NT, NI, G, BS, M/S.	https://tadpole.grand-challenge.org/

Abbreviations: Auth = Authorization; BS = Biospecimen or biosamples; D = Demographic; G = Genetic; M/S = Human gait, movements or speech; N/A = Not Available; NT = Neuropsychological tests; NI = Neuroimaging

Table 4.2: Starting date, duration and number of subjects per class in each of the currently available ADNI studies.

Cohort	ADNI1 10/2004 (5 years)	ADNI GO 09/2009 (2 years)	ADNI2 09/2011 (5 years)	ADNI3 09/2016 (5 years)
CN	200	500[1]	150 + 450-500[2]	135-500 + 295-330[3]
SMC			100	
EMCI		200	150 + 200[4]	
MCI	400			150-515 + 275-320[5]
LMCI		500	150	
AD	200 mild		200 mild	85-185 + 130-150[6]

[1] "500 LMCI and Cognitively Normal subjects from ADNI1". This quantity is shared in the Table with the one indicated for LMCI.
[2] "150 Normal Controls (new), and 450-500 CN and MCI (rollover from ADNI1; approximate)"
[3] "135-500 Normal Controls (new), and 295-330 Normal Controls (rollover from ADNI2; approximate)"
[4] "150 EMCI (new) and 200 EMCI (rollover from ADNI GO; approximate)".
[5] "150-515 Mild Cognitive Impairment (MCI) (new), and 275-320 Mild Cognitive Impairment (MCI) (rollover from ADNI2; approximate)".
[6] "85-185 Mild Alzheimers Disease dementia (AD) (new), and 130-150 Mild Alzheimers Disease dementia (AD) (rollover from ADNI2; approximate)".

4.2 CN-AD

Among all the classification problems that we have worked with, CN-AD can be considered, by far, the easiest one due to the fact that the characteristics of both kind of subjects are much more different.

4.2.1 Input data and preprocessing

In this and the other classification tasks the same methodology for choosing the data imputation and the feature ranking methods was followed.

Two approaches were followed to check how having missing values in the samples changed the features being selected by the feature ranking methods: using the original non-imputed data, or imputing data with the median value per class. The latter was preferred.

The large number of features in the original dataset, 202, was the main reason for applying a feature ranking or selection method. For each of the three ADNI subsets[6] analyzed - ADNI1, ADNI2 and ADNI3 - two very different techniques to rank the features were compared: Fast Correlation-Based Filter (FCBF) [Yu and Liu, 2003] and Extreme

[6] ADNIGO was discarded as it only added EMCI subjects, the CN and AD participants were rollover from ADNI1. At the time of data acquisition, ADNI4 data were not available.

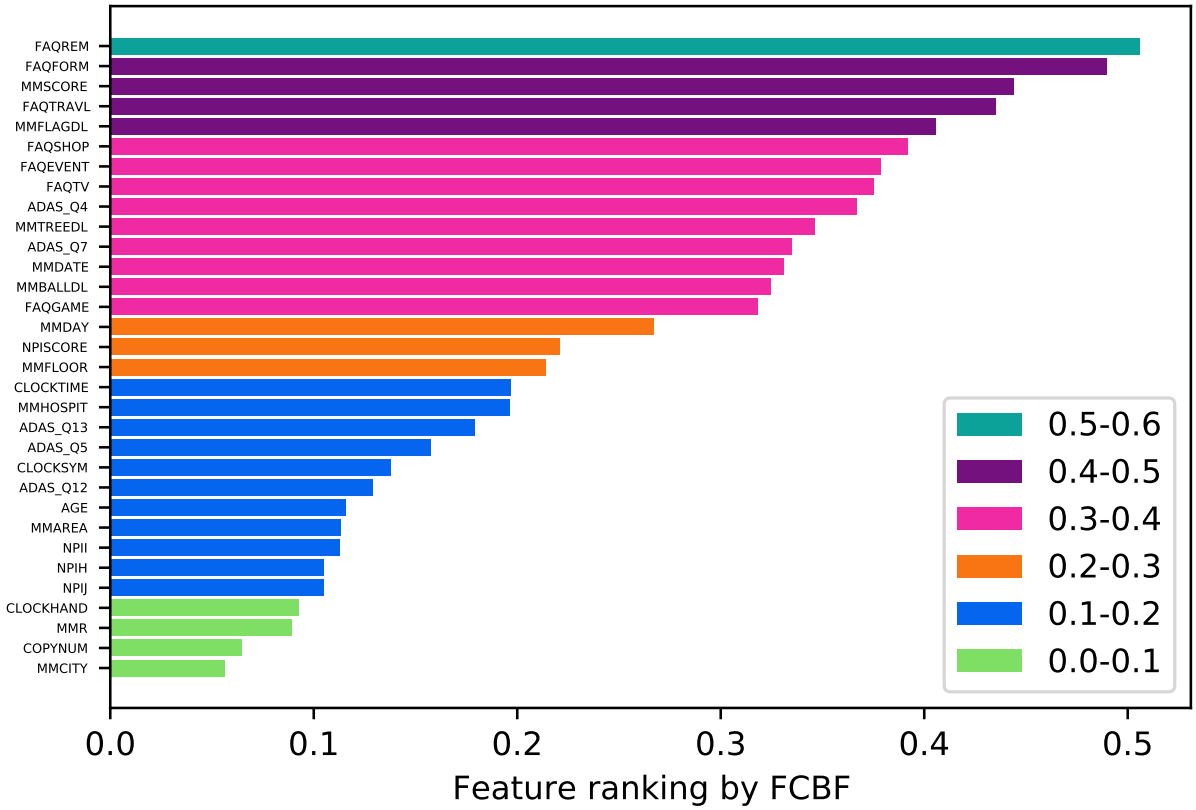


Figure 4.2: FCBF ranking of the features used in the non-invasive approach, CN-AD classification task, ADNI1 phase.

Gradient Boosting (XGBoost) [Chen and Guestrin, 2016], already described in section 2.2. The former was preferred because, generally, it allows having a reduced final set of relevant features with good correlation to the classes but low or no redundancy between the relevant features. The latter had a complementary role.

Data from 379 subjects - 212 CN and 167 AD - that started participating since the ADNI1 phase were used.

Two different sets of features were analyzed, which were chosen by a different methodology:

- The "invasive" approach Table 4.3, which contains the same multimodal features used in the CN-MCI-AD classification task, subsection 4.5.1, and in [Cabrera-León et al., 2024b]. This feature set was built from an initial set based on the FCBF score and several refining steps which involved adding and deleting features based on data visualization techniques and feature statistics [Cabrera-León et al., 2024b]. It contains two features that can be considered invasive, hence the name given to this feature set: ABETA (as its name indicates, refers to the Amyloid beta ($A\beta$) obtained with Cerebrospinal Fluid (CSF)), and VENTRICLES (a Quantitative Magnetic Resonance Imaging (qMRI) measurement of the sum of the volumes of the ventricles).

- The "non-invasive" approach Table 4.4, which includes the age of the patient and several neuropsychological items and a single total score. All these features but the age were selected based on the FCBF score, Figure 4.2.

Table 4.3: Characteristics of the subjects in the CN-AD classification problem, "invasive" approach.

	AD	CN
Number of subjects	167	212
VENTRICLES: mean	50941.89	35555.30
VENTRICLES: StD	26395.31	20383.57
VENTRICLES: interval	[9166 - 147064]	[5834 - 118875]
VENTRICLES: p-value	1.0427e-08 ****	
ABETA: mean	627.94	1254.76
ABETA: StD	266.75	447.83
ABETA: interval	[265.6 - 2568.0]	[200.0 - 3592.0]
ABETA: p-value	1.8514e-42 ****	
FAQTOTAL: mean	12.86	0.14
FAQTOTAL: StD	6.75	0.61
FAQTOTAL: interval	[0 - 30]	[0 - 6]
FAQTOTAL: p-value	1.8102e-80 ****	
MMSCORE: mean	23.34	29.12
MMSCORE: StD	2.03	0.95
MMSCORE: interval	[18 - 27]	[26 - 30]
MMSCORE: p-value	4.7072e-65 ****	
AGE: mean	75.55	75.96
AGE: StD	7.42	4.98
AGE: interval	[55.1 - 90.9]	[59.9 - 89.6]
AGE: p-value	0.4388 ns	

Acronyms: ns (not significant), StD (Standard Deviation).

Table 4.4: Characteristics of the subjects in the CN-AD classification problem, "non-invasive" approach.

	AD	CN
Number of subjects	167	212
FAQREM: mean	3.62	0.12
FAQREM: StD	1.36	0.62
FAQREM: interval	[0 - 5]	[0 - 4]
FAQREM: *p*-value	1.1064e-77 ****	
FAQFORM: mean	3.60	0.12
FAQFORM: StD	1.59	0.54
FAQFORM: interval	[0 - 5]	[0 - 5]
FAQFORM: *p*-value	3.5384e-71 ****	
FAQTRAVL: mean	3.11	0.03
FAQTRAVL: StD	1.92	0.29
FAQTRAVL: interval	[0 - 5]	[0 - 3]
FAQTRAVL: *p*-value	5.3548e-61 ****	
MMSCORE: mean	23.34	29.12
MMSCORE: StD	2.03	0.95
MMSCORE: interval	[18 - 27]	[26 - 30]
MMSCORE: *p*-value	4.7072e-65 ****	
MMFLAGDL: mean	1.84	1.13
MMFLAGDL: StD	0.36	0.34
MMFLAGDL: interval	[1 - 2]	[1 - 2]
MMFLAGDL: *p*-value	3.0753e-47 ****	
AGE: mean	75.55	75.96
AGE: StD	7.42	4.98
AGE: interval	[55.1 - 90.9]	[59.9 - 89.6]
AGE: *p*-value	0.4388 ns	

Acronyms: ns (not significant), StD (Standard Deviation).

As what was done with the data imputation and feature ranking methods, in this and the rest of classification tasks it was required to analyze several scaling techniques, projection methods and number of projected components. The next scaling techniques were studied, section 2.3: Unscaled, Standard, Robust, MaxAbs, MinMax, Normalizer, and Yeo-Johnson Power transformer. Regarding projection methods, this set was analyzed, albeit only the winner ones were explained in section 2.4: Unprojected, Principal Component Analysis (PCA), Incremental PCA, FastICA, Kernel PCA with linear, Radial Basis Function (RBF), polynomial, sigmoid or cosine kernels, Neighborhood Component Analysis (NCA), Factor Analysis, t-distributed stochastic neighbor embedding (t-SNE), dictionary learning, and Linear Discriminant Analysis.

Choosing the most adequate combinations of them involved mixing the rates given by data visualization techniques and clustering quality metrics. Scatter plots of the scaled and projected data together with an enhanced version of box-and-whisker plots called "boxenplots"[7] [Hofmann et al., 2017] of each selected feature were used for data visualization. Clustering quality metrics were limited to silhouette [Rousseeuw, 1987], Davies-Bouldin [Davies and Bouldin, 1979] and Caliński-Harabasz [Caliński and Harabasz, 1974], which were used for other purposes in [Pellicer Sarmiento, 2024]. Both silhouette and Davies-Bouldin are recommended for optimizing feature selection in unsupervised clustering tasks [McCrory and Thomas, 2025]. On the other hand, statistical tests were used to discard non-significant features, although this was deemed secondary, non-mandatory.

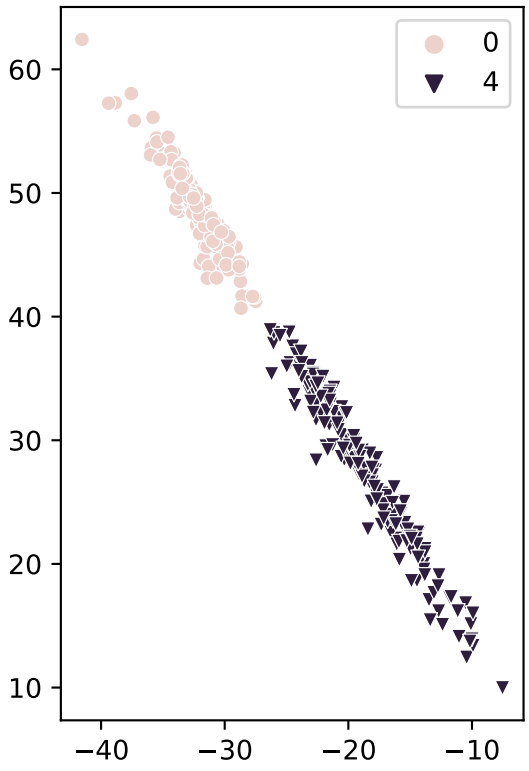
Four different preprocessed datasets were selected according to their scatter plots and boxenplots, Figure 4.3 and Figure 4.4. Data in all of them were projected to two projection components with NCA [Goldberger et al., 2004] and the parameter regarding the initialization of the linear transformation set to a value of "identity" (that is, the identity matrix was truncated to the first number of projection components rows). Three of the four made use of the "invasive" approach, and only one used the "non-invasive" features. The other main differences were the addition of the age of the subject in one case, and the use of whether the "MaxAbs" or the "Robust" scaling methods, already described in section 2.3, each one used by half of the preprocessed datasets. Figure 4.3 and Figure 4.4 show that the four preprocessed datasets can be considered linearly separable.

The relation between each component and the features is shown in Figure 4.5 and Figure 4.5. The "non-invasive" approach differs as MMSCORE is the only feature with a negative relation in both components, whereas the rest are all positive. In the approaches that are the most similar and which only different in using or not AGE, the relation is identical in the non-AGE features.

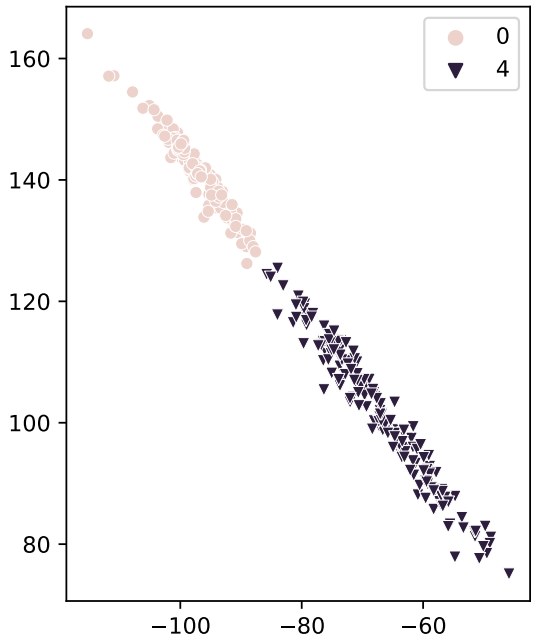
4.2.2 Processing and results with the SupeRGNG

Data partitioning, or splitting, is key for the reliable evaluation of a model as it is recommended that no samples provided to the model during its training are found in the data used for testing it. That is, data partitioning is used for model validation. The Stratified K-Folds cross-validation method, with 5 folds as it kept the same 80-20 training-test subsets proportion used in [Sosa-Marrero et al., 2021; Cabrera-León et al., 2023, 2024b], was used for data partitioning in all the applications but MCI-AD. Using it allowed making a quantitative

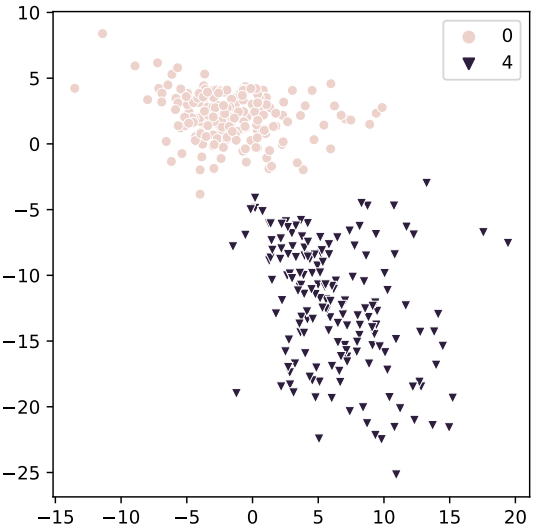
[7]Due to choosing the value "area" for the "width_method" parameter in the implementation of the boxenplots of the "seaborn" Python library used, the width of each letter value box represents the density of data points in that box [Waskom, 2021].



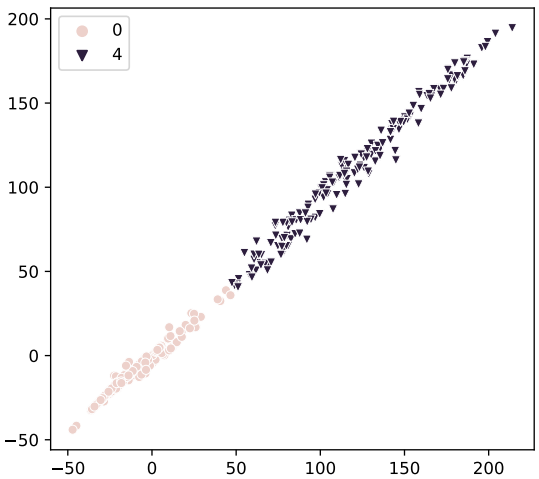
a) "Invasive" features, with AGE, 2 projection components, MaxAbsScaler, NeighborhoodCA identity.



b) "Invasive" features, without AGE, 2 projection components, MaxAbsScaler, NeighborhoodCA identity.

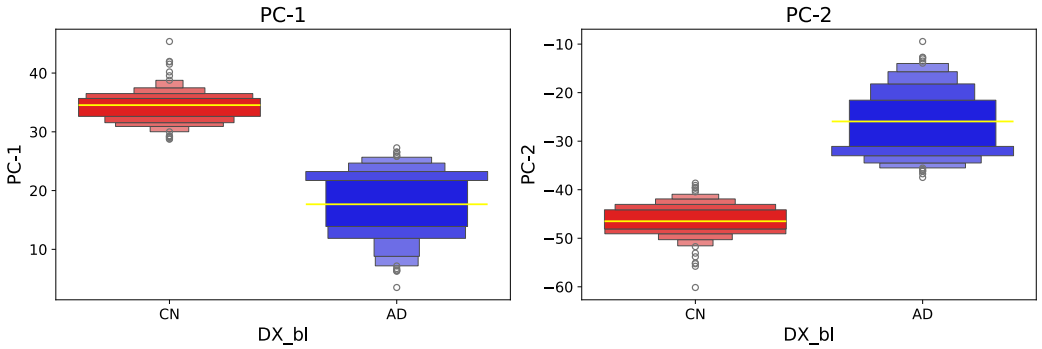


c) "Invasive" features, with AGE, 2 projection components, RobustScaler, NeighborhoodCA identity.

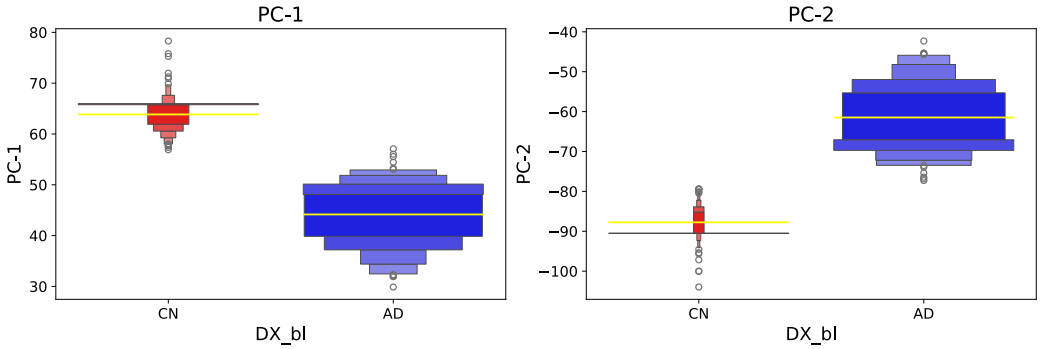


d) "Non-invasive" features, with AGE, 2PC, RobustScaler, NeighborhoodCA identity.

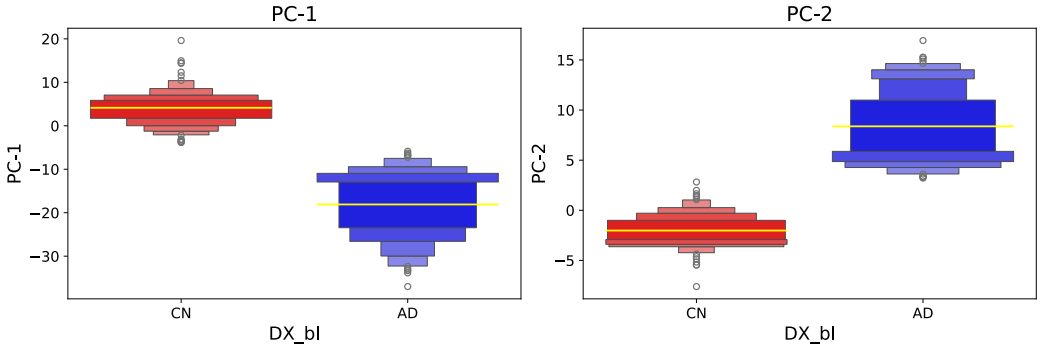
Figure 4.3: Scatter plots of the four preprocessed datasets for the CN-AD classification task.



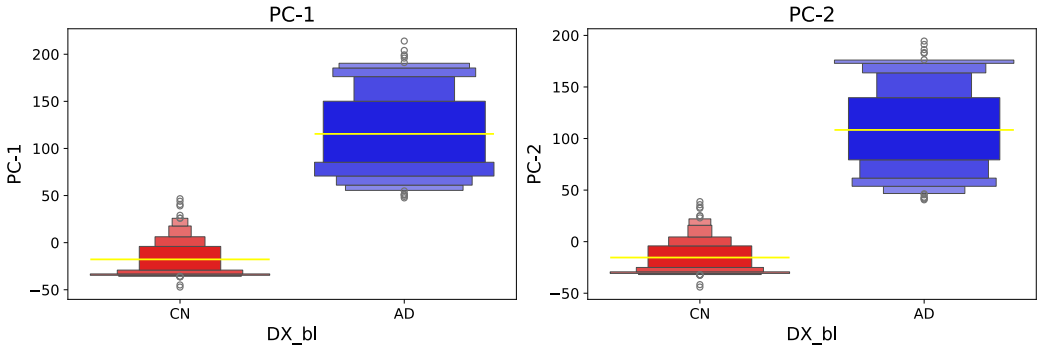
a) "Invasive" features, with AGE, 2 projection components, MaxAbsScaler, NeighborhoodCA identity.



b) "Invasive" features, without AGE, 2 projection components, MaxAbsScaler, NeighborhoodCA identity.

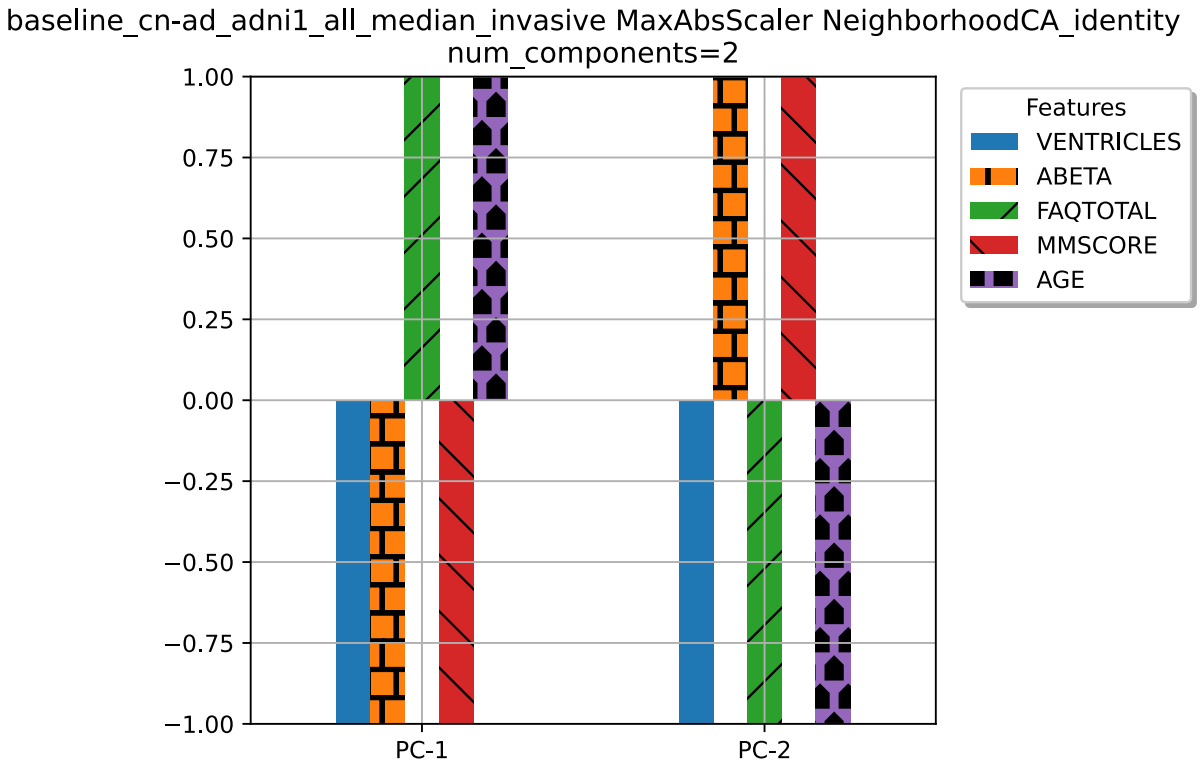


c) "Invasive" features, with AGE, 2 projection components, RobustScaler, NeighborhoodCA identity.

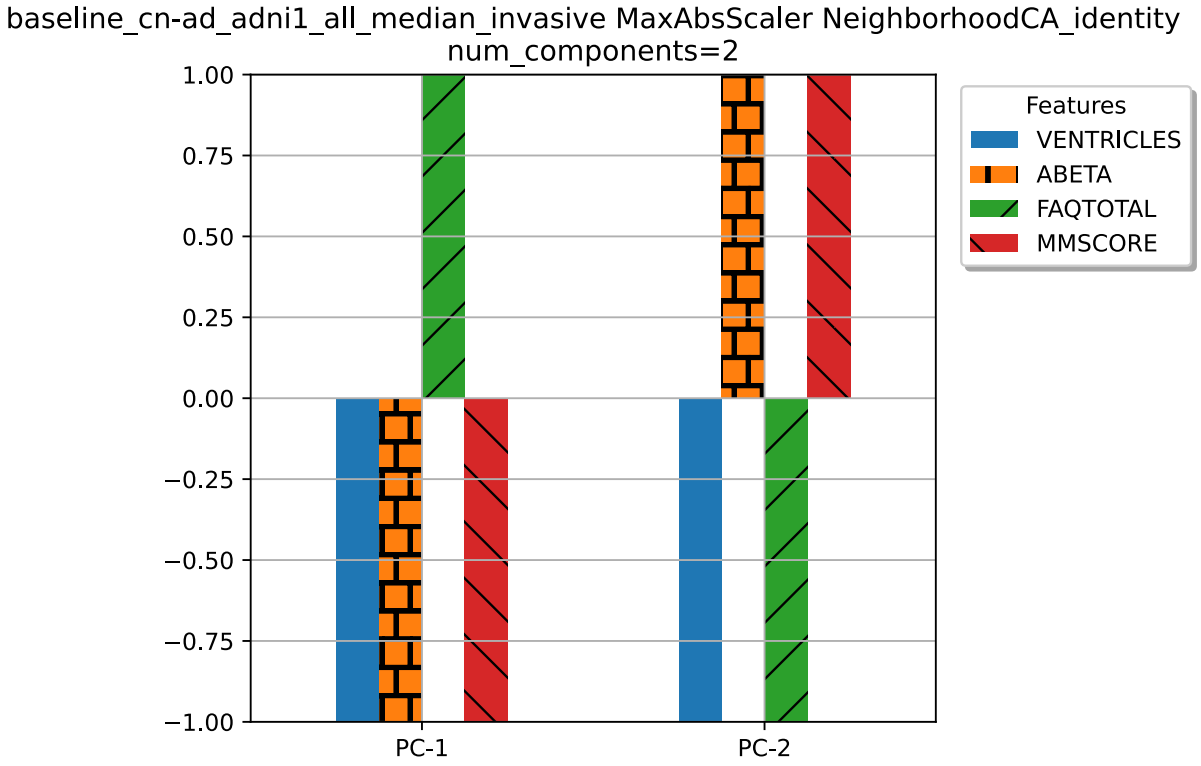


d) "Non-invasive" features, with AGE, 2 projection components, RobustScaler, NeighborhoodCA identity.

Figure 4.4: Boxenplots of the four preprocessed datasets for the CN-AD classification task.

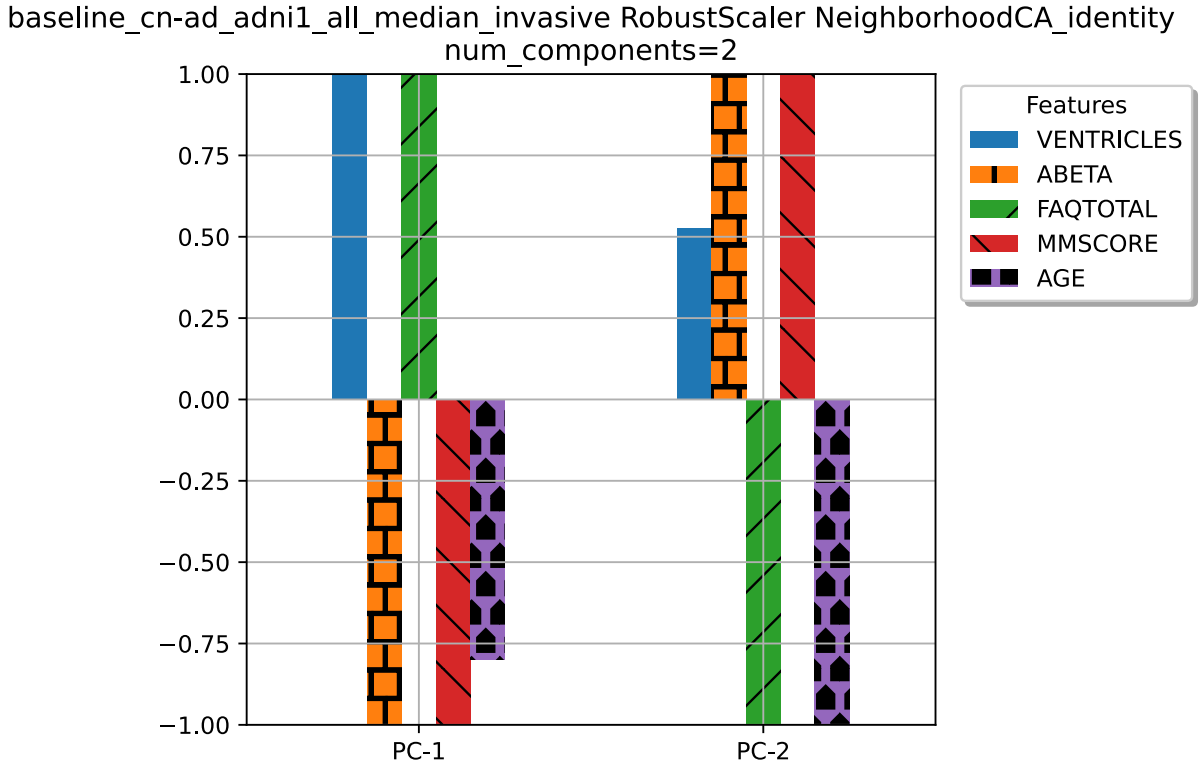


a) "Invasive" features, with AGE, 2 projection components, MaxAbsScaler, NeighborhoodCA identity.

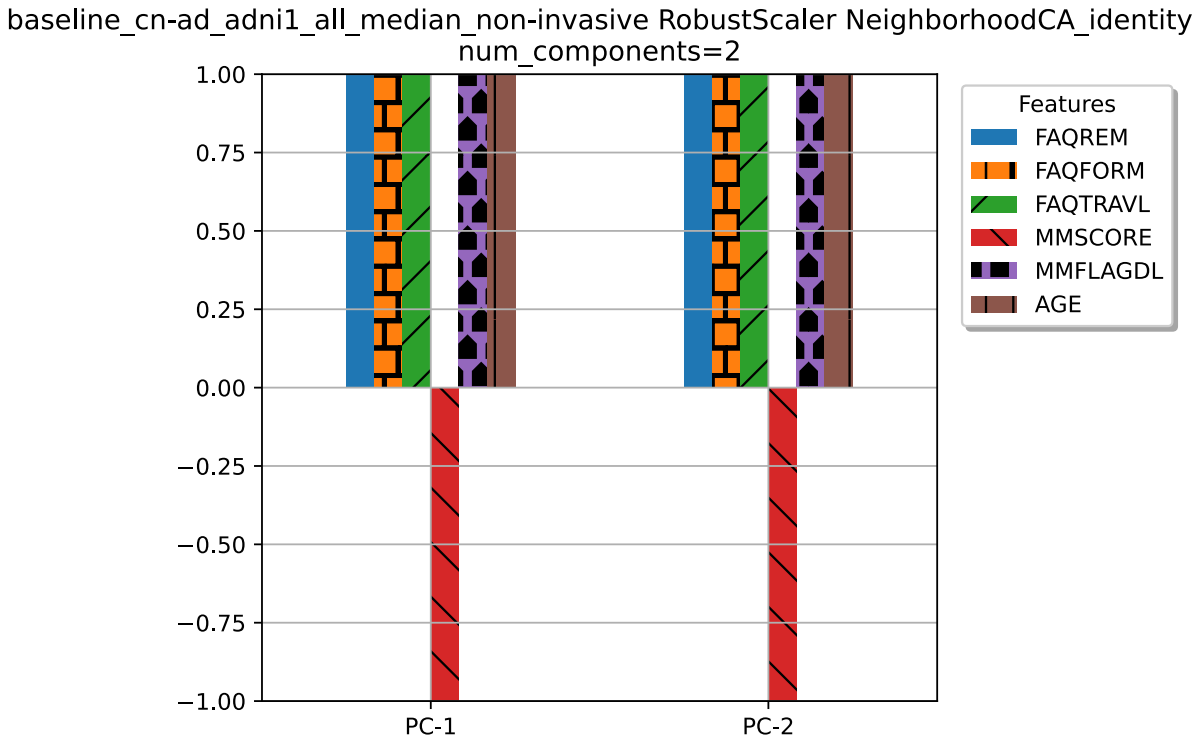


b) "Invasive" features, without AGE, 2 projection components, MaxAbsScaler, NeighborhoodCA identity.

Figure 4.5: Relations between each projected component and the real features of the first two of four preprocessed datasets for the CN-AD classification task.



c) "Invasive" features, with AGE, 2 projection components, RobustScaler, NeighborhoodCA identity.



d) "Non-invasive" features, with AGE, 2 projection components, RobustScaler, NeighborhoodCA identity.

Figure 4.5: Relations between each projected component and the real features of the last two of four preprocessed datasets for the CN-AD classification task.

Table 4.5: Set of 225 configurations that was tested in all the datasets used in the CN-AD classification task.

Hyperparameter	Values
λ	[100, 500, 900]
a_{max}	[100, 500, 900]
max_nodes	[50, 100, 150, 200, 250]
η	[25, 50, 75, 100, 125]
ι	[303]

comparison with all the classification tasks tackled by the intelligent system to aid in AD diagnosis made with the MyGNG [Cabrera-León et al., 2024b]. Cross-validation refers to using different portions of the data to train and test a model in each fold so that a more accurate assessment of the generalization power of the model with unseen data can be inferred. Compared with the normal K-Folds, the main advantage of the Stratified K-Folds is that it keeps the same proportion of samples for each class in all the folds.

The same methodology to analyze and select the most optimal network configurations was followed in all the classification tasks. As choosing an optimal set of hyperparameters is a difficult task [Hastie et al., 2009], a comprehensive approach was followed. A grid search starting with values considered good during the study of hyperparameters of the SupeRGNG found in the previous section was the initial step. Grid searches were then carried out with increasingly refined values chosen from the best candidates from previous searches. After several searches have already been conducted and no further performance improvement could be obtained if there was still room for improvement (that is, the perfect value of the metric AUC was not obtained yet), the searching process was stopped.

The same set of SupeRGNG configurations was tested with the four preprocessed datasets, whose values of the hyperparameters[8] are shown in Table 4.5. The value of ι was calculated as the 80% of the total number of patients so that the disconnection and reconnection processes in the SupeRGNG were carried out just after a full epoch ended. The value of 80% equals the data partitioning percentage that was used for training. The values of ι in the other classification tasks were calculated the same way, and they differed because of the sizes of the input data.

No further sets of configurations were tested as several configurations in this set already obtained perfect performance values in all the four preprocessed datasets: after unifying the values of the 5 folds, a maximum value of 1 in AUC and the rest of performance metrics, and a mean of 1 and standard deviation of 0. Six configurations achieved the perfect performance, which where characterized by the next values of the hyperparameters: $\lambda = 100, a_{max} = \{100, 500, 900\}, max\_nodes = 100, \eta = \{100, 125\}$, in all the four preprocessed datasets, Table 4.6. It was noticed that values of $\eta \leq 75$ did not allow to obtain perfect performance in any of them.

Considering the good values, the only difference between the 4 preprocessed datasets was the quantity of configurations that yielded the perfect values. The least number was with the non-invasive one, followed by the one that did not use AGE. This might be explained by both datasets having the least inter-class distances, Figure 4.3, an expected result, which

[8] As already indicated, fixed values were used for certain hyperparameters of the SupeRGNG, which were indicated in Table 3.4.

Table 4.6: Best configurations in the CN-AD classification task.

Configuration	λ	a_{max}	max_nodes	η
config8	100	100	100	100
config9	100	100	100	125
config33	100	500	100	100
config34	100	500	100	125
config58	100	900	100	100
config59	100	900	100	125

Configuration	Accu	Sens	Spec	Prec	AUC	CUI+	CUI-
config8	1	1	1	1	1	1	1
config9	1	1	1	1	1	1	1
config33	1	1	1	1	1	1	1
config34	1	1	1	1	1	1	1
config58	1	1	1	1	1	1	1
config59	1	1	1	1	1	1	1

points out the necessity to always tune appropriately the hyperparameters.

4.3 CN-MCI

As indicated above, the same methodologies were followed to choose the input data, preprocessing and processing.

4.3.1 Input data and preprocessing

Data from 585 subjects - 212 CN and 373 MCI - that started participating since the ADNI1 phase were used.

The set of features shown in Table 4.7 was analyzed, which was built based on their FCBF scores, Figure 4.6. It includes a biomarker obtained via CSF, ABETA, which was also used in the "invasive" approach in CN-AD and in CN-MCI-AD, the age of the participant, 3 items of the Mini-Mental State Examination (MMSE) neuropsychological scale, and a subscore of Neuropsychiatric Inventory (NPI).

Among all the scaling and projection methods, two combinations were selected *a priori*: Robust scaler and t-SNE [Hinton and Roweis, 2002; van der Maaten and Hinton, 2008] for data projection, and Unscaled data and Kernel PCA [Schölkopf et al., 1998b] with a RBF kernel, Figure 4.7. t-SNE has a parameter called "perplexity" which is similar to the number of nearest neighbors that is used in other manifold learning algorithms. Among the possible values of perplexity, 100 yielded the best scatter plots and boxenplots although it was slower. These plots looked more promising than those from the other combination of scaling and projection methods that was considered a good candidate: Unscaled data and Kernel PCA [Schölkopf et al., 1998b] with a RBF kernel. In both cases 6 projected components were deemed better, Figure 4.8 and Figure 4.9. However, after analyzing the

Table 4.7: Characteristics of the subjects in the CN-MCI classification problem.

	MCI	CN
Number of subjects	373	212
ABETA: mean	725.24	1254.76
ABETA: StD	328.29	447.83
ABETA: interval	[210.9 - 2809.0]	[200.0 - 3592.0]
ABETA: *p*-value	2.8213e-45 ****	
AGE: mean	74.96	75.96
AGE: StD	7.42	4.98
AGE: interval	[55.2 - 89.3]	[59.9 - 89.6]
AGE: *p*-value	0.2204 ns	
MMBALLDL: mean	1.24	1.03
MMBALLDL: StD	0.42	0.18
MMBALLDL: interval	[1 - 2]	[1 - 2]
MMBALLDL: *p*-value	3.1878e-12 ****	
MMDAY: mean	1.11	1
MMDAY: StD	0.31	0.07
MMDAY: interval	[1 - 2]	[1 - 2]
MMDAY: *p*-value	2.3008e-07 ****	
NPIL: mean	0.11	0
NPIL: StD	0.31	0.07
NPIL: interval	[0 - 1]	[0 - 1]
NPIL: *p*-value	2.3008e-07 ****	
MMYEAR: mean	1.03	1
MMYEAR: StD	0.16	0
MMYEAR: interval	[1 - 2]	[1 - 1]
MMYEAR: *p*-value	0.0164 *	

Acronyms: ns (not significant), StD (Standard Deviation).

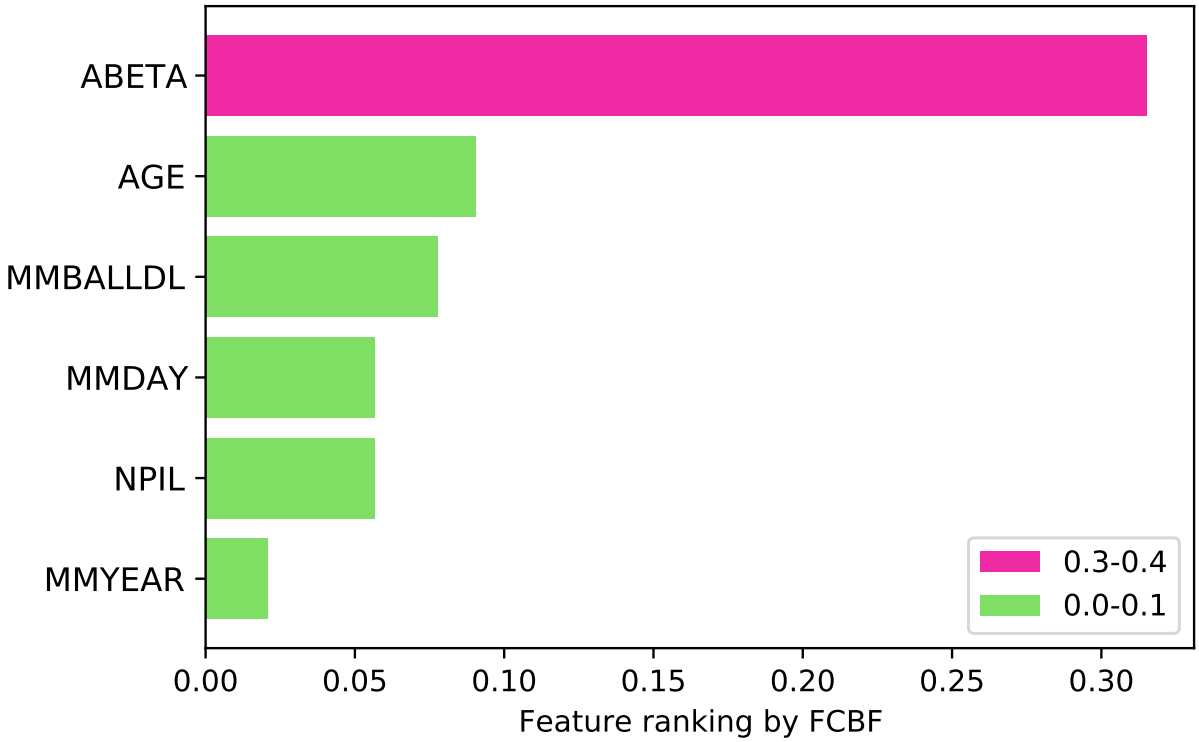


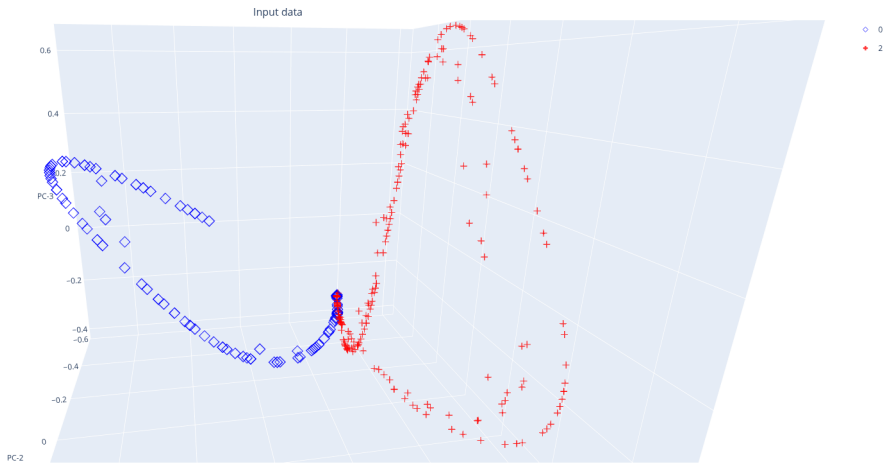
Figure 4.6: FCBF ranking of the features used in the CN-MCI classification task, ADNI1 phase.

scatterplots and boxenplots, only the one that used the Robust scaling method and the t-SNE for data projection was finally studied.

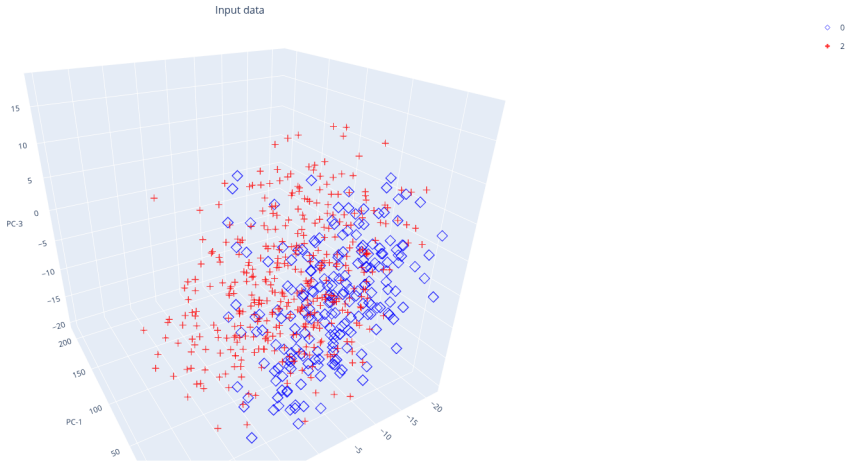
4.3.2 Processing and results with the SupeRGNG

Several sets of configurations with different values of the hyperparameters were tested, Table 4.8. The value of ι was set to 468, that is, the 80% of the size of the input data that was used for training. The search process started with the first set and, after choosing the best network configuration, it continues by building a new set of features where configurations have values close to the one considered the best. This was done iteratively until no further improvements was found. In some cases, the second or third configurations were also tried. At the end, the optimal configurations of each set were compared and analyzed.

There were two configurations that achieved similarly good performance results, Table 4.9. The Receiver Operating Characteristic (ROC) curves of both are shown in Figure 4.10. It should be mentioned that "config248" yielded almost the same performance with just $\eta = 75$ as "config714" did with 380. That is, "config248" was trained for 5 folds in 1min 50s (mean time per fold was 22s aprox.) and always early stopped before epoch 90, whereas the latter, 7min 26s (mean time per fold was 1min 29s aprox.) and before epoch 409. This is an example of how metrics other than performance ones, section 2.5, should also be taken into account when comparing models.



a) 6 projection components, Unscaled, KernelPCA with “rbf” kernel.



b) 6 projection components, RobustScaler, T-SNE with perplexity=100.

Figure 4.7: Scatter plots of the two preprocessed datasets used for the CN-MCI classification task.

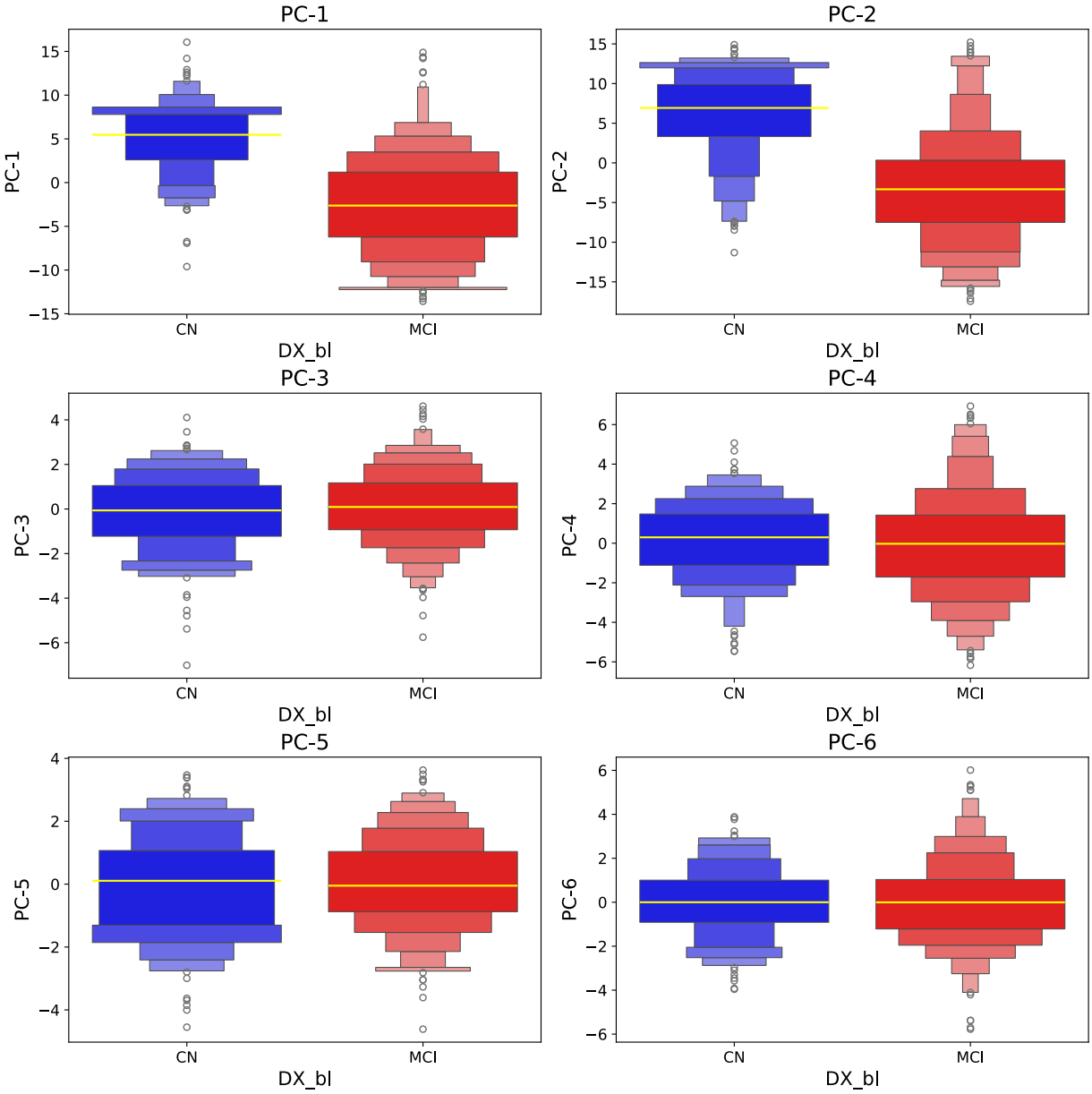


Figure 4.8: Boxenplots of the preprocessed dataset used for the CN-MCI classification task (6 projection components, RobustScaler, t-SNE with perplexity=100).

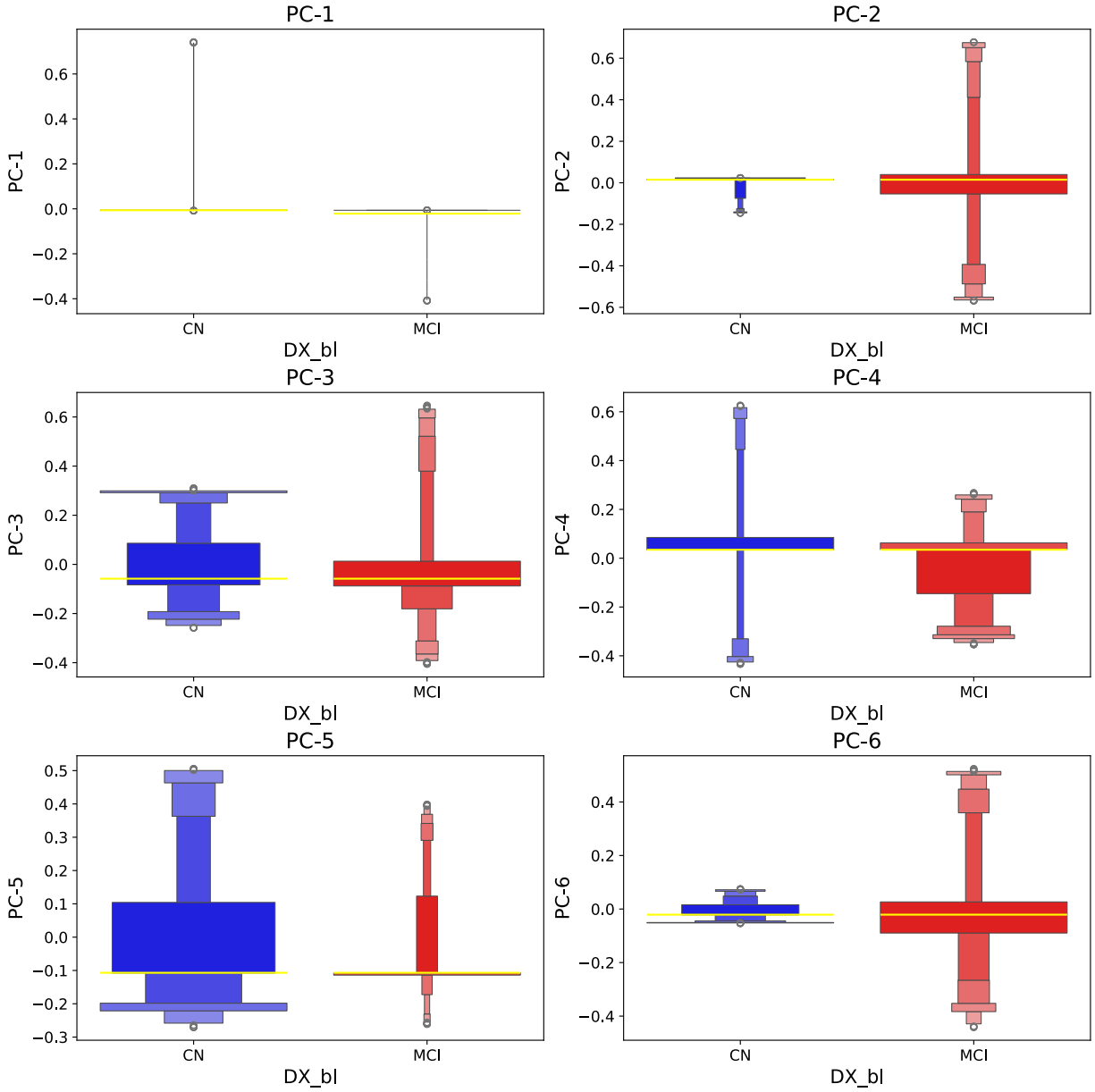


Figure 4.9: Boxenplot of the first preprocessed dataset for the CN-MCI classification task: 6 projection components, Unscaled, KernelPCA with "rbf" kernel.

Table 4.8: Sets of configurations that were tested in the dataset used in the CN-MCI classification task. The number of configurations in each of these tests was: 4536, 2205, 1890, 1225 and 1050.

Hyperparameter	Values
λ	[25, 50, 75, 100, 125, 150, 175, 200]
a_{max}	[25, 50, 75, 100, 125, 150, 175, 200]
max_nodes	[40, 90, 140, 190, 240, 290, 340]
η	[75, 125, 175, 225, 275, 325, 375, 425]
ι	[468]
λ	[130, 135, 140, 145, 150, 155, 160, 165, 170]
a_{max}	[60, 65, 70, 75, 80, 85, 90]
max_nodes	[60, 70, 80, 90, 100, 110, 120]
η	[355, 365, 375, 385, 395]
ι	[468]
λ	[130, 135, 140, 145, 150, 155, 160, 165, 170]
a_{max}	[60, 65, 70, 75, 80, 85, 90]
max_nodes	[40, 50, 85, 95, 200, 250]
η	[355, 365, 375, 385, 395]
ι	[468]
λ	[133, 135, 137, 140, 143, 145, 147]
a_{max}	[88, 90, 92, 94, 96, 98, 100]
max_nodes	[180, 190, 200, 210, 220]
η	[370, 373, 375, 377, 380]
ι	[468]
λ	[130, 131, 132, 133, 134, 135]
a_{max}	[90, 91, 92, 93, 94]
max_nodes	[212, 215, 218, 220, 222, 225, 237]
η	[375, 376, 377, 378, 379]
ι	[468]

Table 4.9: Best configurations in the CN-MCI classification task.

Configuration	λ	a_{max}	max_nodes	η
config714	143	88	200	380
config248	25	125	190	75

Configuration	Accu	Sens	Spec	Prec	AUC	CUI+	CUI-
config714	0.86	0.86	0.86	0.86	0.88	0.74	0.74
config248	0.85	0.85	0.85	0.85	0.88	0.73	0.73

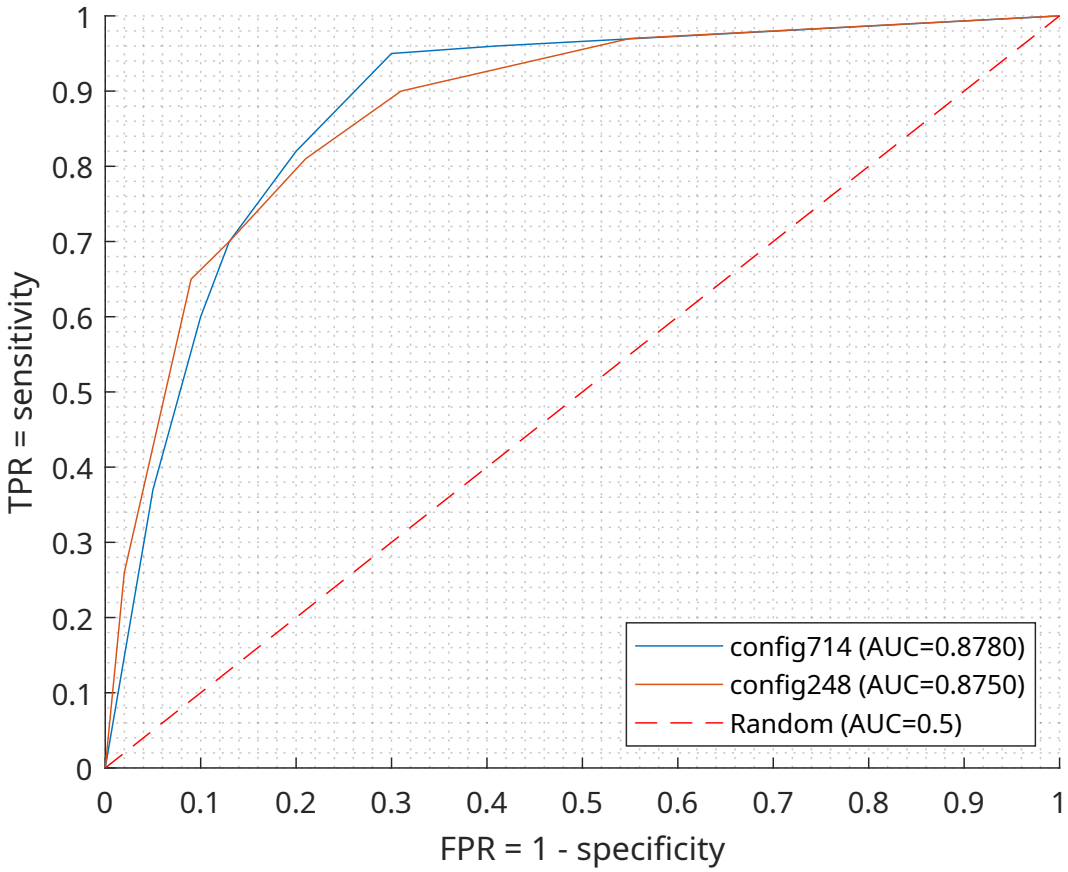


Figure 4.10: ROC curves of the best SupeRGNG configurations for the CN-MCI classification task (6 projection components, RobustScaler, t-SNE with perplexity=100).

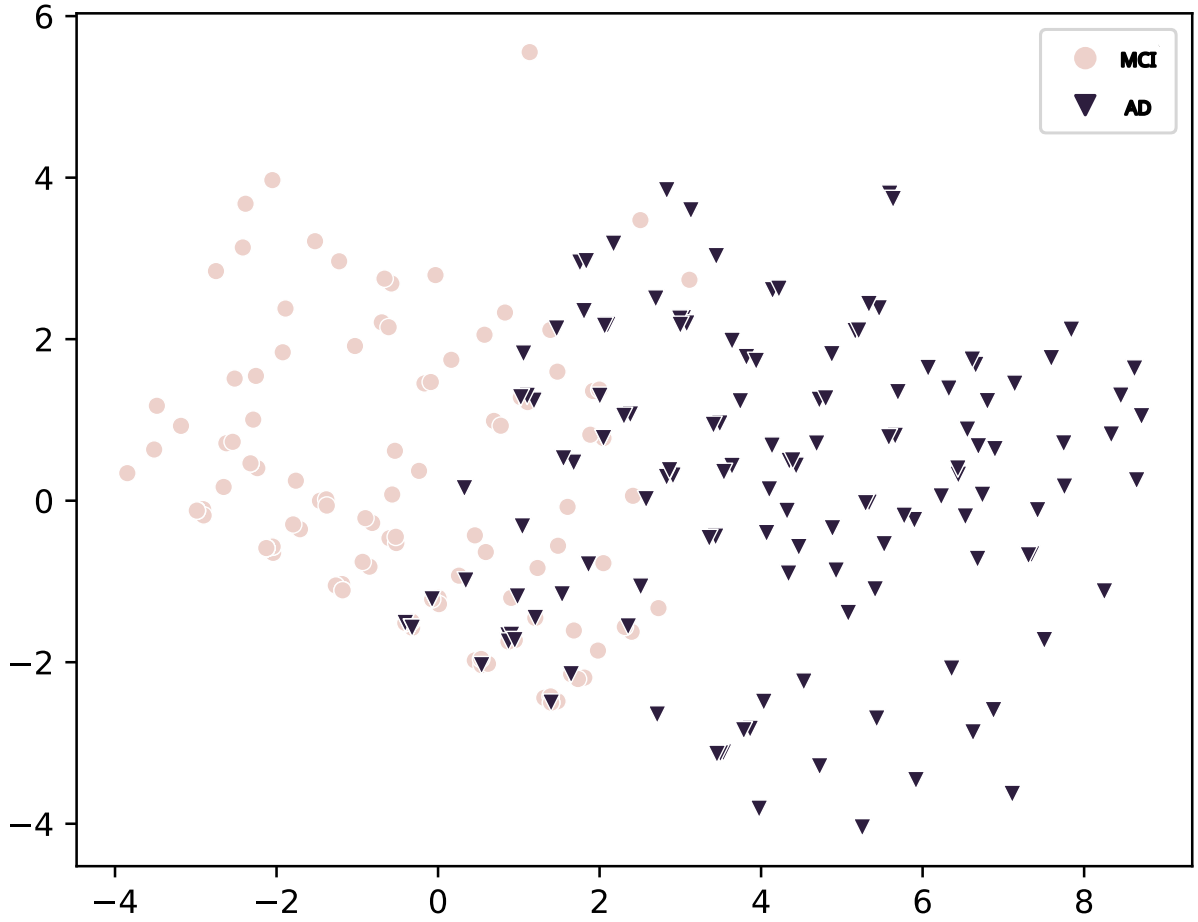


Figure 4.11: Scatterplot of the dataset used for MCI-AD: Unscaled and PCA with 2 projected components.

4.4 MCI-AD

In order to compare the previous version of the SupeRGNG, where the disconnection and reconnection mechanisms were only performed when the Growing Neural Gas (GNG) layer finished clustering at the end of the epochs, the same dataset and preprocessing were used [Cabrera-León et al., 2023]. Also, the data partitioning was different to the one used in [Cabrera-León et al., 2024b].

4.4.1 Input data and preprocessing

Data from 495 subjects — 345 MCI and 150 AD — that belonged to the ADNI2 phase were used. A different ADNI subset was used because the selected one showed less number of missing values in the originally non-imputed data so the FCBF ranking of the imputed data was considered more realistic. The characteristics of the subject in this dataset can be seen in Table 4.10 and a scatterplot in Figure 4.11.

For the same reason explained in the first paragraph of this subsection, the same data preprocessing as the study in [Cabrera-León et al., 2023] was used: FCBF feature ranking, PCA with 2 projected components, and the data was not scaled. This is different to the one

Table 4.10: Characteristics of the subjects in the MCI-AD classification problem: a demographic feature, and the six attributes used by the model as input, sorted according to their FCBF score.

	AD	MCI
Number of subjects	150	345
AGE: mean	74.67	71.56
AGE: StD	8.18	7.38
AGE: interval	[55.6 - 90.3]	[55.0 - 91.4]
AGE: *p*-value	0.3305 ns	
MMSCORE: mean	23.07	27.98
MMSCORE: StD	2.08	1.74
MMSCORE: interval	[19 - 26]	[24 - 30]
MMSCORE: *p*-value	1.9637e-54 ****	
MMDATE: mean	1.6	1.08
MMDATE: StD	0.49	0.26
MMDATE: interval	[1 - 2]	[1 - 2]
MMDATE: *p*-value	9.3794e-35 ****	
MMBALLDL: mean	1.67	1.14
MMBALLDL: StD	0.47	0.35
MMBALLDL: interval	[1 - 2]	[1 - 2]
MMBALLDL: *p*-value	1.4736e-31 ****	
ADAS_Q7: mean	2.4	0.43
ADAS_Q7: StD	1.75	0.85
ADAS_Q7: interval	[0 - 7]	[0 - 8]
ADAS_Q7: *p*-value	1.1044e-31 ****	
MMYEAR: mean	1.27	1.0
MMYEAR: StD	0.44	0.05
MMYEAR: interval	[1 - 2]	[1 - 2]
MMYEAR: *p*-value	9.3548e-22 ****	
FAQSHOP: mean	2.79	0.5
FAQSHOP: StD	1.82	1.14
FAQSHOP: interval	[0 - 5]	[0 - 5]
FAQSHOP: *p*-value	3.1878e-12 ****	

Acronyms: ns (not significant), StD (Standard Deviation).

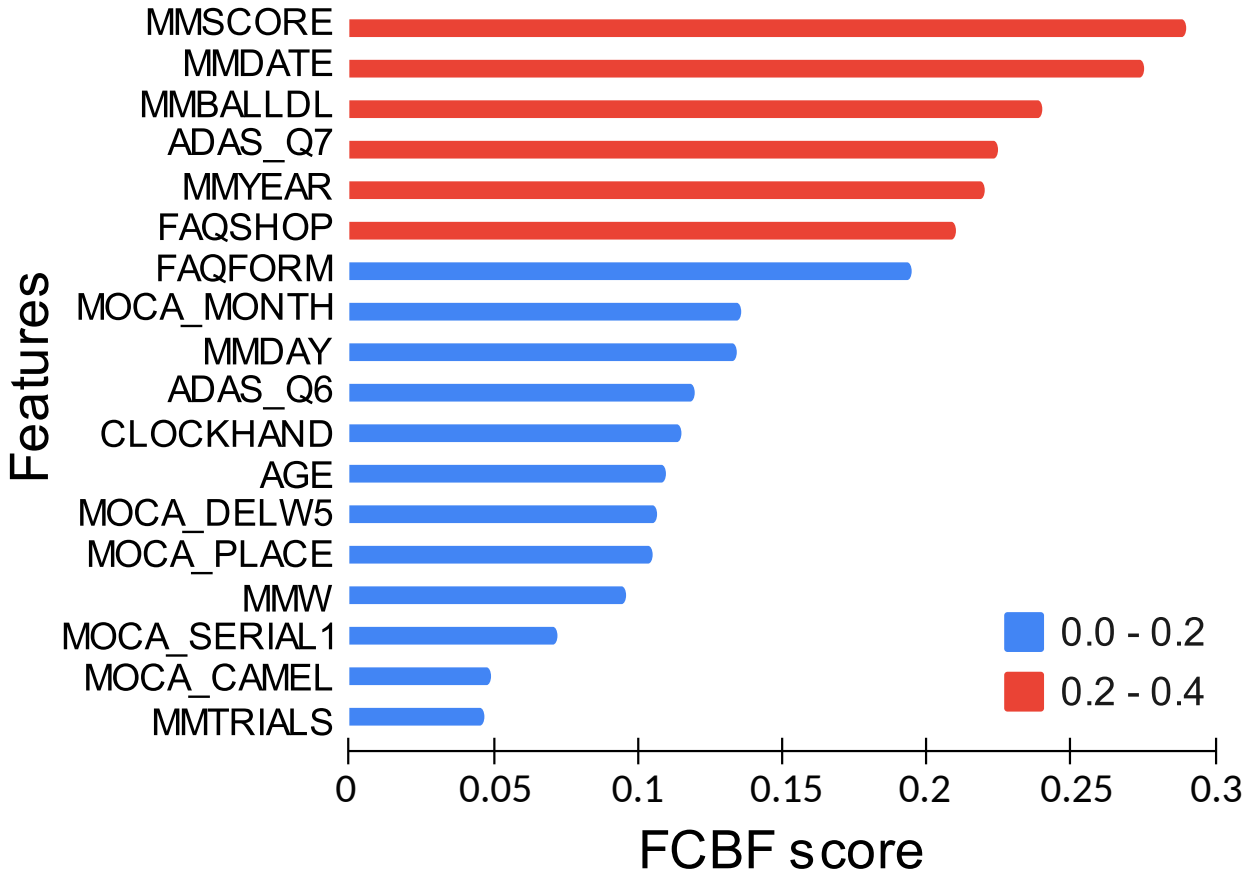


Figure 4.12: Ranking of features according to the FCBF method for MCI and AD subjects from ADNI2.

used in [Cabrera-León et al., 2024b]. The ranking of the features is shown in Figure 4.12, and those with the highest score (in red) were selected.

4.4.2 Processing and results with the MyGNG

Partitioning of the data was done according to the training-test method[9], using 80% for training the models and the remaining 20% for testing them. This same proportion was used in [Sosa-Marrero et al., 2021; Cabrera-León et al., 2023, 2024b].

The values of the hyperparameters of the GNG that were considered appropriate were those found after the search stated in subsection 3.2.2: $max\_nodes = 100, a_{max} = 5, \lambda = 25, \varepsilon_b = 0.9, \varepsilon_n = 0.01, \beta = 0.7, d = 0.18, epochs = 7$. Albeit ρ does not affect the numerical stability of the learning algorithm of the perceptron, it does affect the convergence [Widrow and Lehr, 1990]. In this case, these values were assigned: $\rho = 0.01, epochs\_perceptron = 50$.

The ROC curve of the best MyGNG configuration is shown in Figure 4.13, which yielded 0.9292 accuracy, 0.9710 sensitivity, 0.8333 specificity, 0.9305 precision and 0.963 AUC.

[9]It can be referred as the "hold-out method" [Arlot and Celisse, 2010], which can be considered the simplest form of cross-validation method, with only 1 split of the data.

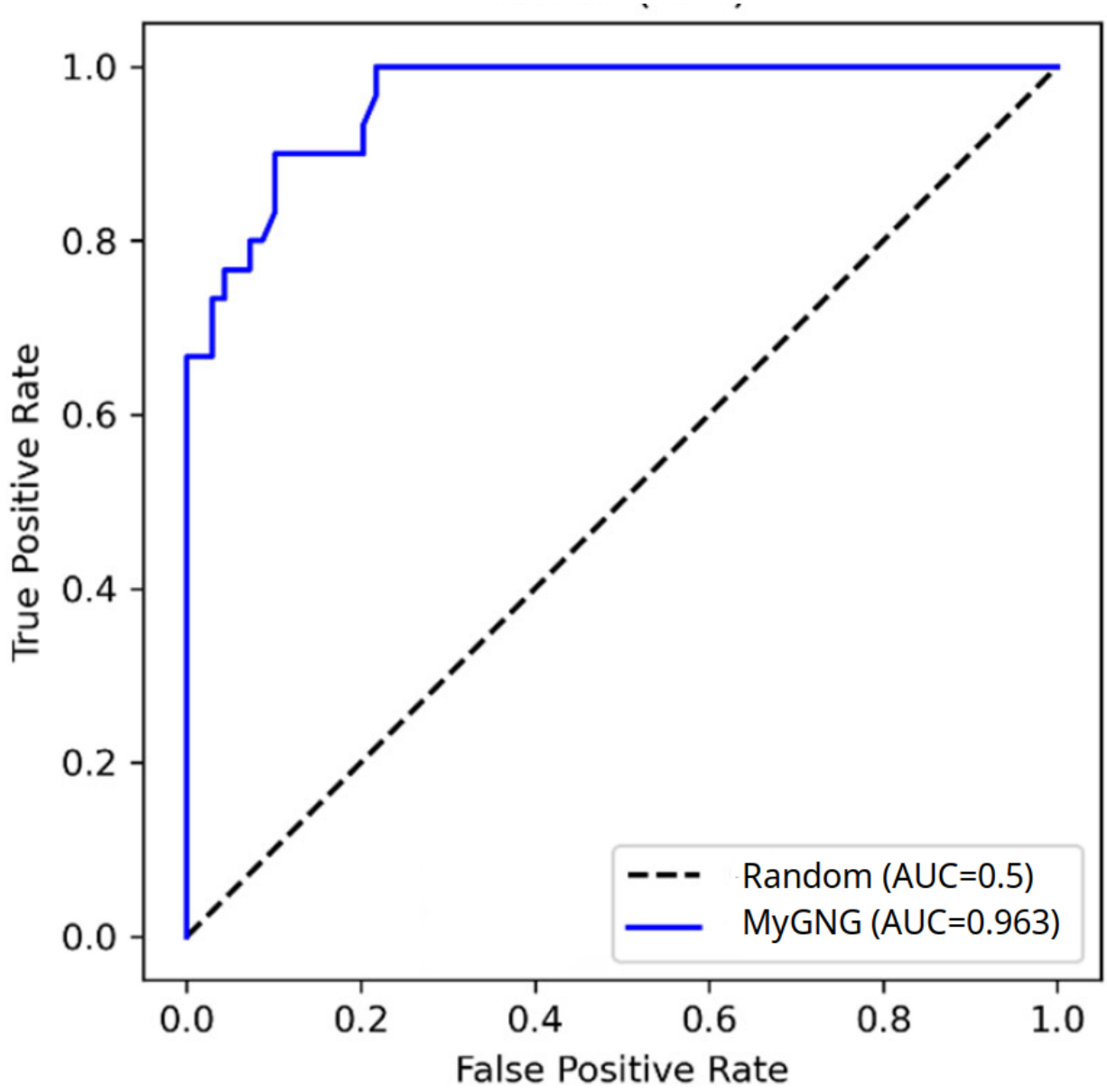


Figure 4.13: ROC curve of the best MyGNG configuration for the MCI-AD classification task (2 projection components, Unscaled, PCA).

Table 4.11: Initial sets of configurations that were tested in the MCI-AD classification task. The number of configurations in each of these tests was: 1715, 1715 and 1715.

Hyperparameter	Values
λ	[635, 640, 645, 650, 655, 660, 665]
a_{max}	[140, 145, 150, 155, 160, 165, 170]
max_nodes	[35, 40, 45, 50, 55, 60, 65]
η	[10, 20, 30, 40, 50] [100, 200, 300, 400, 500] [600, 700, 800, 900, 950]
ι	[396]

Table 4.12: Sets of configurations that were tested in the MCI-AD classification task. The number of configurations in each of these tests was: 1715, 1715 and 1715.

Hyperparameter	Values
λ	[1070, 1075, 1080, 1085, 1090, 1095, 1100]
a_{max}	[845, 850, 855, 860, 865, 870, 875]
max_nodes	[35, 40, 45, 50, 55, 60, 65]
η	[10, 20, 30, 40, 50] [100, 200, 300, 400, 500] [600, 700, 800, 900, 950]
ι	[396]

4.4.3 Processing and results with the SupeRGNG

As with the MyGNG, data partitioning was 80% for training and 20% for testing.

The optimal set of hyperparameters stated in [Cabrera-León et al., 2023] was used to define the initial values around which the grid search started, Table 4.11. However, more experiments were carried out not only because there are more hyperparameters in this version of the SupeRGNG but also in order to try to improve the performance values, Table 4.12.

Results from the latter set induced a new seed of values for several hyperparameters, so closer values to the optimal one in that set were then explored. With the new set, Table 4.13, several SupeRGNG configurations yielded results equivalent to the ones reported as optimal in [Cabrera-León et al., 2023]. The ROC curves of the two best configurations are depicted in Figure 4.14, and the values of the performance metrics in Table 4.14.

4.5 CN-MCI-AD

Unlike the previous ones, this is the only multiclass task that was tackled in this work, where the model should be able to distinguish CN, MCI and AD subjects. As already mentioned

Table 4.13: Set of configurations that was tested in the MCI-AD classification task. The number of configurations was 5760.

Hyperparameter	Values
λ	[1081, 1082, 1083, 1084, 1085, 1086, 1087, 1088, 1089]
a_{max}	[856, 857, 859, 860, 861, 862, 863, 864, 866]
max_nodes	[41, 42, 43, 44, 45, 46, 47, 48, 49, 50]
η	[485, 490, 495, 500, 505, 510, 515]
ι	[396]

Table 4.14: Best configurations in the MCI-AD classification task.

Configuration	λ	a_{max}	max_nodes	η
config2866	1085	860	50	500
config5459	1089	861	50	515

Configuration	Accu	Sens	Spec	Prec	AUC	CUI+	CUI-
config2866	0.97	0.97	0.97	0.97	0.97	0.94	0.94
config5459	0.98	0.98	0.98	0.98	0.97	0.96	0.96

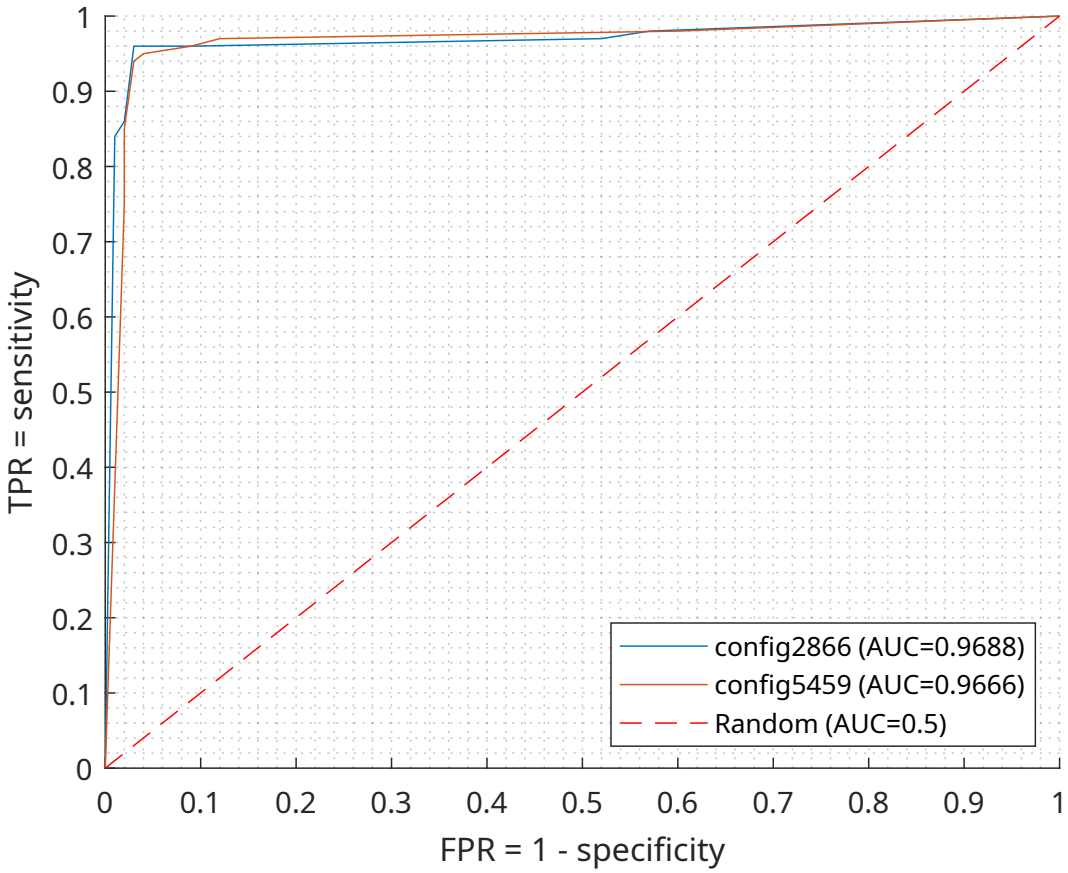


Figure 4.14: ROC curves of the best SupeRGNG configurations for the MCI-AD classification task (2 projection components, Unscaled, PCA).

in previous classification tasks, the same methodologies were followed to select input data, preprocessing and processing.

4.5.1 Input data and preprocessing

Data from 752 subjects — 212 CN, 373 MCI and 167 AD — that belonged to the ADNI1 phase were collated, Table 4.15.

The feature ranking process started with the features ranked with FCBF and keeping those with the highest FCBF score for this classification task. The ranked features were, from greater to lesser significance: ABETA, MMDATE, AGE, MMDAY, CLOCKTIME, MMMONTH, MMFLOOR, NPIG, MMHOSPIT, MMYEAR and MMREAD. It includes the age of the patient, the CSF measure of the Aβ biomarker, 7 subscales of the MMSE test, an item of the clock drawing test, and a subtest of the NPI scale.

An iterative refining process of this feature set was then performed to increase its quality (that is, higher inter-class and lower intra-class distances) and reduce its size, which derived in choosing other features. This extra process was done because the task is more difficult due to being multiclass.

Different sources of information on the features were used to perform this refining. Before including a new feature among those available, its biological relevance was analyzed according to AD-related clinical bibliography Budelier and Bateman [2020]; Gunes et al. [2022], and its relevance for the diagnostic problem addressed by means of analyzing descriptive statistics (mean, standard deviation and interval), and clustering quality metrics (silhouette, Davies-Bouldin and Calinski-Harabasz scores) da Silva et al. [2010]; Gutoski et al. [2018]; McCrory and Thomas [2025]. The usage of the latter is supported by the fact that such clustering quality metrics can help identify relevant features for clustering tasks and, ultimately, classification ones as is our case too da Silva et al. [2010]; Gutoski et al. [2018]; McCrory and Thomas [2025]. Generally, clustering quality metrics can help identify feature sets that lead to more meaningful and well-separated clusters, this way helping in the selection of relevant features for clustering tasks and, ultimately, classification ones as is our case too. On the contrary, only the adequacy for the diagnostic problem was used to discard a feature.

The final set of features included, Table 4.15: a demographic data, a quantitative neuroimaging measurement that measures the volume of the ventricles, a CSF value that measures the quantity of Aβ, and the main scores of two neuropsychological scales. As it can be observed, all these final features but AGE and ABETA were inserted during the refining process. Several references confirmed the biological adequacy of this new feature set and its possible extrapolation to other non-ADNI MCI and AD populations Budelier and Bateman [2020]; Gunes et al. [2022]

Among the many scaling methods that were analyzed, section 2.3, Standard and Robust looked more promising. Only one candidate of the projection methods tried, section 2.4, NCA with "identity" initialization, was selected as the components ended with similar ranges of values. As per the number of projected components, 3 and 4 PCs were analyzed. As noted in [Cabrera-León et al., 2024b], there are 8 possible scenarios to be studied given the pair of values of these three preprocessing steps. However, in this thesis only two scenarios were preferred after analyzing the boxenplots: with AGE, Standard scaling and 3 PCs, and with AGE, Robust scaling and 4 PCs.

Table 4.15: Characteristics of the subjects in the CN-MCI-AD classification problem: a demographic feature, and the six attributes used by the model as input, sorted according to their FCBF score.

	AD	MCI	CN
Number of subjects	167	373	212
AGE: mean	75.55	74.96	75.96
AGE: StD	7.42	7.35	4.98
AGE: interval	[55.1 - 90.9]	[55.2 - 89.3]	[59.9 - 89.6]
AGE: *p*-value		0.6312 ns	
VENTRICLES: mean	50941.89	45719.62	35555.31
VENTRICLES: StD	26395.31	24555.92	20383.57
VENTRICLES: interval	[9166 - 147064]	[7801 - 145115]	[5834 - 118875]
VENTRICLES: *p*-value		3.5611e-08 ****	
ABETA: mean	627.94	725.24	1254.76
ABETA: StD	266.75	328.29	447.83
ABETA: interval	[265.6 - 2568.0]	[210.9 - 2809.0]	[200.0 - 3592.0]
ABETA: *p*-value		’1.7529e-61 ****	
FAQTOTAL: mean	12.86	3.78	0.14
FAQTOTAL: StD	6.75	4.39	0.61
FAQTOTAL: interval	[0 - 30]	[0 - 21]	[0 - 6]
FAQTOTAL: *p*-value		6.5382e-70 ****	
MMSCORE: mean	23.34	27.04	29.12
MMSCORE: StD	2.03	1.75	0.95
MMSCORE: interval	[18 - 27]	[24 - 30]	[26 - 30]
MMSCORE: *p*-value		7.0326e-75 ****	

Acronyms: ns (not significant), StD (Standard Deviation).

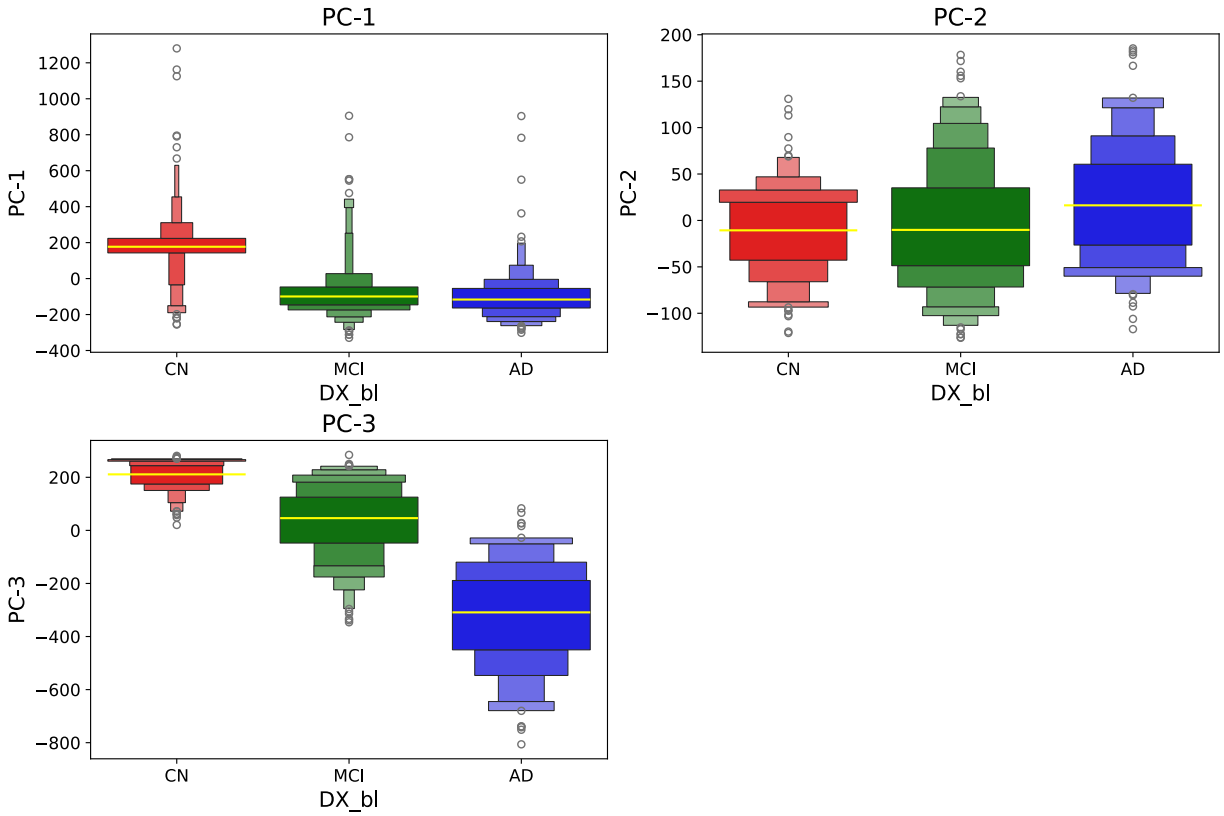


Figure 4.15: Boxenplot of the features (AGE included) for the CN-MCI-AD classification task: 3 projected components, Standard scaling, NCA with “identity” initialization.

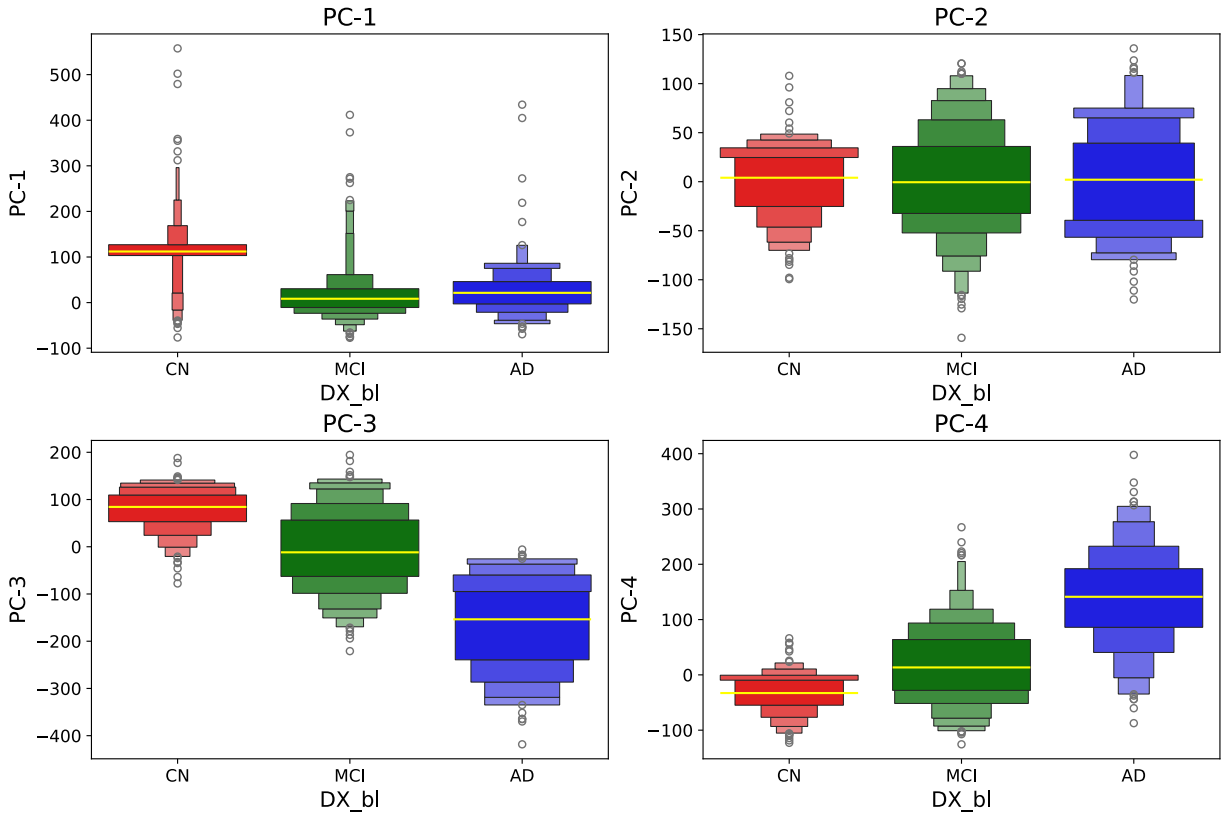


Figure 4.16: Boxenplot of the features (AGE included) for the CN-MCI-AD classification task: 4 projected components, Robust scaling, NCA with "identity" initialization.

Table 4.16: Performance results of the MyGNG in each of the two scenarios studied for the CN-MCI-AD classification task.

Scenario			Accu	Sens	Spec	Prec	AUC	CUI+	CUI-
With AGE	Scaling method	No. PCs							
Yes	Standard	3	0.83	0.89	0.78	0.8	0.83	0.71	0.69
Yes	Robust	4	0.86	0.89	0.79	0.82	0.83	0.73	0.72

Acronyms: Accu (accuracy), AUC (Area Under the ROC Curve), PC (principal component), Prec (precision), Sens (sensitivity), Spec (specificity).

4.5.2 Processing and results with the MyGNG

StratifiedKFold with 5 folds was used in the experiments with both models.

The values of the hyperparameters of the GNG that were considered appropriate for the MCI-AD classification task were used in the grid search as the initial ones, subsection 4.4.2: $max\_nodes = 100, a_{max} = 5, \lambda = 25, \varepsilon_b = 0.9, \varepsilon_n = 0.01, \beta = 0.7, d = 0.18, epochs = 7, \rho = 0.01, epochs\_perceptron = 50$. These initial values were then varied to find the optimal combination of the hyperparameters of the MyGNG. One was found of interest when 3 principal components (PCs) were used: $max\_nodes = 75, a_{max} = 5, \lambda = 50, \varepsilon_b = 0.65, \varepsilon_n = 0.015, \beta = 0.7, d = 0.8, epochs = 8, \rho = 0.01, epochs\_perceptron = 50$. A different one when there were 4 PCs: $max\_nodes = 75, a_{max} = 5, \lambda = 50, \varepsilon_b = 0.5, \varepsilon_n = 0.015, \beta = 0.6, d = 0.7, epochs = 8, \rho = 0.01, epochs\_perceptron = 50$. Adding AGE to the feature set was beneficial in both cases and the optimal configurations remained the same.

In Table 4.16 the performance results of the MyGNG in each of the two scenarios that were studied are shown.

4.5.3 Processing and results with the SupeRGNG

Compared to the previous binary problems and considering the complexity of this multiclass task, substantially more combinations of the hyperparameters were studied. For this reason, in Table 4.17 there only will be tabulated those sets of configurations where the optimal ones were found: "config1144" and "config848", respectively for each scenario, Table 4.18. The value of ι was calculated as before, as the 80% of the number of samples: 656.

The performance results of the best SupeRGNG configurations in each of the two scenarios selected for this multiclass task are depicted in Table 4.19.

4.6 Comparative studies of the Modular Hybrid Growing Neural Gas and the Supervised Reconfigurable Growing Neural Gas with other Machine Learning methods

A quantitative comparative study of SupeRGNG and MyGNG with ML and DL models was conducted.

Table 4.17: Sets of configurations that were tested where the optimal ones are found, one per scenario (CN-MCI-AD classification task). The number of configurations in each of these tests was: 1800 and 1600.

Scenario	Hyperparameter	Values
With AGE, Standard, 3PC	λ	[25, 50, 75, 100, 125, 150, 175, 200]
	a_{max}	[600, 700, 800, 900, 1000]
	max_nodes	[40, 60, 80, 120, 180, 280, 380, 480, 580]
	η	[100, 300, 500, 700, 900]
	ι	[656]
With AGE, Robust, 4PC	λ	[405, 455, 505, 555, 605]
	a_{max}	[405, 455, 505, 555, 605]
	max_nodes	[60, 80, 120, 180, 280, 380, 480, 580]
	η	[200, 300, 400, 500, 600, 700, 800, 900]
	ι	[656]

Table 4.18: Best configurations found for the CN-MCI-AD classification task.

Configuration	λ	a_{max}	max_nodes	η
config1144	150	600	120	900
config5459	505	555	120	200

Table 4.19: Performance results of the SupeRGNG in each of the two scenarios studied for the CN-MCI-AD classification task. First line: mean, Second line: CN-AD, Third line: CN-MCI, Fourth line: MCI-AD.

Scenario			Accu	Sens	Spec	Prec	AUC	CUI+	CUI-
With AGE	Scaling method	No. PCs							
Yes	Standard	3	0.88	0.88	0.88	0.88	0.95	0.77	0.77
			0.88	0.88	0.88	0.88	0.99	0.78	0.78
			0.88	0.88	0.88	0.88	0.97	0.77	0.77
			0.87	0.87	0.87	0.87	0.87	0.76	0.76
Yes	Robust	4	0.89	0.89	0.89	0.89	0.94	0.79	0.79
			0.94	0.94	0.94	0.94	1	0.89	0.89
			0.83	0.83	0.83	0.83	0.91	0.69	0.69
			0.89	0.89	0.89	0.89	0.92	0.8	0.8

Acronyms: Accu (accuracy), AUC (Area Under the Curve), CUI (Clinical Utility Index), PC (principal component), Prec (precision), Sens (sensitivity), Spec (specificity).

4.6.1 Quantitative comparative study

Both shallow neural networks and non-neural ML approaches were selected for this comparative study.

Shallow neural networks refers to the most paradigmatic supervised neural architecture, the Multilayer Perceptron (MLP) with Backpropagation Network (BPN) learning. This will allow us to conduct a comparative study with deep architectures whose underlying architectural framework is also BPN, i.e., Convolutional Neural Networks (CNNs).

On the one hand, a quantitative comparative study between our methods and other popular supervised ML models was carried out, with the same models as in [Cabrera-León et al., 2024b]: a Decision Tree (DT) (flowchart-like structure; easier to interpret than Artificial Neural Networks (ANNs)) [Kotsiantis, 2013], a Naïve Bayes (NB) classifier (based on applying the Bayes' theorem and assuming that the features are strongly independent given class) [Rish, 2001], a Random Forest (RF) (an ensemble of DTs, each trained with a random subset of features; the class that is returned is the one chosen by most DTs) [Kotsiantis, 2013], a Support Vector Machine (SVM) (builds an hyperplane usually in a high-dimensional space to separate classes; using certain kernel functions allow separation of non-linear data) [Cortes and Vapnik, 1995], and a MLP (a feedforward ANN able to separate non-linear data, unlike the single-layer perceptron) [Haykin, 2009]. Another shallow architecture that we have used in this comparative study is a hybrid neural network named Counterpropagation Network (CPN) [Hecht-Nielsen, 1987; Freeman and Skapura, 1991], such as our proposed architectures.

The previous ML classifiers were implemented with "scikit-learn" [Pedregosa et al., 2011] and "Keras" [Chollet et al., 2015], two very popular ML and DL Python modules, respectively. Best results were yielded by models with these combinations of hyperparameters. For DT, Pruning=at least 2 instances in leaves; at least 5 instances in internal nodes; maximum depth=100; Splitting: Stop splitting when majority reaches 95% (classification only); Binary trees: Yes For NB, scikit-learn's default values. For RF, Number of trees=10; Maximal number of considered features=unlimited; Replicable training=No; Maximal tree depth=unlimited; Stop splitting nodes with maximum instances=5. For SVM, C=1.0, ε=0.1; Kernel: RBF; exp(-auto$|x-y|^2$); Numerical tolerance: 0.001; Iteration limit: 100. For MLP, hidden neurons=(16, 8); activation function="relu"; solver="rmsprop".

Additionally, a comparison of the SupeRGNG with a parallel CNN called ParallelNet [Sharma et al., 2019] and some variants of the Residual Network (ResNet) CNN [He et al., 2016] was carried out. The variants selected were ResNet-18, ResNet-50 and ResNet-101. The main differences of these three variants of the ResNet are the number of layers (18, 50 and 101, respectively), the number of parameters (11.7, 25.6 and 44.6, respectively, in million), and the memory used for the parameters (45, 98 and 171, respectively, in MB). In common they have the size of the input image: 224 x 224.

Firstly, the reason for choosing the DeepInsight architecture, in which the ParallelNet is proposed as one of its classification models, is its parallel nature and the good results it provides in [Sharma et al., 2019] for different types of input data, especially in the accuracy metric, with outstanding values compared to those obtained with other classifiers (*i.e.* RF). On the other hand, the choice of the other three pre-trained ResNet networks is mainly due to their good classification accuracy-Graphics Processing Unit (GPU) computing power balance on the ImageNet validation set, as depicted in Figure 4.17, where the size of the circle is proportional to the number of parameters. It should be noted that a high accuracy

on ImageNet does not necessarily mean that the same will also occur with other tasks or datasets.

As we are addressing these diagnostic problems with non-neuroimaging biomarkers and considering the characteristics of the CNNs, which are essentially designed for a two-dimensional input space, it was required that our feature vectors were converted to that type of data. More specifically, the conversion to 2D images was performed via the DeepInsight pipeline [Sharma et al., 2019] so that these deep architectures were able to work with our feature vectors. Both DeepInsight and ParallelNet were already described in subsubsection 2.1.3.1.

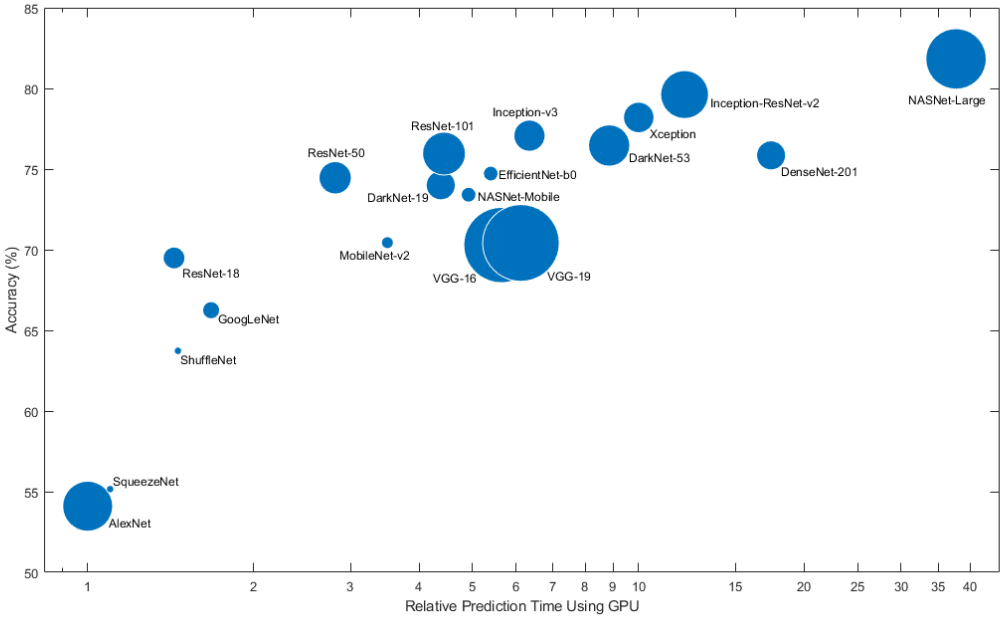


Figure 4.17: Accuracy vs relative prediction time using GPU of several Deep Neural Networks on the ImageNet validation set. Source: MathWorks [2025]

4.6.1.1 CN-AD

The best SupeRGNG configurations yielded the best possible value with all performance metrics, Table 4.20. Although results are quite good and similar in all classifiers, SupeRGNG and CPN are the only ones with perfect performance with all performance metrics, followed by SVM. Despite data being linearly separable in all cases, MLP was unable to obtain the best possible results.

Deep Neural Networks (DNNs) yielded very good but not perfect results in all these preprocessed datasets but the non-invasive one. Especially low were the values of sensitivity and CUI+ of the ParallelNet, which might be explained by their need for more training data due to being a more complex network than the other CNNs.

Most of the architectures used yielded perfect results, being the SupeRGNG at the same level as ensembles (RF) and hybrid models such as CPN. This confirms that ontogenetic architectures as the one proposed are optimal computational solutions for distinguishing subjects with AD from those cognitively healthy. We can conclude that for this type of

Table 4.20: Comparison of the SupeRGNG and several popular neural and non-neural ML methods (CN-AD classification task).

Dataset	Method	Accu	Sens	Spec	AUC	CUI+	CUI-
"Invasive", with AGE, 2PC, MaxAbsScaler, NCA identity	DT	0.99	0.99	0.99	0.99	0.98	0.98
	RF	0.99	0.99	0.99	0.99	0.98	0.98
	NB	0.99	0.99	0.99	0.99	0.99	0.99
	SVM	1	1	1	1	1	1
	MLP	0.97	0.95	0.98	0.97	0.96	0.95
	CPN	1	1	1	1	1	1
	ResNet-18	0.99	1	0.99	1	0.98	0.99
	ResNet-50	0.99	1	0.99	1	0.98	0.99
	ParallelNet	0.99	0.99	0.99	0.99	0.98	0.99
	SupeRGNG	1	1	1	1	1	1
"Invasive", without AGE, 2PC, MaxAbsScaler, NCA identity	DT	0.99	0.99	0.99	1	0.98	0.98
	RF	1	1	1	1	0.99	0.99
	NB	1	1	1	1	1	1
	SVM	1	1	1	1	1	1
	MLP	0.94	0.88	0.99	0.93	0.97	0.90
	CPN	1	1	1	1	1	1
	ResNet-18	0.99	1	0.97	1	0.98	0.99
	ResNet-50	0.99	0.99	0.99	1	0.98	0.99
	ParallelNet	0.99	0.99	0.99	0.99	0.98	0.98
	SupeRGNG	1	1	1	1	1	1
"Invasive", with AGE, 2PC, RobustScaler, NCA identity	DT	0.99	0.99	0.99	0.99	0.98	0.98
	RF	0.99	0.99	0.99	0.99	0.98	0.98
	NB	0.99	0.99	0.99	0.99	0.99	0.99
	SVM	1	1	1	1	0.99	0.99
	MLP	1	1	1	1	0.99	0.99
	CPN	1	1	1	1	1	1
	ResNet-18	0.99	1	0.99	1	0.99	0.99
	ResNet-50	0.99	1	0.99	1	0.99	0.99
	ParallelNet	0.99	1	0.99	0.99	0.99	0.99
	SupeRGNG	1	1	1	1	1	1
"Non-invasive", with AGE, 2PC, RobustScaler, NCA identity	DT	1	1	1	1	0.99	0.99
	RF	0.99	0.99	0.99	1	0.99	0.99
	NB	0.99	0.99	0.99	0.99	0.98	0.98
	SVM	1	1	1	1	1	1
	MLP	0.98	0.99	0.96	0.98	0.91	0.96
	CPN	1	1	1	1	1	1
	ResNet-18	0.84	0.73	0.93	0.90	0.64	0.76
	ResNet-50	0.85	0.76	0.92	0.90	0.67	0.77
	ParallelNet	0.66	0.21	1	0.71	0.27	0.63
	SupeRGNG	1	1	1	1	1	1

Acronyms: Accu (accuracy), AUC (Area Under the Curve), CUI (Clinical Utility Index), DT (Decision Tree), MLP (Multilayer Perceptron), MyGNG (Modular Hybrid Growing Neural Gas), NB (Naïve Bayes), RF (Random Forest), Sens (sensitivity), Spec (specificity), SupeRGNG (Supervised Reconfigurable Growing Neural Gas), SVM (Support Vector Machine).

diagnostic problem it is sufficient a monolithic and shallow ANN, which are less complex and computationally intensive than ensembles and deep architectures, respectively.

On the other hand, as there are four set of features that were considered good candidates by the methodology used, it is generally advisable to propose to clinicians a single candidate of feature set to be used as the optimal clinical criteria for the CN-AD classification task. After the analysis done, we can conclude that the "non-invasive" set of features (that is, FAQREM, FAQFORM, FAQTRAVL, MMSCORE, MMFLAGDL and AGE) is preferable due to the features being non-invasive, fast to obtain, easily repeatable and non-expensive. These characteristics make these features especially interesting for primary care.

4.6.1.2 CN-MCI

Compared to other ML solutions, the SupeRGNG behaved similarly, yielding better results in most metrics but AUC, Table 4.21. Comparing these values with the also hybrid architecture CPN, the difference is not statistically significant. Values of sensitivity with CPN were the best, albeit it had much lower specificity and CUI-. The better CUI+ values of SupeRGNG may indicate that this network is a good option for translational medicine.

Most of the architectures compared yielded good results, being the SupeRGNG the one with good yet balanced values in the pairs sensitivity-specificity and CUI+-CUI-. Sensitivity, AUC and CUI+ in CPN and DNNs were higher or slightly higher than those of the SupeRGNG. However, their values of specificity and CUI- were quite lower, in the ParallelNet case up to 0.16 and 0.19 less, respectively. This confirms that ontogenetic architectures as the one proposed are good computational solutions for distinguishing subjects with MCI from those cognitively normal.

On the other hand, considering that only one set of features have been proposed and good results have been achieved, it can be concluded that the proposed set of features (that is, ABETA, AGE, MMBALLDL, MMDAY, NPIL and MMYEAR) can be recommended to clinicians as the optimal clinical criteria for the CN-MCI classification task. As it includes the ABETA, which requires performing an invasive CSF extraction, it is not as recommended for primary care.

4.6.1.3 MCI-AD

A comparison with other ML algorithms is shown in Table 4.22. Same best results were achieved in both versions of the SupeRGNG but the newer version is much faster thanks to its new capabilities: its training finished in 529 epochs instead of the originally used 1000 epochs.

The SupeRGNG overtook the other models by a big margin, achieving the same performance as the the previous variant [Cabrera-León et al., 2023], followed by the MyGNG. The MLP and the DNNs behaved not as good, especially the sensitivity of the latter. This confirms that ontogenetic architectures as the one proposed are good computational solutions for distinguishing subjects with MCI from with AD.

On the other hand, considering that only one set of features have been proposed and very good results have been yielded, it can be concluded that the proposed set of features (that is, MMSCORE, MMDATE, MMBALLDL, ADAS_Q7, MMYEAR and FAQSHOP) can be recommended to practitioners as the optimal clinical criteria for the MCI-AD classification task. Another advantage of this feature set is that only inexpensive, non-invasive and easy to obtain features are included, making it interesting to be applied even in primary care.

Table 4.21: Comparison of the SupeRGNG and several popular neural and non-neural ML methods (CN-MCI classification task).

Dataset	Method	Accu	Sens	Spec	AUC	CUI+	CUI-
6 projection components, RobustScaler, t-SNE with perplexity=100	DT	0.74	0.74	0.74	0.73	0.54	0.54
	NB	0.79	0.79	0.79	0.75	0.62	0.62
	RF	0.80	0.80	0.80	0.74	0.65	0.65
	SVM	0.84	0.84	0.84	0.79	0.71	0.71
	MLP	0.79	0.79	0.79	0.77	0.63	0.63
	CPN	0.86	0.91	0.76	0.91	0.80	0.63
	ResNet-18	0.83	0.88	0.74	0.87	0.75	0.58
	ResNet-50	0.83	0.89	0.74	0.88	0.76	0.59
	ResNet-101	0.83	0.89	0.72	0.87	0.75	0.57
	ParallelNet	0.82	0.89	0.70	0.87	0.75	0.55
	SupeRGNG	0.86	0.86	0.86	0.88	0.74	0.74

Acronyms: Accu (accuracy), AUC (Area Under the Curve), CUI (Clinical Utility Index), DT (Decision Tree), MLP (Multilayer Perceptron), MyGNG (Modular Hybrid Growing Neural Gas), NB (Naïve Bayes), RF (Random Forest), Sens (sensitivity), Spec (specificity), SupeRGNG (Supervised Reconfigurable Growing Neural Gas), SVM (Support Vector Machine).

Table 4.22: Comparison of the SupeRGNG, the MyGNG and several popular neural and non-neural ML methods (MCI-AD classification task).

Dataset	Method	Accu	Sens	Spec	AUC	CUI+	CUI-
2PC, Unscaled, PCA	DT	0.91	0.91	0.91	0.92	0.83	0.83
	NB	0.91	0.91	0.91	0.92	0.83	0.83
	RF	0.90	0.90	0.90	0.92	0.81	0.81
	SVM	0.94	0.94	0.94	0.94	0.88	0.88
	MLP	0.89	0.89	0.89	0.90	0.79	0.79
	CPN	0.87	0.87	0.87	0.96	0.64	0.82
	ResNet-18	0.90	0.77	0.96	0.94	0.68	0.86
	ResNet-50	0.89	0.80	0.93	0.94	0.66	0.85
	ResNet-101	0.90	0.73	0.97	0.95	0.67	0.87
	ParallelNet	0.89	0.73	0.96	0.95	0.65	0.85
	MyGNG	0.93	0.97	0.83	0.96	N/A	N/A
	"Static" SupeRGNG	0.98	0.98	0.98	0.97	N/A	N/A
	SupeRGNG	0.98	0.98	0.98	0.97	0.96	0.96

Acronyms: Accu (accuracy), AUC (Area Under the Curve), CUI (Clinical Utility Index), DT (Decision Tree), MLP (Multilayer Perceptron), MyGNG (Modular Hybrid Growing Neural Gas), N/A (Not Available), NB (Naïve Bayes), RF (Random Forest), Sens (sensitivity), Spec (specificity), SupeRGNG (Supervised Reconfigurable Growing Neural Gas), SVM (Support Vector Machine).

4.6.1.4 CN-MCI-AD

In the 2 scenarios that were analyzed the SupeRGNG has behaved quite good, outperforming the ML approaches. Although its values of AUC were on par with or slightly superior than the those from the CPN and the four DL approaches, the values of the rest of metrics were not. On the contrary, MyGNG was outperformed by the DNNs and CPN but yielded performance results at the same level as those from the most popular ML classifiers.

Additionally, the proposed set of features (that is, VENTRICLES, ABETA, FAQTOTAL, MMSCORE, and AGE) has demonstrated to be a good set of features for the CN-MCI-AD classification problem. Therefore, it is suitable for presentation to physicians for diagnostic use, although the presence of ABETA makes this set not the most recommended for primary care.

In Table 4.23 it is shown a quantitative comparative for the multiclass task.

4.7 Towards an e-Health solution for the diagnosis of Alzheimer's Disease

Geriatric and neurological assessments should be diagnostic instruments to analyze biological, mental, functional and psychosocial disorders of the older people, to achieve an adequate treatment plan and an optimal management of the sociosanitary resources. Primary Care (PC) centers lack specialists in AD and other related pathologies at their disposal. Hence, used procedures, techniques and diagnostic tools are not especially conditioned for the care needs to perform an accurate diagnosis of patient's disease (type of dementia) neither to provide the disease stage. In the particular case of Specialized Care (SC) centers, diversity and variability of procedures and diagnostic methods employed by clinicians to perform patient-evaluation and assessment are of remarkable importance. Continuity in patient care, a detailed monitoring of any variable associated to the disorder, fomenting collaborative work between specialists, and giving the patient's familiar environment the importance it deserves become a must to build a solid knowledge base on MCI, AD and other dementia.

To achieve this goal, we propose an e-Health solution that involves a virtual clinical environment, delivered through a web application. This virtual clinical station is user-friendly for both primary and specialized care practitioners. It can use the intelligent diagnostic support systems for dementia in general and AD in particular, developed in this thesis. These new tools, also considering the intelligent diagnostic support systems based on the new neural architectures developed in this thesis, should drive better use of healthcare system resources, improving the relationship between healthcare quality, efficiency, and budget.

In this chapter we will describe an e-Health solution for the diagnosis of AD and MCI that is able to integrate, among others, any of the ANNs implemented, which have been described in previous chapters.

4.7.1 EDEVITALZH

Clinical Virtual Environment to aid diagnosis and prognosis of Alzheimer's Disease and other dementias (EDEVITALZH) is a Personalized, Predictive, Preventive, and Participatory Healthcare Delivery System (4P-HCDS) that follows the philosophy of a Clinical Workstation (CW) [Pérez Del Pino, 2015; Pérez-del-Pino et al., 2014; Suárez Araujo et al., 2012; Suárez-

Table 4.23: Comparison of the SupeRGNG, the MyGNG and several popular neural and non-neural ML methods (CN-MCI-AD classification task).

Dataset	Method	Accu	Sens	Spec	AUC	CUI+	CUI-
With AGE, 3PC, StandardScaler, NCA identity	DT	0.86	0.86	0.86	0.90	0.73	0.73
	NB	0.87	0.87	0.87	0.88	0.76	0.76
	RF	0.88	0.88	0.88	0.89	0.77	0.77
	SVM	0.88	0.88	0.88	0.90	0.77	0.77
	MLP	0.71	0.71	0.71	0.75	0.51	0.51
	CPN	0.93	0.93	0.93	0.95	0.88	0.88
	ResNet-18	0.91	0.91	0.91	0.94	0.85	0.85
	ResNet-50	0.91	0.91	0.91	0.94	0.85	0.85
	ResNet-101	0.91	0.91	0.91	0.93	0.85	0.85
	ParallelNet	0.92	0.92	0.92	0.94	0.86	0.86
	MyGNG	0.83	0.89	0.78	0.83	N/A	N/A
	SupeRGNG	0.88	0.88	0.88	0.95	0.77	0.77
With AGE, 4PC, RobustScaler, NCA identity	DT	0.86	0.86	0.86	0.86	0.74	0.74
	NB	0.86	0.86	0.86	0.89	0.74	0.74
	RF	0.88	0.88	0.88	0.91	0.77	0.77
	SVM	0.87	0.87	0.87	0.90	0.76	0.76
	MLP	0.70	0.70	0.70	0.53	0.50	0.50
	CPN	0.93	0.93	0.93	0.94	0.88	0.88
	ResNet-18	0.92	0.92	0.92	0.94	0.86	0.86
	ResNet-50	0.91	0.91	0.91	0.93	0.85	0.85
	ResNet-101	0.91	0.91	0.91	0.94	0.84	0.84
	ParallelNet	0.91	0.91	0.91	0.94	0.85	0.85
	MyGNG	0.86	0.89	0.79	0.83	N/A	N/A
	SupeRGNG	0.89	0.89	0.89	0.94	0.79	0.79

Acronyms: Accu (accuracy), AUC (Area Under the Curve), CUI (Clinical Utility Index), DT (Decision Tree), MLP (Multilayer Perceptron), MyGNG (Modular Hybrid Growing Neural Gas), NB (Naïve Bayes), RF (Random Forest), Sens (sensitivity), Spec (specificity), SupeRGNG (Supervised Reconfigurable Growing Neural Gas), SVM (Support Vector Machine).

Araujo et al., 2004]. EDEVITALZH Environment provides both an Electronic Medical Records Database (EMRDB) and the digital implementation of the Global Clinical Protocol for Dementia (GCPD). A GCPD includes a detailed set of clinical procedures, forms, tests and diagnostic criteria, which have been validated by Geriatrics and Neurology experts.

EDEVITALZH Environment is based on a scalable, secure, robust and fault-tolerant technological architecture that provides users anywhere the capability to connect at any time. The principal set of applications, systems, tools and EDEVITALZH for information management, exploitation and integration inside the EDEVITALZH Environment is known as EDEVITALZH Core (EDV-Core), Figure 4.18 EDEVITALZH Environment is considered to be an iCWs thanks to the Intelligent Clinical Wizard and Assistant (ICWA), which comprises the integration of EDV-Core User Interface (UI) applications (ICWA-App) with our intelligent systems for diagnosis based on MyGNG (MISD) and SupeRGNG (SISD), providing the environment of Computational Intelligence. This way, EDV-Core has embedded the required business logic to manage some Intelligent Systems for Diagnosiss-Intelligent Decision Support Tools (ISDs-IDSTs) working at the same time on different requests according to the clinical criteria selected, Figure 4.18. Moreover, EDEVITALZH incorporates the Electronic Medical Interconsultation (EMI), whose aim it to entice collaboration between clinicians via an internal messaging. Thank to this, physicians can ask for colleagues' opinions, interact in discussions and aid in the monitorization of colleagues' patients. EMI is a powerful and valuable tool for clinicians at primary care and other welfare centers where no specialist physicians in MCI, AD and other dementia are easily available. This gives the primary care clinicians the possibility to assist patients with these neuropathologies with a higher level of quality, accuracy and efficiency in the diagnosis and prognosis processes.

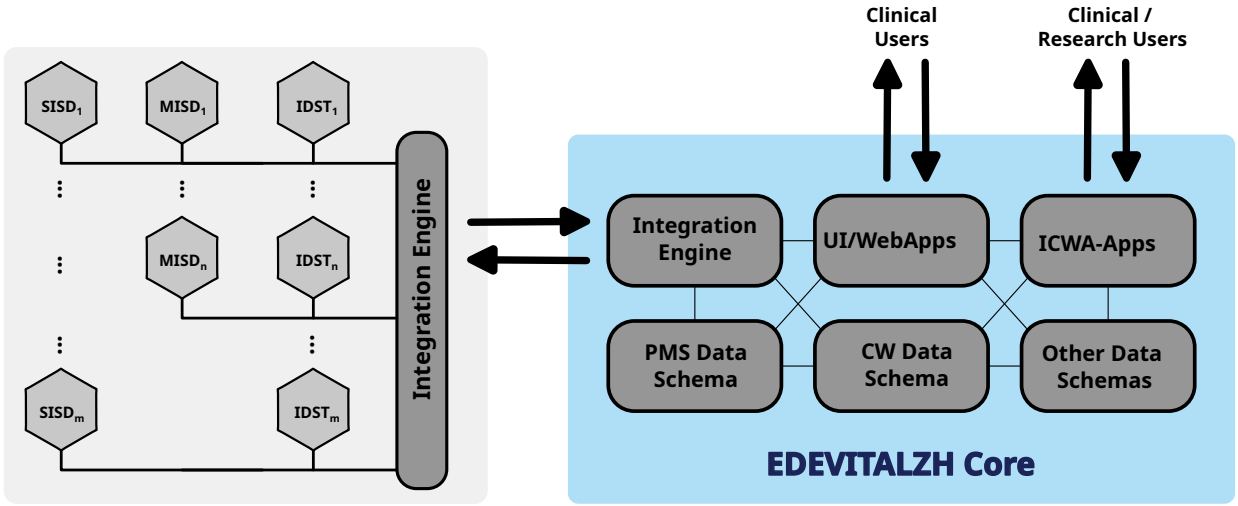


Figure 4.18: High level architectural diagram of the EDEVITALZH environment.

4.7.2 EDEVITALZH systems level

This virtual clinical environment has been built over a EDEVITALZH Systems Level (SL) base [Pérez Del Pino, 2015]. As a multi-user environment, it allows many clinicians to request at the same time aid in the diagnosis process. It provides services of Patient Management Software (PMS) to manage patients' medical records in a computerized way. This is directly related to database systems and storage.

Also, it provides services of CW via the set of clinical procedures, forms, tests and diagnostic criteria used by physicians to diagnose the patient. This has direct implications with UI Applications, database systems and storage.

Additionally, it provides the required technological mechanisms to integrate and communicate EDV-Core with Intelligent Systems for Diagnosis (ISD) and Intelligent Decision Support Tool (IDST), enabling EDEVITALZH to become an intelligent Clinical Workstation (iCW). This is directly related to processing systems, database systems and communication procedures.

The Systems Level Core (SL-Core) is the hardware basis to support the EDV-Core, Figure 4.18. It comprises the group of systems that manage databases, user applications and internal and external communications of the EDEVITALZH environment.

The Decision Support Tools Execution Module (ST-Exec) consists of the group of systems which manage processing load and storage via execution policies regarding ISDs-IDSTs workload. Each ISD-IDST component must be developed as an autonomous entity because every part of the environment has to be developed as self-reliant. Each ISD-IDST component receives input data under a user request and returns results according to that data. This way, one or several ISDs-IDSTs are first selected upon a selection procedure and then asked to generate certain results according to the provided inputs every time a user demands a Decision Supported Operation.

Taking into account that EDEVITALZH is considered a multi-user platform, the environment is able to handle several simultaneous decision supported operations. This means that EDEVITALZH needs to be capable of planning and managing its global processing resources to handle all user decision supported operations efficiently, because it has execution policies to queue, prioritize and execute decision supported operations.

The Database Level (DBL) comprises the EMRDB Infrastructure that stores, manages and relates the social, administrative and clinical data of EDEVITALZH subjects.

EMRDB is the resulting database structure of representing the EDEVITALZH Data Model (DM), which models the patients' electronic medical records and their detailed set of clinical procedures, forms, tests and diagnostic criteria defined by GCPD Pérez-del-Pino et al. [2014]. Validated by medical experts in Geriatrics and Neurology, GCPD reflects schematically specific and relevant data focused on the diagnosis of Alzheimer's Disease and other dementia, correlating clinical and therapeutical parameters at the same time.

On the other hand, the presentation level regards the user interfaces, EDEVITALZH PT-UI, which consists of the web applications that make up the iCW UI. It guides clinical users in their workflow. Two main web applications have been implemented, based on the GCPD. On the one hand, the PC web application (PCapp) is focused on PC clinicians and handles a reduced set of the GCPD that groups the needed medical criteria to perform a basic diagnosis at any not-specialized center. On the other hand, the SC web application (SCapp) is focused on SC clinicians and manages the complete version of GCPD to perform a more accurate differential diagnosis and prognosis. Also, it allows carrying out evolution studies of patients and their pathologies.

With the EMIs capabilities it can be performed shared consultation by several clinicians, at primary care or specialized care, or even a college can be asked for opinion or specialized assessment. Furthermore, since both applications, PCapp and SCapp, share the database environment, doctors can refer their patients to any other specialist physician in a simple and straightforward way.

4.7.2.1 Integration of Intelligent Systems for Diagnosis Mechanisms

EDEVITALZH integration mechanisms are the logical component of the Integration Engine, Figure 4.18. It specifies processes related to the extraction, deidentification (if required), encapsulation and transfer of information between ICWA-Apps, hence the EDV-Core, and ISDs-IDSTs, Figure 4.19.

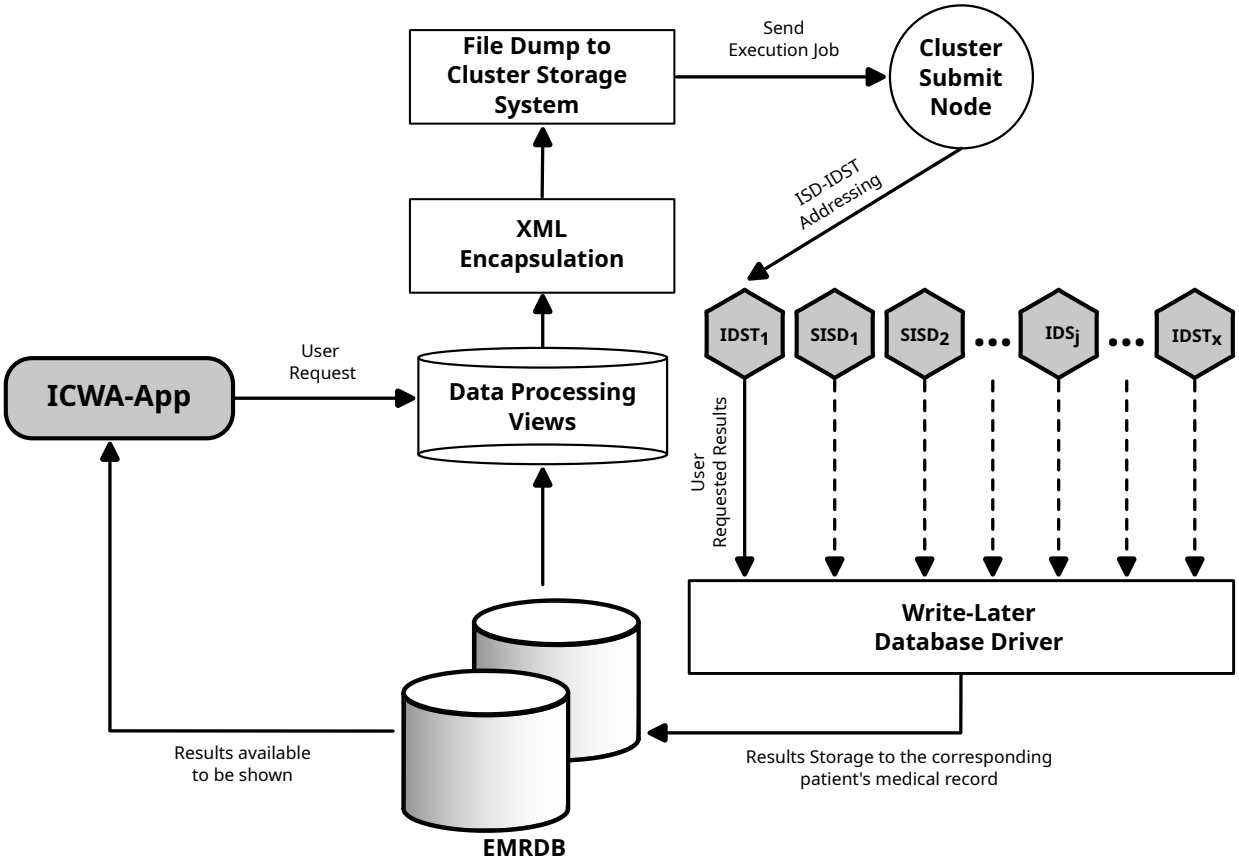


Figure 4.19: Mechanism to integrate Intelligent Systems for Diagnosis with the corresponding ICWAs Application to provide aid in decision making tasks. The flow diagram depicts the operative since the clinical user requests helps in decision making until results are available.

EDEVITALZH processes for data extraction are implemented as SQL views, which are custom developed to extract a specific data set, according to the input needs of the SISDT or SISDs-IDSTs that are going to be requested. Later, the dataset will be encapsulated as an XML structure and saved into a storage system. If there are any restriction (such as special parametrization of any SISD/MISD), every XML files contain an SISD/MISD ID parameter section. This section will allow the different intelligent systems based on our SupeRGNG or MyGNG to identify when they should handle a certain request.

Once the file is generated, the corresponding ICWA-App is responsible for the execution of the set of processes corresponding to each requested SISD. As soon as results are ready, they are written to the corresponding patient's medical record in the EMRDB, Figure 4.19.

This integration protocol has several advantages to the way EDEVITALZH functions: it offers the possibility to request a certain support for decision making, while the user is performing any other task without interfering with their workflow. It is possible to manage requests asynchronously, and results will be notified to EMRDB when available.

Regarding security, an intrusion detection system based on Artificial Neural Networks is included [Pérez del Pino et al., 2012], which is able to detect any type of known and unknown attacks [Suárez-Araujo et al., 2004].

Considering all the above, the intelligent systems developed to comply with the first and second goals of this PhD thesis, in the Abstract, can be used as the intelligent detection engines in a mono- or multi-scheme in the diagnosis module of EDEVITALZH, Figure 4.20. This way, it provides a complete e-Health solution in the fields of Primary Care, Neurology and Gerontology, which is able to deal with all the disease stages found in the AD continuum.

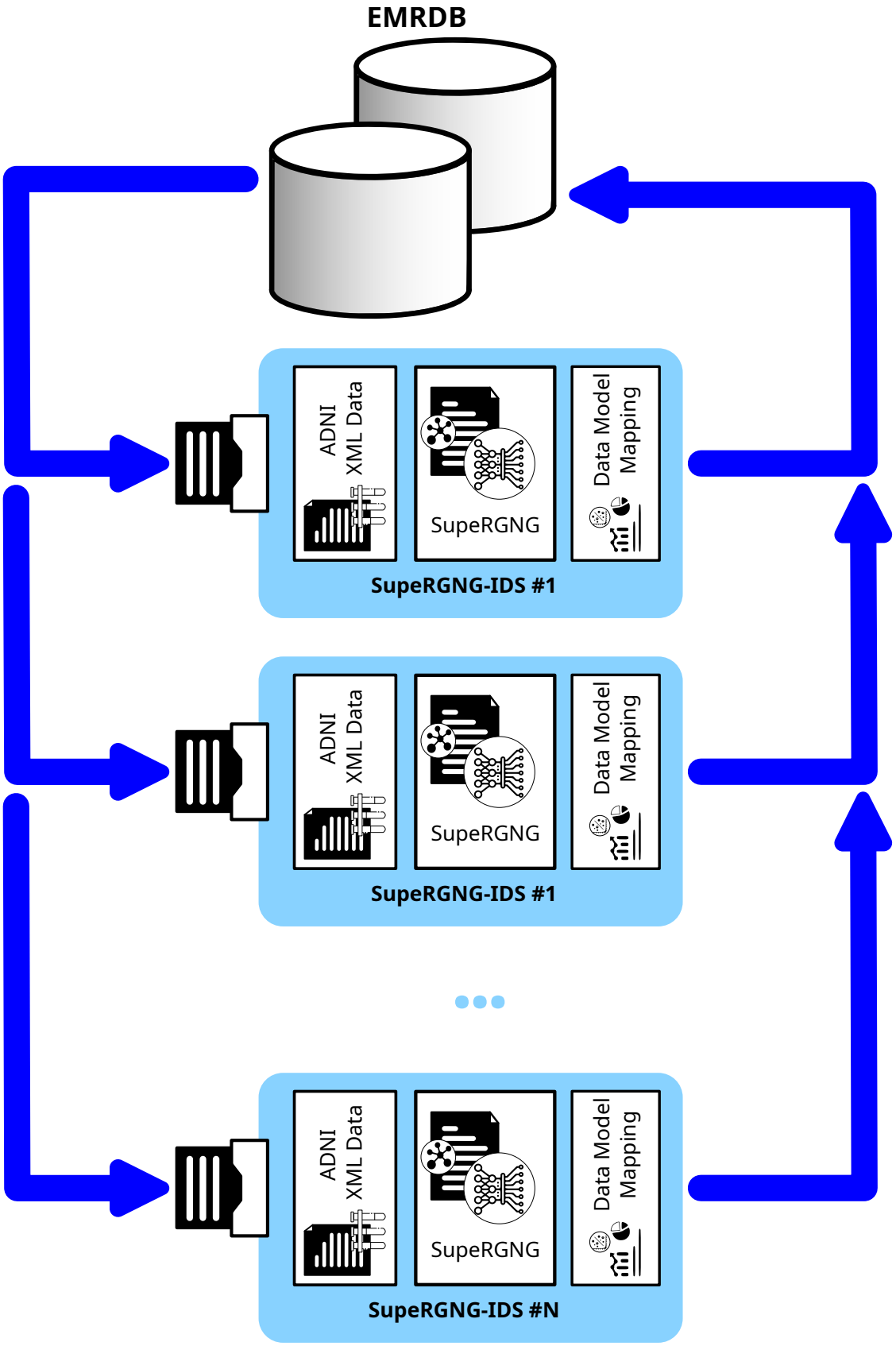


Figure 4.20: Example of a SupeRGNG-based multi-scheme working as intelligent detection engines in the diagnosis module of EDEVITALZH.

Chapter 5

Conclusions and Future Works

5.1 Conclusions

It can be concluded, overall, that the main contributions of this doctoral thesis are its own objectives, which have been completely covered. In the next paragraphs the conclusions and main contributions of this PhD thesis are indicated:

1. The ability of ontogenic architectures, such as the Supervised Reconfigurable Growing Neural Gas (SupeRGNG), to address complex real-world problems such as the differential diagnosis of Alzheimer's Disease (AD) in its early stages, compared to deep architectures, neural ensembles, and other traditional Machine Learning (ML) approaches.

2. The possibility of obtaining highly accurate diagnoses using non-invasive multimodal clinical criteria appropriate for primary care.

3. The possibility of achieving a universal diagnosis in both primary and specialized care (neurology and geriatrics).

4. A dataset has been built by integrating longitudinal data from different modalities such as neuropsychological tests, demographic data, biospecimen, genetics, quantitative neuroimaging, and Cerebrospinal Fluid. These data were extracted from multiple files of the Alzheimer's Disease Neuroimaging Initiative (ADNI) database, and it includes data from Cognitively Normal (CN), Mild Cognitive Impairment (MCI) and AD subjects.

5. Two intelligent diagnostic systems have been put forward for the diagnosis of all the disease stages of the AD continuum, which are based on two novel neural architectures. Also, it is demonstrated that Artificial Neural Networks (ANNs) are good approaches for complex problems, and those that were developed, also with unbalanced datasets, a characteristic frequent in medicine ones.

6. A new hybrid ontogenetic ANN, the Modular Hybrid Growing Neural Gas (MyGNG), has been proposed. It comprises two modules: a Growing Neural Gas (GNG), unsupervised, for clustering the input data followed by a mono-layer perceptron, supervised, to improve the clustering made.

7. A novel supervised, ontogenic ANN, the SupeRGNG, has been developed. Its main characteristics are the dynamic disconnection and reconnection of the clusters, which improves the clustering performed by the GNG on which it is based. These dynamic procedures allow the user to choose when to initiate them and how frequently to run them. As other ANNs can also do, an early stopping mechanism was included in order to shorten training time. The SupeRGNG reliably works with extremely unbalanced datasets (up to 90%) and with non-Gaussian and non-linearly separable datasets. Furthermore, it is capable of producing very good performance results with heavily class overlapped datasets.

8. Performance results with both MyGNG and SupeRGNG in several binary classification problems related to the diagnosis of the different stages of the AD continuum — CN vs MCI, CN vs AD and MCI vs AD — were outstanding, especially the SupeRGNG, which was able to outperform the rest of ML classifiers, including a parallel Convolutional Neural Network (CNN) and several variants of Residual Network (ResNet).

9. Performance results with MyGNG in the CN vs MCI vs AD multiclass problem were on par with the popular ML classifiers.

10. Performance results with SupeRGNG in the CN vs MCI vs AD classification task were on par with the ResNets and the parallel CNN, and greater than the other popular ML models.

11. For each binary and multiclass classification task that clinicians and researchers consider of interest, the minimal and optimal set of diagnostic criteria was obtained, which can be considered sufficient and appropriate for the differential diagnosis of AD, MCI and healthy aging brains.

12. Both neural computation methods developed can be integrated into an e-Health solution as intelligent systems to aid in the diagnosis of AD. They can be used in almost any e-Health solution. In fact, they can be used as the detection engines in the diagnosis module of Clinical Virtual Environment to aid diagnosis and prognosis of Alzheimer's Disease and other dementias (EDEVITALZH). This way, they can provide a complete e-Health solution in the fields of Primary Care, Neurology and Gerontology, which is able to deal with all the disease stages found in the AD continuum. Moreover, EDEVITALZH is securely accessible with inexpensive devices anytime and from anywhere, allowing different clinicians and researchers to work collaboratively and exchange information. All of this indicates that our proposal is capable of providing universal cortical dementia diagnosis.

5.2 Future works

The work presented in this doctoral thesis has opened several lines of research and some future work may be derived:

- Characterization of AD by establishing its Biological-Cognitive Profile.
- Study the evolution of AD by determining an optimal classification of the degree of severity of this neuropathology.

- Development of a prototype of EDEVITALZH that makes use of both intelligent systems developed as its detection engines.
- Creation of a new multisite dataset that integrates data from several non-private databases other than ADNI, such as Open Access Series of Imaging Studies (OASIS) and Sydney Memory and Ageing Study (MAS). Additionally, use data from both public and private databases that include all possible biomarkers that have been used to date for the diagnosis and prognosis of AD.
- Development of intelligent systems, based on the new proposed neural architectures, to aid in the prognosis of neurodegenerative pathologies studied in the thesis.
- Development of new neural computation methods, possibly as improvements of those developed in this PhD thesis, able to work with non-quantitative neuroimaging data, presumably also as a multimodal solution.
- Further studies on other data preprocessing techniques, especially regarding feature selection and data projection.

Appendix A

List of acronyms

4P-HCDS	Personalized, Predictive, Preventive, and Participatory Healthcare Delivery System
aMCI	amnestic Mild Cognitive Impairment
AA	Alzheimer's Association
Aβ	Amyloid beta
AD	Alzheimer's Disease
ADAS	Alzheimer's Disease Assessment Scale
ADNI	Alzheimer's Disease Neuroimaging Initiative
ADRDA	Alzheimer's Disease and Related Disorders Association
AE	Autoencoder
AI	Artificial Intelligence
AIBL	Australian Imaging, Biomarker & Lifestyle Flagship Study of Ageing
ANN	Artificial Neural Network
APA	American Psychiatric Association
APOE	Apolipoprotein E
APP	Amyloid Precursor Protein
ASL	Arterial Spin Labeling
ART	Adaptive Resonance Theory
AUC	Area Under the Curve
AV-45	Florbetapir F 18 amyloid
BMI	Body Mass Index
BMU	Best Matching Unit
BNN	Biological Neural Network
BPN	Backpropagation Network
CAD	Computer-Aided Diagnosis
CCI	Cognitive Change Index

CDR-SB	Clinical Dementia Rating Scale Sum of Boxes
CDR	Clinical Dementia Rating
CFC	Cognitive-Functional Composite
CHL	Competitive Hebbian Learning
CN	Cognitively Normal
CNN	Convolutional Neural Network
CNS	Central Nervous System
CPN	Counterpropagation Network
CPU	Central Processing Unit
CSF	Cerebrospinal Fluid
CT	Computed Tomography
CUI	Clinical Utility Index
CW	Clinical Workstation
DCL	Deterioro Cognitivo Leve
DL	Deep Learning
DLB	Dementia with Lewy bodies
DNN	Deep Neural Network
DSM	Diagnostic and Statistical Manual of Mental Disorders
DT	Decision Tree
DTI	Diffusion Tensor Imaging
DW-MRI	Diffusion-Weighted Magnetic Resonance Imaging
EA	Enfermedad de Alzheimer
ECog	Everyday Cognition
EDEVITALZH	Clinical Virtual Environment to aid diagnosis and prognosis of Alzheimer's Disease and other dementias
EEG	Electroencephalography
EII	Escuela de Ingeniería Informática
ELM	Extreme Learning Machine
EmITIC	Empresa, Internet y Tecnologías de las Comunicaciones
EMCI	Early Mild Cognitive Impairment
EMI	Electronic Medical Interconsultation
EMRDB	Electronic Medical Records Database
fMRI	Functional Magnetic Resonance Imaging
FAQ	Functional Activities Questionnaire
FCBF	Fast Correlation-Based Filter
FDG-PET	Fluorodeoxyglucose-Positron Emission Tomography

FDG	Fluorodeoxyglucose
FN	False Negatives
fNIRS	Functional Near-Infrared Spectroscopy
FP	False Positives
FPR	False Positive Rate
FPGA	Field Programmable Gate Arrays
FTD	Frontotemporal dementia
GAN	Generative Adversarial Network
GCN	Graph Convolutional Network
GCPD	Global Clinical Protocol for Dementia
GCS	Growing Cell Structures
GDS	Geriatric Depression Scale
GEO	Gene Expression Omnibus
GNG	Growing Neural Gas
GNN	Graph Neural Network
GPU	Graphics Processing Unit
HPC	High-Performance Computing
HUMANN	Hybrid Unsupervised Modular Adaptive Neural Network
ICV	Intracranial Volume
iCW	intelligent Clinical Workstation
ICWA	Intelligent Clinical Wizard and Assistant
IDST	Intelligent Decision Support Tool
ISD	Intelligent Systems for Diagnosis
IUCES	Instituto Universitario de Cibernética, Empresa y Sociedad
IWG	International Working Group
IWG-1	International Working Group-1
IWG-2	International Working Group-2
***k*-NN**	k-Nearest Neighbors
LATE	Limbic-predominant Age-related TDP-43 Encephalopathy
LMCI	Late Mild Cognitive Impairment
LR	Logistic Regression
LSTM	Long Short-Term Memory
LVQ	Learning Vector Quantization
MAS	Sydney Memory and Ageing Study
MCC	Matthews Correlation Coefficient

MCI	Mild Cognitive Impairment
MEG	Magnetoencephalography
ML	Machine Learning
MLP	Multilayer Perceptron
MMSE	Mini-Mental State Examination
MoCA	Montreal Cognitive Assessment
MRI	Magnetic Resonance Imaging
MyGNG	Modular Hybrid Growing Neural Gas
naMCI	nonamnestic Mild Cognitive Impairment
NB	Naïve Bayes
NC	Normal Control
NCA	Neighborhood Component Analysis
NFT	neurofibrillary tangle
NG	Neural Gas
NIA-AA	National Institute on Aging and Alzheimer's Association
NIA	National Institute on Aging
NINCDS-ADRDA	National Institute of Neurological and Communicative Disorders and Stroke and Alzheimer's Disease and Related Disorders Association
NINCDS	National Institute of Neurological and Communicative Disorders and Stroke
NPI-Q	Neuropsychiatric Inventory-Questionnaire
NPI	Neuropsychiatric Inventory
OASIS	Open Access Series of Imaging Studies
OCTA	Optical Coherence Tomography Angiography
OS	operating system
OVL	overlapping coefficient
PCA	Principal Component Analysis
PD	Parkinson's Disease
PDD	Parkinson's Disease Dementia
PET	Positron Emission Tomography
PiB	Pittsburgh compound B
pMCI	progressive Mild Cognitive Impairment
PMS	Patient Management Software
PPV	Positive Predictive Value
qMRI	Quantitative Magnetic Resonance Imaging
RAVLT	Rey Auditory Verbal Learning Test

RBF	Radial Basis Function
ResNet	Residual Network
RF	Random Forest
RNA	Red Neuronal Artificial
RNN	Recurrent Neural Network
ROC	Receiver Operating Characteristic
ROI	Region of Interest
SHAP	SHapley Additive exPlanations
sMRI	Structural Magnetic Resonance Imaging
SAE	Stacked Autoencoder
SMC	Significant Memory Concern
SMV	Simple Majority Voting
SNP	Single Nucleotide Polymorphism
SOM	Self-Organizing Map
SPECT	Single-Photon Emission Computed Tomography
SupeRGNG	Supervised Reconfigurable Growing Neural Gas
SVM	Support Vector Machine
TADPOLE	The Alzheimer's Disease Prediction Of Longitudinal Evolution
TMS	Transcranial Magnetic Stimulation
TN	True Negatives
TNR	True Negative Rate
TP	True Positives
TPR	True Positive Rate
t-SNE	t-distributed stochastic neighbor embedding
UI	User Interface
ULL	Universidad de La Laguna
ULPGC	Universidad de Las Palmas de Gran Canaria
VaD	Vascular dementia present
VGG	Visual Geometry Group
ViT	Vision Transformer
WMS	Wechsler Memory Scale
WNN	Wavelet Neural Network
XGBoost	Extreme Gradient Boosting

Appendix B

Growing Neural Gas hyperparameter recommendations from other authors

Few succinct recommendations for selecting the values of the GNG hyperparameters have been found in the literature [Fritzke, 1995; Hamker and Heinke, 1997; Heinke and Hamker, 1998; Holmström, 2002].

GNG has low sensitivity to variation of its parameters, and only a few of them severely influence the behavior of the network [Heinke and Hamker, 1998][1]: learning rate of best (ε_c; 0.1 or 0.2), learning rate of neighbors (ε_n; 0.006 or 0.012), learning rate of output (None[2]; 0.15), adaptation steps (*epochs*; 100 or 200), decreasing counters (d; 0.995), decreasing of signal counters (β; 0.5), and maximum age of edges (a_max; 50). These authors indicated that no insertions for rare classes tend to occur when too high values for the adaptation step are chosen. Having fewer nodes and tuning the other parameters appropriately to obtain similar performance is possible [Heinke and Hamker, 1998]. In their experiments, the maximum number of epochs was 200, and each dataset was run 30 times: six of every parameter set, each with five randomly selected initializations.

Selecting values for the parameters ε_c and ε_n, which are the steps or learning rates for the winner neuron and its neighbors, respectively, depends on both the other GNG parameters and the input distribution. In most cases, ε_c and ε_n should have low values, ideally lower than 0.3 [Holmström, 2002]: for example, 0.05 and 0.0006. Too high values will give unstable GNG networks, while too low values cause GNG to adapt sluggishly. Also, ε_n should use values one or two orders of magnitude smaller than the ones given to ε_c so that the winner neuron adapts faster than its neighbors.

The λ parameter, the neuron insertion rate that has a fixed value in the original GNG, heavily impacts the GNG performance [Holmström, 2002]. Too low values provide non-optimal initial distribution of neurons, where some neurons will become inactive with time as they will be far from the inputs [Holmström, 2002]. This implies that more epochs are required in order to delete these inactive nodes and create new ones where they are more needed to better represent the inputs. Conversely, too high values cause GNG to grow slowly,

[1] According to [Heinke and Hamker, 1998], "only some parameters have a strong influence on the outcome of the training", Table XVIII in that document: learning rate of best, learning rate of neighbors, learning rate of output, adaptation steps, decreasing counters, decreasing of signal counters, and maximal age of edges.

[2] Apparently, these authors used a supervised variant of the GNG, further explained in [Hamker and Heinke, 1997], as the original GNG in [Fritzke, 1995] does not have the parameter called "learning rate of output".

hence needing more epochs, but the distribution of neurons will be correct.

Appendix C

Statistical significance, null hypothesis, significance level and *p*-value

Three concepts are important to be defined beforehand:

- The null hypothesis, which refers to the hypothesis where the phenomenon being studied has no effect. It can be rejected if the result is statistically significant. That is, its p-value is less or equal than the pre-specified significance level α.
- The significance level α, the probability of the study rejecting the null hypothesis, considering the null hypothesis to be true. A value of 0.05 is common, or much lower in some fields of study. It needs to be selected before data collection.
- The p-value, which is the probability of occurrence of a result at least as "extreme", considering the null hypothesis to be true.

Results or an observed effect during a research may have been obtained by chance or due to a sampling error. To conclude that a result has statistical significance, it should be very infrequent to obtain a result at least as "extreme", considering the null hypothesis to be true.

Many methods have been used to statistically compare classifiers over multiple datasets [Demšar, 2006]. However, a few of them can be used to tackle most research questions [Shankar and Singh, 2014]. Choosing which statistical test to use depends on several aspects [University of Minnesota, 2018]: the research design, the distribution of the data, and the type of variable. Not selecting the most appropriate statistical test might make the results obtained become not significant. Hence, the importance of checking the indicated aspects. The most common mistake is to use a parametric test, mainly used with normally distributed data, when the data is non-normal, which requires a non-parametric one. Non-parametric tests equivalent to parametric ones exist [Boslaugh and Watters, 2008; McDonald, 2014; Tronstad and Pripp, 2014; Greenland et al., 2016; Marshall and Russell, 2016].

Evaluating the statistical association of each clinical criteria (feature) with the disease severity of a patient (sample) is another use of some of these statistical methods, especially in a clinical context [Yao et al., 2020].

Based on the tables and schemes found in several articles [De Muth, 2009; Campbell and Shantikumar, 2010; Marusteri and Bacarea, 2010; Nayak and Hazra, 2011; Shankar and Singh, 2014; SmartVision Europe, 2017], we summarize them in a Decision Tree, Figure C.1, that can be used to ease the selection of the most well-suited statistical test. The main questions to be answered are: the type of study to be carried out, the number of classes

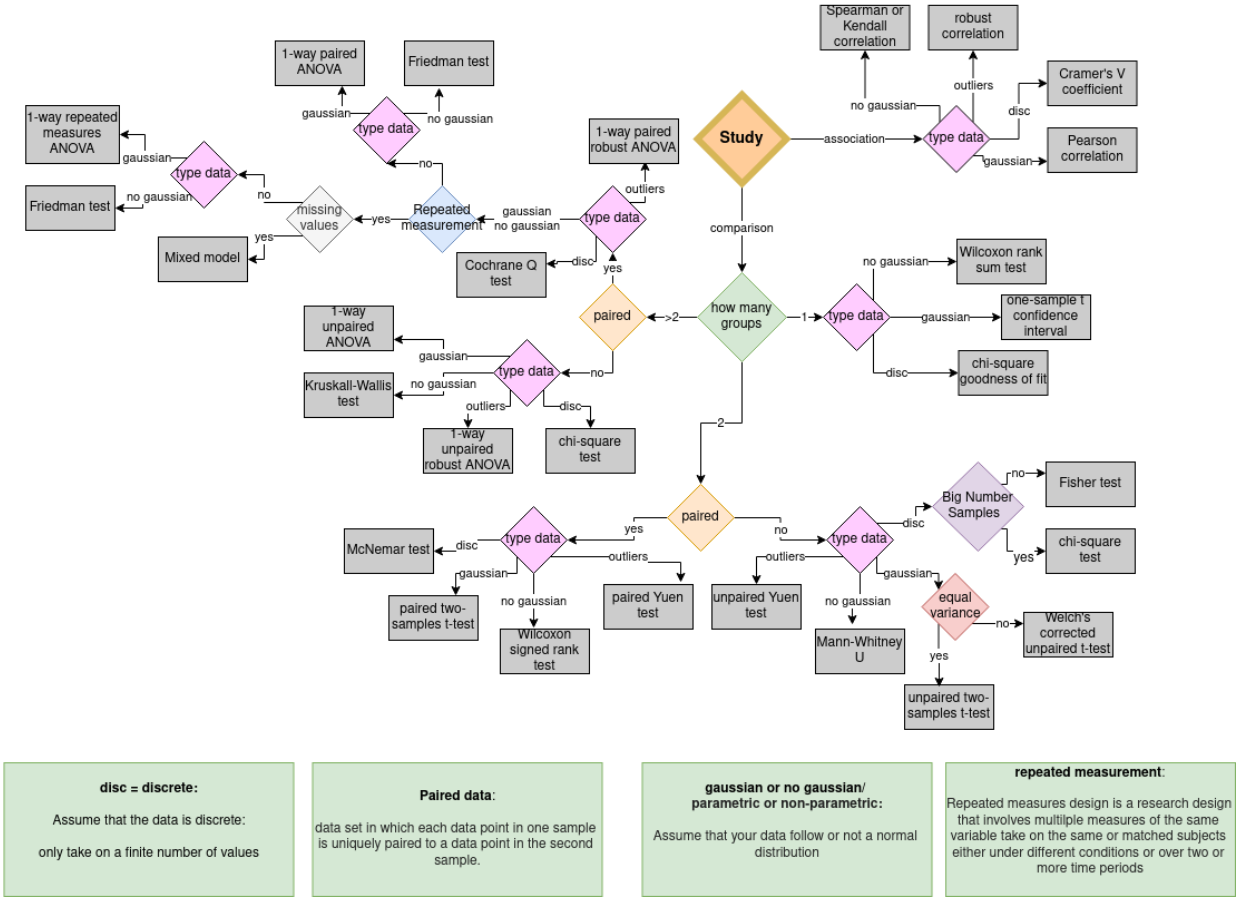


Figure C.1: Decision tree to choose the most appropriate statistical test.

present in the data, the dependence of the data to be analyzed (paired vs unpaired tests), and if these data follow the normal distribution (parametric vs non-parametric tests).

Bibliography

Aberathne, I., Kulasiri, D., and Samarasinghe, S. (2023). Detection of Alzheimer's disease onset using MRI and PET neuroimaging: Longitudinal data analysis and machine learning. *Neural Regeneration Research*, 18(10):2134–2140.

Airola, A., Pahikkala, T., Waegeman, W., De Baets, B., and Salakoski, T. (2010). A comparison of AUC estimators in small-sample studies. In *3rd International Workshop on Machine Learning in Systems Biology (MLSB 09)*, pages 15–23.

Akobeng, A. K. (2007a). Understanding diagnostic tests 1: Sensitivity, specificity and predictive values. *Acta Paediatrica*, 96(3):338–341.

Akobeng, A. K. (2007b). Understanding diagnostic tests 2: Likelihood ratios, pre- and post-test probabilities and their use in clinical practice. *Acta Paediatrica*, 96(4):487–491.

Akobeng, A. K. (2007c). Understanding diagnostic tests 3: Receiver operating characteristic curves. *Acta Paediatrica*, 96(5):644–647.

Alberca Serrano, R. and López-Pousa, S. (2011). *Enfermedad de Alzheimer y Otras Demencias*. Editorial Médica Panamericana, 4 edition.

Alberdi, A., Aztiria, A., and Basarab, A. (2016). On the early diagnosis of Alzheimer's Disease from multimodal signals: A survey. *Artificial Intelligence in Medicine*, 71:1–29.

Alfalahi, H., Dias, S. B., Khandoker, A. H., Chaudhuri, K. R., and Hadjileontiadis, L. J. (2023). A scoping review of neurodegenerative manifestations in explainable digital phenotyping. *npj Parkinson's Disease 2023 9:1*, 9(1):1–22.

Aliño, J. J. L.-I., Association, A. P., and Miyar, M. V. (2008). *DSM-IV-TR: manual diagnóstico y estadístico de los trastornos mentales*. American Psychiatric Pub.

Aljović, A., Badnjević, A., and Gurbeta, L. (2016). Artificial neural networks in the discrimination of Alzheimer's disease using biomarkers data. In *2016 5th Mediterranean Conference on Embedded Computing (MECO)*, pages 286–289.

Alkabawi, E. M., Hilal, A. R., and Basir, O. A. (2017). Computer-aided classification of multi-types of dementia via convolutional neural networks. In *2017 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*, pages 45–50.

Alzheimer's Association (2018). 2018 Alzheimer's disease facts and figures. *Alzheimer's & Dementia*, 14(3):367–429.

Alzheimer's Association (2025). 2025 Alzheimer's Disease Facts and Figures. *Alzheimer's & Dementia*, 21(4).

Alzheimer's Australia (2017). Diagnostic Criteria for Dementia. Help sheet, Alzheimer's Australia.

Alzheimer's Disease Neuroimaging Initiative (ADNI) (2013). ADNI Data and Publications.

Anderson, E. (1936). The Species Problem in Iris. *Annals of the Missouri Botanical Garden*, 23:457–509.

Ardalan, Z. and Subbian, V. (2022). Transfer Learning Approaches for Neuroimaging Analysis: A Scoping Review. *Frontiers in Artificial Intelligence*, 5:1–15.

Arimura, H., Yoshiura, T., Kumazawa, S., Tanaka, K., Koga, H., Mihara, F., Honda, H., Sakai, S., Toyofuku, F., and Higashida, Y. (2008). Automated Method for Identification of Patients With Alzheimer's Disease Based on Three-dimensional MR Images. *Academic Radiology*, 15(3):274–284.

Arlot, S. and Celisse, A. (2010). A survey of cross-validation procedures for model selection. *Statistics Surveys*, 4:40–79.

Arya, A. D., Verma, S. S., Chakarabarti, P., Chakrabarti, T., Elngar, A. A., Kamali, A. M., and Nami, M. (2023). A systematic review on machine learning and deep learning techniques in the effective diagnosis of Alzheimer's disease. *Brain Informatics*, 10(1):1–15.

Bagyinszky, E., Youn, Y. C., An, S. S. A., and Kim, S. (2014). Diagnostic methods and biomarkers for Alzheimer's disease. *Toxicology and Environmental Health Sciences*, 6(3):133–147.

Baloyannis, S. I., Mavroudis, I., Mitilineos, D., Baloyannis, I. S., and Costa, V. G. (2019). The Hypothalamus in Alzheimer's Disease: A Golgi and Electron and Microscope Study. *Cognitive Disorders*.

Barkhof, F., Fox, N. C., Bastos-Leite, A. J., and Scheltens, P. (2011). *Neuroimaging in Dementia*. Springer-Verlag, Berlin Heidelberg.

Basogain Olabe, X. (1998). Redes Neuronales Artificiales y sus Aplicaciones.

Battineni, G., Chintalapudi, N., Amenta, F., and Traini, E. (2020). A Comprehensive Machine-Learning Model Applied to Magnetic Resonance Imaging (MRI) to Predict Alzheimer's Disease (AD) in Older Subjects. *Journal of Clinical Medicine*, 9(7):2146.

Battineni, G., Chintalapudi, N., Hossain, M. A., Losco, G., Ruocco, C., Sagaro, G. G., Traini, E., Nittari, G., and Amenta, F. (2022). Artificial Intelligence Models in the Diagnosis of Adult-Onset Dementia Disorders: A Review. *Bioengineering*, 9(8):1–15.

Baumgart, M., Snyder, H. M., Carrillo, M. C., Fazio, S., Kim, H., and Johns, H. (2015). Summary of the evidence on modifiable risk factors for cognitive decline and dementia: A population-based perspective. *Alzheimer's & Dementia*, 11(6):718–726.

Bevilacqua, R., Barbarossa, F., Fantechi, L., Fornarelli, D., Paci, E., Bolognini, S., Giammarchi, C., Lattanzio, F., Paciaroni, L., Riccardi, G. R., Pelliccioni, G., Biscetti, L., and Maranesi, E. (2023). Radiomics and artificial intelligence for the diagnosis and monitoring of alzheimer's disease: A systematic review of studies in the field. *Journal of Clinical Medicine 2023, Vol. 12, Page 5432*, 12(16):5432.

Bhachawat, S., Shriram, E., Srinivasan, K., and Hu, Y. C. (2023). Leveraging computational intelligence techniques for diagnosing degenerative nerve diseases: A comprehensive review, open challenges, and future research directions. *Diagnostics 2023, Vol. 13, Page 288*, 13(2):288.

Bonny, J.-M. (2005). Methods and Applications of Quantitative MRI. In Webb, G. A., editor, *Annual Reports on NMR Spectroscopy*, volume 56, pages 213–229. Academic Press.

Borgio, J. G. F., Baldaçara, L., Moraes, W. d. S., Lacerda, A. L. T., Montaño, M. B., Jackowski, A. P., Tufik, S., Ramos, L. R., and Bressan, R. A. (2012). Hippocampal volume and CDR-SB can predict conversion to dementia in MCI patients. *Arquivos de Neuro-Psiquiatria*, 70(11):839–842.

Boslaugh, S. and Watters, P. A. (2008). *Statistics in a Nutshell*. O'Reilly Media, Inc.

Bradley, A. P. (1997). The Use of the Area Under the ROC Curve in the Evaluation of Machine Learning Algorithms. *Pattern Recognition*, 30(7):1145–1159.

Bredesen, D. E. (1995). Neural apoptosis. *Annals of Neurology*, 38(6):839–851.

Breijyeh, Z. and Karaman, R. (2020). Comprehensive Review on Alzheimer's Disease: Causes and Treatment. *Molecules*, 25(24):5789.

Brown, C. D. and Davis, H. T. (2006). Receiver operating characteristics curves and related decision measures: A tutorial. *Chemometrics and Intelligent Laboratory Systems*, 80(1):24–38.

Budelier, M. M. and Bateman, R. J. (2020). Biomarkers of Alzheimer Disease. *The Journal of Applied Laboratory Medicine*, 5(1):194–208.

Byun, S., Han, J. W., Kim, T. H., Kim, K., Kim, T. H., Park, J. Y., Suh, S. W., Seo, J. Y., So, Y., Lee, K. H., Lee, J. R., Jeong, H., Jeong, H.-G., Han, K., Hong, J. W., and Kim, K. W. (2018). Gait Variability Can Predict the Risk of Cognitive Decline in Cognitively Normal Older People. *Dementia and Geriatric Cognitive Disorders*, 45:251–261.

Cabrera León, Y. (2015). *Análisis del Uso de las Redes Neuronales Artificiales en el Diseño de Filtros Antispam: una Propuesta Basada en Arquitecturas Neuronales No Supervisadas*. Final Project, Universidad de Las Palmas de Gran Canaria, Las Palmas de Gran Canaria.

Cabrera-León, Y., Báez, P. G., Fernández-López, P., and Suárez-Araujo, C. P. (2023). Study on Mild Cognitive Impairment and Alzheimer's Disease Classification using a New Ontogenic Neural Architecture, the Supervised Reconfigurable Growing Neural Gas. In *2023 Annual Modeling and Simulation Conference (ANNSIM 2023)*, pages 425–436, Mohawk College, ON, Canada. IEEE.

Cabrera-León, Y., Báez, P. G., Fernández-López, P., and Suárez-Araujo, C. P. (2024a). Neural Computation-Based Methods for the Early Diagnosis and Prognosis of Alzheimer's Disease Not Using Neuroimaging Biomarkers: A Systematic Review. *Journal of Alzheimer's Disease*, 98(3):793–823.

Cabrera-León, Y., Báez, P. G., Fernández-López, P., and Suárez-Araujo, C. P. ((UNDER REVIEW) 2025). Neural computation methods for the early diagnosis and prognosis of Alzheimer's Disease based on neuroimaging. A Systematic Review.

Cabrera-León, Y., Báez, P. G., Ruiz-Alzola, J., and Suárez-Araujo, C. P. (2018a). Classification of Mild Cognitive Impairment Stages Using Machine Learning Methods. In *2018 IEEE 22nd International Conference on Intelligent Engineering Systems (INES)*, pages 67–72, Las Palmas de Gran Canaria, Spain. IEEE.

Cabrera-León, Y., Fernández-López, P., García Báez, P., Kluwak, K., Navarro-Mesa, J. L., Suárez-Araujo, C. P., and for the Alzheimer's Disease Neuroimaging Initiative (2024b). Toward an intelligent computing system for the early diagnosis of Alzheimer's disease based on the modular hybrid growing neural gas. *Digital Health*, 10:1–16.

Cabrera-León, Y., Fernández-López, P., García Báez, P., and Suárez-Araujo, C. P. (2024c). Early diagnosis of Alzheimer's disease based on the Supervised Reconfigurable Growing Neural Gas. Towards an eHealth solution. In *1st Congress Bridge to Africa*, pages 171–176, Las Palmas de Gran Canaria. Servicio de Publicaciones y Difusión Científica de la Universidad de Las Palmas de Gran Canaria.

Cabrera-León, Y., García Báez, P., and Suárez-Araujo, C. P. (2018b). E-Mail Spam Filter Based on Unsupervised Neural Architectures and Thematic Categories: Design and Analysis. In Merelo, J. J., Melício, F., Cadenas, J. M., Dourado, A., Madani, K., Ruano, A., and Filipe, J., editors, *Computational Intelligence: International Joint Conference, IJCCI 2016 Porto, Portugal, November 9–11, 2016 Revised Selected Papers*, volume 792 of *Studies in Computational Intelligence*, pages 239–262. Springer International Publishing, Cham.

Cabrera-León, Y., García Báez, P., and Suárez-Araujo, C. P. (2018c). Non-email Spam and Machine Learning-Based Anti-spam Filters: Trends and Some Remarks. In Moreno-Díaz, R., Pichler, F., and Quesada-Arencibia, A., editors, *Computer Aided Systems Theory – EUROCAST 2017*, volume 10671 of *Lecture Notes in Computer Science*, pages 245–253. Springer, Cham.

Caliński, T. and Harabasz, J. (1974). A dendrite method for cluster analysis. *Communications in Statistics - Theory and Methods*, 3(1):1–27.

Campbell, MJ. and Shantikumar, S. (2010). Parametric and Non-parametric tests for comparing two or more groups.

Carpenter, G. A. and Grossberg, S. (2003). Adaptive Resonance Theory. In Arbib, M. A., editor, *The Handbook of Brain Theory and Neural Networks*. MIT Press, Cambridge, MA, USA, 2nd edition.

Cassani, R., Estarellas, M., San-Martin, R., Fraga, F. J., and Falk, T. H. (2018). Systematic Review on Resting-State EEG for Alzheimer's Disease Diagnosis and Progression Assessment. *Disease Markers*, 2018:1–26.

Cedarbaum, J. M., Jaros, M., Hernandez, C., Coley, N., Andrieu, S., Grundman, M., Vellas, B., and Alzheimer's Disease Neuroimaging Initiative (2013). Rationale for use of the Clinical Dementia Rating Sum of Boxes as a primary outcome measure for Alzheimer's disease clinical trials. *Alzheimer's & Dementia: The Journal of the Alzheimer's Association*, 9(1 Suppl):S45–S55.

Chan, M., Andre, B., Herrera, A., and Celsis, P. (1993). Incremental learning in a multilayer neural network as an aid to Alzheimer's disease diagnosis. In *Proceedings of IEEE Systems Man and Cybernetics Conference - SMC*, volume 4, pages 1–4 vol.4.

Chandra, A., Dervenoulas, G., Politis, M., and for the Alzheimer's Disease Neuroimaging Initiative (2019). Magnetic resonance imaging in Alzheimer's disease and mild cognitive impairment. *Journal of Neurology*, 266(6):1293–1302.

Chapelle, O., Schölkopf, B., and Zien, A. (2006). *Semi-Supervised Learning*, volume 2. MIT Press, Cambridge, MA, USA.

Chen, T. and Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '16*, pages 785–794, San Francisco, California, USA. ACM Press.

Chollet, F. et al. (2015). Keras: Deep Learning for humans.

Clark, J. and Allan Holton, D. (1991). *A First Look at Graph Theory*. WORLD SCIENTIFIC.

Cooper, C., Sommerlad, A., Lyketsos, C. G., and Livingston, G. (2015). Modifiable Predictors of Dementia in Mild Cognitive Impairment: A Systematic Review and Meta-Analysis. *American Journal of Psychiatry*, 172(4):323–334.

Coppedè, F., Grossi, E., Buscema, M., and Migliore, L. (2013). Application of Artificial Neural Networks to Investigate One-Carbon Metabolism in Alzheimer's Disease and Healthy Matched Individuals. *PLoS ONE*, 8(8):e74012.

Cortes, C. and Vapnik, V. (1995). Support-Vector Networks. *Machine Learning*, 20(3):273–297.

Crary, J. F. (2023). Neurodegeneration: 2023 update. *Free neuropathology*, 4.

Cui, R., Liu, M., and Li, G. (2018). Longitudinal analysis for Alzheimer's disease diagnosis using RNN. In *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, pages 1398–1401, Washington, DC, USA. IEEE.

Cummings, J. (2018). The National Institute on Aging—Alzheimer's Association Framework on Alzheimer's disease: Application to clinical trials. *Alzheimer's & Dementia*.

Cummings, J. L. (1997). The Neuropsychiatric Inventory Assessing psychopathology in dementia patients. *Neurology*, 48(5 Suppl 6):S10–S16.

Cummings, J. L., Mega, M., Gray, K., Rosenberg-Thompson, S., Carusi, D. A., and Gornbein, J. (1994). The Neuropsychiatric Inventory: Comprehensive assessment of psychopathology in dementia. *Neurology*, 44(12):2308–2308.

da Silva, A. B. and Rosa, J. L. G. (2013). Biological Plausibility in an Artificial Neural Network Applied to Real Predictive Tasks. In Rutkowski, L., Korytkowski, M., Scherer, R., Tadeusiewicz, R., Zadeh, L. A., and Zurada, J. M., editors, *Artificial Intelligence and Soft Computing*, pages 183–192, Berlin, Heidelberg. Springer.

da Silva, S. F., Brandoli, B., Eler, D. M., Neto, J. B., and Traina, A. J. M. (2010). Silhouette-based feature selection for classification of medical images. In *2010 IEEE 23rd International Symposium on Computer-Based Medical Systems (CBMS)*, pages 315–320.

Davies, D. L. and Bouldin, D. W. (1979). A Cluster Separation Measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-1(2):224–227.

Davison, C. M. and O'Brien, J. T. (2013). A comparison of FDG-PET and blood flow SPECT in the diagnosis of neurodegenerative dementias: A systematic review. *International Journal of Geriatric Psychiatry*, 29(6):551–561.

De Muth, J. E. (2009). Overview of biostatistics used in clinical research. *American Journal of Health-System Pharmacy*, 66(1):70–81.

DeMarshall, C. A., Nagele, E. P., Sarkar, A., Acharya, N. K., Godsey, G., Goldwaser, E. L., Kosciuk, M., Thayasivam, U., Han, M., Belinka, B., and Nagele, R. G. (2016). Detection of Alzheimer's disease at mild cognitive impairment and disease progression using autoantibodies as blood-based biomarkers. *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring*, 3:51–62.

Demšar, J. (2006). Statistical Comparisons of Classifiers over Multiple Data Sets. *Journal of Machine Learning Research*, 7:30.

Dolph, C. V., Alam, M., Shboul, Z., Samad, M. D., and Iftekharuddin, K. M. (2017). Deep learning of texture and structural features for multiclass Alzheimer's disease classification. In *2017 International Joint Conference on Neural Networks (IJCNN)*, pages 2259–2266.

Dubois, B., Feldman, H. H., Jacova, C., DeKosky, S. T., Barberger-Gateau, P., Cummings, J., Delacourte, A., Galasko, D., Gauthier, S., Jicha, G., Meguro, K., O'Brien, J., Pasquier, F., Robert, P., Rossor, M., Salloway, S., Stern, Y., Visser, P. J., and Scheltens, P. (2007). Research criteria for the diagnosis of Alzheimer's disease: Revising the NINCDS–ADRDA criteria. *The Lancet Neurology*, 6(8):734–746.

Dubois, B., Hampel, H., Feldman, H. H., Scheltens, P., Aisen, P., Andrieu, S., Bakardjian, H., Benali, H., Bertram, L., Blennow, K., Broich, K., Cavedo, E., Crutch, S., Dartigues, J.-F., Duyckaerts, C., Epelbaum, S., Frisoni, G. B., Gauthier, S., Genthon, R., Gouw, A. A., Habert, M.-O., Holtzman, D. M., Kivipelto, M., Lista, S., Molinuevo, J.-L., O'Bryant, S. E., Rabinovici, G. D., Rowe, C., Salloway, S., Schneider, L. S., Sperling, R., Teichmann, M., Carrillo, M. C., Cummings, J., and Jack, C. R. (2016). Preclinical Alzheimer's disease: Definition, natural history, and diagnostic criteria. *Alzheimer's & Dementia*, 12(3):292–323.

Duda, R. O., Hart, P. E., and Stork, D. G. (2001). *Pattern Classification*. Wiley, 2 edition.

Elazab, A., Wang, C., Abdelaziz, M., Zhang, J., Gu, J., Gorriz, J. M., Zhang, Y., and Chang, C. (2024). Alzheimer's disease diagnosis from single and multimodal data using machine and deep learning models: Achievements and future directions. *Expert Systems with Applications*, 255:124780.

Eluyode, O. S. and Akomolafe, D. T. (2013). Comparative study of biological and artificial neural networks. *European Journal of Applied Engineering and Scientific Research*, 2(1).

Erasmus, L. J., Hurter, D., Naude, M., Kritzinger, H. G., and Acho, S. (2004). A short overview of MRI artefacts. *South African Journal of Radiology*, 8(2):13.

Fagotto, F. (2014). The cellular basis of tissue separation. *Development*, 141(17):3303–3318.

Fan, J., Upadhye, S., and Worster, A. (2006). Understanding receiver operating characteristic (ROC) curves. *Canadian Journal of Emergency Medicine*, 8(1):19–20.

Farias, S. T., Mungas, D., Reed, B. R., Cahn-Weiner, D., Jagust, W., Baynes, K., and DeCarli, C. (2008). The Measurement of Everyday Cognition (ECog): Scale Development and Psychometric Properties. *Neuropsychology*, 22(4):531–544.

Farina, N., Llewellyn, D., Isaac, M. G. E. K. N., and Tabet, N. (2017). Vitamin E for Alzheimer's dementia and mild cognitive impairment. In *The Cochrane Library*. John Wiley & Sons, Ltd.

Fearing, M. A., Bigler, E. D., Norton, M., Tschanz, J. A., Hulette, C., Leslie, C., Welsh-Bohmer, K., and the Cache County Investigators (2007). Autopsy-confirmed Alzheimer's disease versus clinically diagnosed Alzheimer's disease in the Cache County Study on Memory and Aging: A comparison of quantitative MRI and neuropsychological findings. *Journal of Clinical and Experimental Neuropsychology*, 29(5):553–560.

Fei, Z., Yang, E., Li, D., Butler, S., Ijomah, W., and Mackin, N. (2017). A Survey of the State-of-the-Art Techniques for Cognitive Impairment Detection in the Elderly. In *Advanced Computational Methods in Life System Modeling and Simulation*, volume 761 of *Communications in Computer and Information Science*, pages 143–161. Springer Nature Singapore.

Ferrari, M. and Quaresima, V. (2012). A brief review on the history of human functional near-infrared spectroscopy (fNIRS) development and fields of application. *NeuroImage*, 63(2):921–935.

Fiesler, E. and Beale, R., editors (1997). *Handbook of Neural Computation*. Computational Intelligence Library. Taylor & Francis.

Fisher, R. A. (1936). The Use of Multiple Measurements in Taxonomic Problems. *Annals of Eugenics*, 7(2):179–188.

Folstein, M., Folstein, S., and McHugh, P. (1975). "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*, 12(3):189–198.

Forcada, V. R. (2003). *Clasificación supervisada basada en redes bayesianas, aplicación en biología computacional*. Doctoral thesis, Universidad Politécnica de Madrid.

Freeman, J. A. and Skapura, D. M. (1991). *Neural Networks: Algorithms, Applications, and Programming Techniques*. Computation and Neural Systems Series. Addison-Wesley, Reading, Mass.

Frisoni, G. B., Boccardi, M., Barkhof, F., Blennow, K., Cappa, S., Chiotis, K., Démonet, J.-F., Garibotto, V., Giannakopoulos, P., Gietl, A., Hansson, O., Herholz, K., Jack, C. R., Nobili, F., Nordberg, A., Snyder, H. M., Ten Kate, M., Varrone, A., Albanese, E., Becker, S., Bossuyt, P., Carrillo, M. C., Cerami, C., Dubois, B., Gallo, V., Giacobini, E., Gold, G., Hurst, S., Lönneborg, A., Lovblad, K.-O., Mattsson, N., Molinuevo, J.-L., Monsch, A. U., Mosimann, U., Padovani, A., Picco, A., Porteri, C., Ratib, O., Saint-Aubert, L., Scerri, C., Scheltens, P., Schott, J. M., Sonni, I., Teipel, S., Vineis, P., Visser, P. J., Yasui, Y., and Winblad, B. (2017). Strategic roadmap for an early diagnosis of Alzheimer's disease based on biomarkers. *The Lancet Neurology*, 16(8):661–676.

Frisoni, G. B., Fox, N. C., Jack, C. R., Scheltens, P., and Thompson, P. M. (2010). The clinical use of structural MRI in Alzheimer disease. *Nature reviews. Neurology*, 6(2):67–77.

Fritzke, B. (1994). Growing cell structures—A self-organizing network for unsupervised and supervised learning. *Neural Networks*, 7(9):1441–1460.

Fritzke, B. (1995). A Growing Neural Gas Network Learns Topologies. In *Advances in Neural Information Processing Systems 7 (NIPS 1994)*, pages 625–632. MIT Press, Cambridge, MA, USA.

Fritzke, B. (1997a). Some Competitive Learning Methods. Draft, Institute for Neural Computation. Ruhr-Universität Bochum.

Fritzke, B. (1997b). Unsupervised ontogenic networks. In *Handbook of Neural Computation*, Computational Intelligence Library, page 16. Taylor & Francis, Boca Raton, FL, USA.

Frizzell, T. O., Glashutter, M., Liu, C. C., Zeng, A., Pan, D., Hajra, S. G., D'Arcy, R. C. N., and Song, X. (2022). Artificial intelligence in brain MRI analysis of Alzheimer's disease over the past 12 years: A systematic review. *Ageing Research Reviews*, 77(101614):1–17.

Furcila, D., DeFelipe, J., and Alonso-Nanclares, L. (2018). A Study of Amyloid-β and Phosphotau in Plaques and Neurons in the Hippocampus of Alzheimer's Disease Patients. *Journal of Alzheimer's Disease*, 64(2):417–435.

Gabrielsson, S. and Gabrielsson, S. (2006). The Use of self-organizing maps in recommender systems. A survey of the Recommender Systems field and a presentation of a State of the Art Highly Interactive Visual Movie Recommender System. Master's thesis, Department of Information Technology at the Division of Computer Systems, Uppsala University.

Galasko, D., Chang, L., Motter, R., Clark, C. M., Kaye, J., Knopman, D., Thomas, R., Kholodenko, D., Schenk, D., Lieberburg, I., Miller, B., Green, R., Basherad, R., Kertiles, L., Boss, M. A., and Seubert, P. (1998). High Cerebrospinal Fluid Tau and Low Amyloid β42 Levels in the Clinical Diagnosis of Alzheimer Disease and Relation to Apolipoprotein E Genotype. *Archives of Neurology*, 55(7):937–945.

Gale, R. P., Hochhaus, A., and Zhang, M.-J. (2016). What is the (p-) value of the P-value? *Leukemia*, 30(10):1965–1967.

Galletly, C. A., Clark, C. R., and McFarlane, A. C. (1996). Artificial neural networks: A prospective tool for the analysis of psychiatric disorders. *Journal of Psychiatry and Neuroscience*, 21(4):239–247.

Gao, S. and Lima, D. (2022). A review of the application of deep learning in the detection of Alzheimer's disease. *International Journal of Cognitive Computing in Engineering*, 3:1–8.

García Abad, J. (2012). Calibración Local de Predicciones Numéricas de Viento con Técnicas Estadísticas no Lineales (Downscaling Estadístico). Master's thesis, Universidad de Cantabria, Santander.

García Báez, P. (2005). *HUMANN: una nueva red neuronal artificial adaptativa, no supervisada, modular y jerárquica. Aplicaciones en neurociencia y medioambiente*. PhD thesis, Universidad de Las Palmas de Gran Canaria.

García Báez, P., del Pino, M. A. P., Viadero, C. F., and Araujo, C. P. S. (2009). Artificial intelligent systems based on supervised HUMANN for differential diagnosis of cognitive impairment: Towards a 4P-HCDS. In *International Work-Conference on Artificial Neural Networks*, pages 981–988. Springer.

García Báez, P., Fernández-Viadero, C., Pérez Del Pino, M. Á., Procházka, A., and Suárez-Araujo, C. P. (2010). HUMANN-based systems for differential diagnosis of dementia using neuropsychological tests. In *2010 IEEE 14th International Conference on Intelligent Engineering Systems*, pages 67–72.

García Báez, P., Fernández Viadero, C., Rodríguez Espinosa, N., Pérez Del Pino, M. Á., and Suárez-Araujo, C. P. (2015). Detection of mild cognitive impairment using a counterpropagation network based system. An e-Health solution. In *2015 International Workshop on Computational Intelligence for Multimedia Understanding (IWCIM)*, pages 1–5.

García Báez, P., Suarez Araujo, C., and Fernández López, P. (2011). Neural Network Ensembles with Missing Data Processing and Data Fusion Capacities: Applications in Medicine and in the Environment. In *IWANN 2011*, volume 6692, pages 169–176.

García Báez, P., Suárez Araujo, C. P., Fernández Viadero, C., and Procházka, A. (2012). Differential Diagnosis of Dementia Using HUMANN-S Based Ensembles. In Kacprzyk, J., Fodor, J., Klempous, R., and Suárez Araujo, C. P., editors, *Recent Advances in Intelligent Engineering Systems*, volume 378, pages 305–324. Springer Berlin Heidelberg, Berlin, Heidelberg.

García Báez, P., Suárez Araujo, C. P., Fernández Viadero, C., and Regidor García, J. (2007). Automatic Prognostic Determination and Evolution of Cognitive Decline Using Artificial Neural Networks. In Yin, H., Tino, P., Corchado, E., Byrne, W., and Yao, X., editors, *Intelligent Data Engineering and Automated Learning - IDEAL 2007*, volume 4881 of *Lecture Notes in Computer Science*, pages 898–907. Springer Berlin Heidelberg, Berlin, Heidelberg.

Gillain, S., Drame, M., Lekeu, F., Wojtasik, V., Ricour, C., Croisier, J.-L., and Petermans, J. (2016). Gait speed or gait variability, which one to use as a marker of risk to develop Alzheimer's disease? A pilot study. *Aging Clinical and Experimental Research*, 28(2).

Goceri, E. and Goceri, N. (2017). Deep Learning in Medical Image Analysis: Recent Advances and Future Trends. In *International Conference Big Data Analytics, Data Mining and Computational Intelligence 2017*, pages 305–310, Portugal.

Goldberger, J., Roweis, S., Hinton, G., and Salakhutdinov, R. (2004). Neighbourhood Components Analysis. In *NIPS'04: Proceedings of the 17th International Conference on Neural Information Processing Systems*, pages 513–520, Vancouver, British Columbia, Canada. MIT Press.

Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep Learning*. MIT Press.

Gosztolya, G., Vincze, V., Tóth, L., Pákáski, M., Kálmán, J., and Hoffmann, I. (2019). Identifying Mild Cognitive Impairment and mild Alzheimer's disease based on spontaneous speech using ASR and linguistic features. *Computer Speech & Language*, 53:181–197.

Greenland, S., Senn, S. J., Rothman, K. J., Carlin, J. B., Poole, C., Goodman, S. N., and Altman, D. G. (2016). Statistical tests, P values, confidence intervals, and power: A guide to misinterpretations. *European Journal of Epidemiology*, 31(4):337–350.

Gunes, S., Aizawa, Y., Sugashi, T., Sugimoto, M., and Rodrigues, P. P. (2022). Biomarkers for Alzheimer's Disease in the Current State: A Narrative Review. *International Journal of Molecular Sciences*, 23(9):4962.

Gutoski, M., Ribeiro, M., Aquino, N. M. R., Hattori, L. T., Lazzaretti, A. E., and Lopes, H. S. (2018). Feature Selection Using Differential Evolution for Unsupervised Image Clustering. In Rutkowski, L., Scherer, R., Korytkowski, M., Pedrycz, W., Tadeusiewicz, R., and Zurada, J. M., editors, *Artificial Intelligence and Soft Computing*, pages 376–385, Cham. Springer International Publishing.

Hagan, M. T., Demuth, H. B., Beale, M. H., and De Jesús, O. (2014). *Neural Network Design*. Martin Hagan, 2nd edition.

Hamker, F. and Heinke, D. (1997). Implementation and Comparison of Growing Neural Gas, Growing Cell Structures and Fuzzy Artmap. Technical Report 1/97, Technische Universität Ilmenau.

Hampel, H., Prvulovic, D., Teipel, S., Jessen, F., Luckhaus, C., Frölich, L., Riepe, M. W., Dodel, R., Leyhe, T., Bertram, L., Hoffmann, W., and Faltraco, F. (2011). The future of Alzheimer's disease: The next 10 years. *Progress in Neurobiology*, 95(4):718–728.

Harris, T. J. C. and Tepass, U. (2010). Adherens junctions: From molecules to morphogenesis. *Nature Reviews Molecular Cell Biology*, 11(7):502–514.

Hastie, T., Tibshirani, R., and Friedman, J. (2009). *The Elements of Statistical Learning. Data Mining, Inference, and Prediction*. Springer Series in Statistics. Springer, 2nd edition.

Haykin, S. S. (1999). *Neural Networks. A Comprehensive Foundation*. Prentice-Hall International, Ontario, Canada, 2nd edition.

Haykin, S. S. (2009). *Neural Networks and Learning Machines*. Prentice Hall.

He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep Residual Learning for Image Recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778, Las Vegas, NV, USA. IEEE.

Hebb, D. O. (1949). *The Organization of Behavior: A Neuropsychological Theory*. Psychology Press.

Hecht-Nielsen, R. (1987). Counterpropagation networks. *Applied Optics*, 26(23):4979–4984.

Hecht-Nielsen, R. (1990). *Neurocomputing*. Reading, Mass. : Addison-Wesley Pub. Co.

Heinke, D. and Hamker, F. H. (1998). Comparing neural networks: A benchmark on growing neural gas, growing cell structures, and fuzzy ARTMAP. *IEEE Transactions on Neural Networks*, 9(6):1279–1291.

Hinton, G. E. and Roweis, S. (2002). Stochastic Neighbor Embedding. In *Advances in Neural Information Processing Systems*, volume 15. MIT Press.

Hofmann, H., Wickham, H., and Kafadar, K. (2017). Letter-Value Plots: Boxplots for Large Data. *Journal of Computational and Graphical Statistics*, 26(3):469–477.

Holmström, J. (2002). Experiments with GNG, GNG with Utility and Supervised GNG. Master's thesis, Uppsala University, Sweden.

Hosseini-Asl, E., Ghazal, M., Mohmoud, A., Aslantas, A., Shalaby, A., Casanova, M., Barnes, G., Gimel'farb, G., Keynton, R., and El-Baz, A. (2018). Alzheimer's disease diagnostics by a 3D deeply supervised adaptable convolutional network. *Frontiers in Bioscience*, 23:584–596.

Huang, W., Zeng, J., Wan, C., Ding, H., and Chen, G. (2017). Image-based dementia disease diagnosis via deep low-resource pair-wise learning. *Multimedia Tools and Applications*.

Hughes, C. P., Berg, L., Danziger, W. L., Coben, L. A., and Martin, R. L. (1982). A new clinical scale for the staging of dementia. *The British Journal of Psychiatry*, 140(6):566–572.

Hurd, M. D., Martorell, P., Delavande, A., Mullen, K. J., and Langa, K. M. (2013). Monetary Costs of Dementia in the United States. *New England Journal of Medicine*, 368(14):1326–1334.

Huynh, R. A. and Mohan, C. (2017). Alzheimer's Disease: Biomarkers in the Genome, Blood, and Cerebrospinal Fluid. *Frontiers in Neurology*, 8(102):1–15.

Inman, H. F. and Bradley Jr, E. L. (1989). The overlapping coefficient as a measure of agreement between probability distributions and point estimation of the overlap of two normal densities. *Communications in Statistics - Theory and Methods*, 18(10):3851–3874.

Iqbal, M. S., Heyat, M. B. B., Parveen, S., Hayat, M. A. B., Roshanzamir, M., Alizadehsani, R., Akhtar, F., Sayeed, E., Hussain, S., Hussein, H. S., and Sawan, M. (2024). Progress and trends in neurological disorders research based on deep learning. *Computerized medical imaging and graphics : the official journal of the Computerized Medical Imaging Society*, 116.

Ivanova, O., Meilán, J. J. G., Martínez-Sánchez, F., Martínez-Nicolás, I., Llorente, T. E., and González, N. C. (2022). Discriminating speech traits of Alzheimer's disease assessed through a corpus of reading task for Spanish language. *Computer Speech & Language*, 73(101341):1–13.

Jack, C. R., Bennett, D. A., Blennow, K., Carrillo, M. C., Dunn, B., Haeberlein, S. B., Holtzman, D. M., Jagust, W., Jessen, F., Karlawish, J., Liu, E., Molinuevo, J. L., Montine, T., Phelps, C., Rankin, K. P., Rowe, C. C., Scheltens, P., Siemers, E., Snyder, H. M., Sperling, R., Elliott, C., Masliah, E., Ryan, L., and Silverberg, N. (2018). NIA-AA Research Framework: Toward a biological definition of Alzheimer's disease. *Alzheimer's & Dementia*, 14(4):535–562.

Jack, C. R., Bernstein, M. A., Fox, N. C., Thompson, P., Alexander, G., Harvey, D., Borowski, B., Britson, P. J., Whitwell, J. L., Ward, C., Dale, A. M., Felmlee, J. P., Gunter, J. L., Hill, D. L., Killiany, R., Schuff, N., Fox-Bosetti, S., Lin, C., Studholme, C., DeCarli, C. S., Krueger, G., Ward, H. A., Metzger, G. J., Scott, K. T., Mallozzi, R., Blezek, D., Levy, J., Debbins, J. P., Fleisher, A. S., Albert, M., Green, R., Bartzokis, G., Glover, G., Mugler, J., and Weiner, M. W. (2008). The Alzheimer's Disease Neuroimaging Initiative (ADNI): MRI Methods. *Journal of magnetic resonance imaging : JMRI*, 27(4):685–691.

Jack Jr., C. R., Andrews, J. S., Beach, T. G., Buracchio, T., Dunn, B., Graf, A., Hansson, O., Ho, C., Jagust, W., McDade, E., Molinuevo, J. L., Okonkwo, O. C., Pani, L., Rafii, M. S., Scheltens, P., Siemers, E., Snyder, H. M., Sperling, R., Teunissen, C. E., and Carrillo, M. C. (2024). Revised criteria for diagnosis and staging of Alzheimer's disease: Alzheimer's Association Workgroup. *Alzheimer's & Dementia*, 20(8):5143–5169.

Jain, A. K., Mao, J., and Mohiuddin, K. M. (1996). Artificial Neural Networks: A Tutorial. *IEEE Computational Science & Engineering*, 29(3):31–44.

Jenkins, R., Fox, N. C., Rossor, A. M., Harvey, R. J., and Rossor, M. N. (2000). Intracranial Volume and Alzheimer Disease: Evidence Against the Cerebral Reserve Hypothesis. *Archives of Neurology*, 57(2):220–224.

Jo, T., Nho, K., and Saykin, A. J. (2019). Deep Learning in Alzheimer's Disease: Diagnostic Classification and Prognostic Prediction Using Neuroimaging Data. *Frontiers in Aging Neuroscience*, 11(220):1–14.

Johnson, K. A., Fox, N. C., Sperling, R. A., and Klunk, W. E. (2012). Brain Imaging in Alzheimer Disease. *Cold Spring Harbor Perspectives in Medicine*, 2(4).

Jolliffe, I. T. and Cadima, J. (2016). Principal component analysis: A review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 374(2065):20150202.

Jutten, R. J., Harrison, J., Lee Meeuw Kjoe, P. R., Opmeer, E. M., Schoonenboom, N. S. M., de Jong, F. J., Ritchie, C. W., Scheltens, P., and Sikkes, S. A. M. (2018). A novel cognitive-functional composite measure to detect changes in early Alzheimer's disease: Test–retest reliability and feasibility. *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring*, 10:153–160.

Kamimura, R. (2010). Information enhancement for interpreting competitive learning. *International Journal of General Systems*, 39(7):705–728.

Katz, S., Ford, A. B., Moskowitz, R. W., Jackson, B. A., and Jaffe, M. W. (1963). Studies of Illness in the Aged: The Index of ADL: A Standardized Measure of Biological and Psychosocial Function. *JAMA*, 185(12):914–919.

Kaufer, D. I., Cummings, J. L., Ketchel, P., Smith, V., MacMillan, A., Shelley, T., Lopez, O. L., and DeKosky, S. T. (2000). Validation of the NPI-Q, a brief clinical form of the Neuropsychiatric Inventory. *The Journal of neuropsychiatry and clinical neurosciences*, 12(2):233–239.

Kerdels, J. (2016). *A Computational Model of Grid Cells Based on a Recursive Growing Neural Gas*. PhD thesis, Fernuniversität in Hagen, Hagen.

Khaliq, F., Oberhauser, J., Wakhloo, D., and Mahajani, S. (2023). Decoding degeneration: The implementation of machine learning for clinical detection of neurodegenerative disorders. *Neural Regeneration Research*, 18(6):1235.

Killin, L. O. J., Starr, J. M., Shiue, I. J., and Russ, T. C. (2016). Environmental risk factors for dementia: A systematic review. *BMC Geriatrics*, 16(1):175.

Kippenhan, J. S., Barker, W. W., Pascal, S., Nagel, J., and Duara, R. (1992). Evaluation of a Neural-Network Classifier for PET Scans of Normal and Alzheimer's Disease Subjects. *The Journal of Nuclear Medicine*, 33(8):1459–1467.

Kiran, A., Alsaadi, M., Dutta, A. K., Raparthi, M., Soni, M., Alsubai, S., Byeon, H., Kulkarni, M. H., and Asenso, E. (2024). Bio-inspired deep learning-personalized ensemble Alzheimer's diagnosis model for mental well-being. *SLAS Technology*, page 100161.

Klekociuk, S. Z. and Summers, M. J. (2014). Exploring the validity of mild cognitive impairment (MCI) subtypes: Multiple-domain amnestic MCI is the only identifiable subtype at longitudinal follow-up. *Journal of Clinical and Experimental Neuropsychology*, 36(3):290–301.

Kohonen, T. (1988). An Introduction to Neural Computing. *Neural Networks*, 1(1):3–16.

Kohonen, T. (2001). *Self-Organizing Maps*. Springer-Verlag New York, Secaucus, NJ, USA, 3 edition.

Korolev, I. O. (2014). Alzheimer's Disease: A Clinical and Basic Science Review. *Medical Student Research Journal*, 04(Fall):24–33.

Korolev, S., Safiullin, A., Belyaev, M., and Dodonova, Y. (2017). Residual and plain convolutional neural networks for 3D brain MRI classification. In *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*, pages 835–838.

Kotsiantis, S. B. (2013). Decision trees: A recent overview. *Artificial Intelligence Review*, 39(4):261–283.

Koutroumbas, K. and Theodoridis, S. (2008). *Pattern Recognition*. Academic Press, fourth edition.

Kriesel, D. (2007). *A Brief Introduction to Neural Networks*. Online.

Kumar, R. and Indrayan, A. (2011). Receiver Operating Characteristic (ROC) Curve for Medical Researchers. *Indian pediatrics*, 48(4):277–287.

Landis, J. R. and Koch, G. G. (1977). The Measurement of Observer Agreement for Categorical Data. *Biometrics*, 33(1):159.

Laske, C., Sohrabi, H. R., Frost, S. M., López-de-Ipiña, K., Garrard, P., Buscema, M., Dauwels, J., Soekadar, S. R., Mueller, S., Linnemann, C., Bridenbaugh, S. A., Kanagasingam, Y., Martins, R. N., and O'Bryant, S. E. (2015). Innovative diagnostic tools for early detection of Alzheimer's disease. *Alzheimer's & Dementia*, 11(5):561–578.

Lawton, M. P. and Brody, E. M. (1969). Assessment of Older People: Self-Maintaining and Instrumental Activities of Daily Living. *The Gerontologist*, 9(3_Part_1):179–186.

Lazzaro, V. D. (2018). Emergence of neurophysiological biomarkers of Alzheimer disease. *J Neurol Neurosurg Psychiatry*, pages jnnp–2018–318867.

LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *Nature*, 521(7553):436–444.

Leuzy, A., Mattsson-Carlgren, N., Palmqvist, S., Janelidze, S., Dage, J. L., and Hansson, O. (2022). Blood-based biomarkers for Alzheimer's disease. *EMBO Molecular Medicine*, 14(e14408):1–15.

Lewis, T. J. and Trempe, C. L. (2014). *The End of Alzheimer's*. Tjlphd, LLC.

Li, F., Cheng, D., and Liu, M. (2017). Alzheimer's disease classification based on combination of multi-model convolutional networks. In *2017 IEEE International Conference on Imaging Systems and Techniques (IST)*, pages 1–5.

Li, R., Wang, X., Lawler, K., Garg, S., Bai, Q., and Alty, J. (2022). Applications of artificial intelligence to aid early detection of dementia: A scoping review on current capabilities and future directions. *Journal of Biomedical Informatics*, 127(104030):1–13.

Li, R., Zhang, W., Suk, H.-I., Wang, L., Li, J., Shen, D., and Ji, S. (2014). Deep Learning Based Imaging Data Completion for Improved Brain Disease Diagnosis. *Medical image computing and computer-assisted intervention : MICCAI ... International Conference on Medical Image Computing and Computer-Assisted Intervention*, 17(0 3):305–312.

Liampas, I., Folia, V., Morfakidou, R., Siokas, V., Yannakoulia, M., Sakka, P., Scarmeas, N., Hadjigeorgiou, G., Dardiotis, E., and Kosmidis, M. H. (2022). Language Differences Among Individuals with Normal Cognition, Amnestic and Non-Amnestic MCI, and Alzheimer's Disease. *Archives of Clinical Neuropsychology*, page acac080.

Lim, A. S. P., Gaiteri, C., Yu, L., Sohail, S., Swardfager, W., Tasaki, S., Schneider, J. A., Paquet, C., Stuss, D. T., Masellis, M., Black, S. E., Hugon, J., Buchman, A. S., Barnes, L. L., Bennett, D. A., and Jager, P. L. D. (2018). Seasonal plasticity of cognition and related biological measures in adults with and without Alzheimer disease: Analysis of multiple cohorts. *PLOS Medicine*, 15(9):e1002647.

Lista, S., Garaci, F. G., Ewers, M., Teipel, S., Zetterberg, H., Blennow, K., and Hampel, H. (2014). CSF Aβ1-42 combined with neuroimaging biomarkers in the early detection, diagnosis and prediction of Alzheimer's disease. *Alzheimer's & Dementia*, 10(3):381–392.

Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., van der Laak, J. A., van Ginneken, B., and Sánchez, C. I. (2017). A Survey on Deep Learning in Medical Image Analysis. *Medical Image Analysis*, 42:60–88.

Liu, M., Cheng, D., Wang, K., Wang, Y., and the Alzheimer's Disease Neuroimaging Initiative (2018a). Multi-Modality Cascaded Convolutional Neural Networks for Alzheimer's Disease Diagnosis. *Neuroinformatics*, 16(3-4):295–308.

Liu, M., Liu, B., Ye, Z., and Wu, D. (2023). Bibliometric analysis of electroencephalogram research in mild cognitive impairment from 2005 to 2022. *Frontiers in Neuroscience*, 17:1128851.

Liu, S., Liu, S., Cai, W., Pujol, S., Kikinis, R., and Feng, D. (2014). Early diagnosis of Alzheimer's disease with deep learning. In *2014 IEEE 11th International Symposium on Biomedical Imaging (ISBI)*, pages 1015–1018.

Liu, X., Chen, K., Wu, T., Weidman, D., Lure, F., and Li, J. (2018b). Use of multimodality imaging and artificial intelligence for diagnosis and prognosis of early stages of Alzheimer's disease. *Translational Research*, 194:56–67.

Livieratos, L. (2012). Basic Principles of SPECT and PET Imaging. In Fogelman, I., Gnanasegaran, G., and van der Wall, H., editors, *Radionuclide and Hybrid Bone Imaging*, pages 345–359. Springer, Berlin, Heidelberg.

Lobo, J. M., Jiménez-Valverde, A., and Real, R. (2008). AUC: A misleading measure of the performance of predictive distribution models. *Global Ecology and Biogeography*, 17(2):145–151.

Logan, R., Williams, B. G., Ferreira da Silva, M., Indani, A., Schcolnicov, N., Ganguly, A., and Miller, S. J. (2021). Deep Convolutional Neural Networks With Ensemble Learning and Generative Adversarial Networks for Alzheimer's Disease Image Data Classification. *Frontiers in Aging Neuroscience*, 13(720226):1–12.

López, V., Fernández, A., Moreno-Torres, J. G., and Herrera, F. (2012). Analysis of preprocessing vs. cost-sensitive learning for imbalanced classification. Open problems on intrinsic data characteristics. *Expert Systems with Applications*, 39(7):6585–6608.

Lopez-Bastida, J., Serrano-Aguilar, P., Perestelo-Perez, L., and Oliva-Moreno, J. (2006). Social-economic costs and quality of life of Alzheimer disease in the Canary Islands, Spain. *Neurology*, 67(12):2186–2191.

Magnin, B., Mesrob, L., Kinkingnéhun, S., Pélégrini-Issac, M., Colliot, O., Sarazin, M., Dubois, B., Lehéricy, S., and Benali, H. (2009). Support vector machine-based classification of Alzheimer's disease from whole-brain anatomical MRI. *Neuroradiology*, 51(2):73–83.

Mahley, R. W., Weisgraber, K. H., and Huang, Y. (2006). Apolipoprotein E4: A causative factor and therapeutic target in neuropathology, including Alzheimer's disease. *Proceedings of the National Academy of Sciences*, 103(15):5644–5651.

Mahoney, F. I. and Barthel, D. W. (1965). Functional Evaluation: The Barthel Index. *Maryland State Medical Journal*, 14:61–65.

Malik, I., Iqbal, A., Gu, Y. H., and Al-antari, M. A. (2024). Deep learning for alzheimer's disease prediction: A comprehensive review. *Diagnostics 2024, Vol. 14, Page 1281*, 14(12):1281.

Mantzavinos, V. and Alexiou, A. (2017). Biomarkers for Alzheimer's Disease Diagnosis. *Current Alzheimer Research*, 14(11):1149–1154.

Marshall, E. and Russell, J. (2016). The Statistics Tutor's Quick Guide to Commonly Used Statistical Tests. Quick guide, University of Sheffield.

Martí-Nicolovius, M. and Arévalo-García, R. (2018). Envejecimiento y memoria: efectos de la restricción calórica. *Revista de Neurología*, 66(12):415–422.

Martinetz, T. (1993). Competitive Hebbian Learning Rule Forms Perfectly Topology Preserving Maps. In Gielen, S. and Kappen, B., editors, *ICANN '93*, pages 427–434. Springer London.

Martinetz, T. and Schulten, K. (1991). A "neural-gas" network learns topologies. In *Artificial Neural Networks*, pages 397–402. Elsevier Science Publishers B.V.

Marusteri, M. and Bacarea, V. (2010). Comparing groups for statistical differences: How to choose the right statistical test? *Biochemia Medica*, 20(1):15–32.

Maruta, C., Guerreiro, M., de Mendonça, A., Hort, J., and Scheltens, P. (2011). The use of neuropsychological tests across Europe: The need for a consensus in the use of assessment tools for dementia. *European Journal of Neurology*, 18(2):279–285.

Mathotaarachchi, S., Pascoal, T. A., Shin, M., Benedet, A. L., Kang, M. S., Beaudry, T., Fonov, V. S., Gauthier, S., and Rosa-Neto, P. (2017). Identifying incipient dementia individuals using machine learning and amyloid imaging. *Neurobiology of Aging*, 59:80–90.

MathWorks (2025). Pretrained Deep Neural Networks.

Mazarakis, N. D., Edwards, A. D., and Mehmet, H. (1997). Apoptosis in neural development and disease. *Archives of Disease in Childhood - Fetal and Neonatal Edition*, 77(3):F165–F170.

McBride, J. C., Zhao, X., Munro, N. B., Smith, C. D., Jicha, G. A., Hively, L., Broster, L. S., Schmitt, F. A., Kryscio, R. J., and Jiang, Y. (2014). Spectral and complexity analysis of scalp EEG characteristics for mild cognitive impairment and early Alzheimer's disease. *Computer Methods and Programs in Biomedicine*, 114(2):153–163.

McCrory, M. N. F. and Thomas, S. A. (2025). Cluster Metric Sensitivity to Irrelevant Features. In Mastorakis, N. E., Rudas, I. J., and Shmaliy, Y. S., editors, *Computational Problems in Science and Engineering II*, pages 85–95. Springer Nature Switzerland, Cham.

McDonald, J. H. (2014). Choosing the right statistical test. In *Handbook of Biological Statistics*. Sparky House Publishing, Baltimore, Maryland, 3rd edition.

McKhann, G., Drachman, D., Folstein, M., Katzman, R., Price, D., and Stadlan, E. M. (1984). Clinical diagnosis of Alzheimer's disease: Report of the NINCDS-ADRDA Work Group* under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. *Neurology*, 34(7):939–939.

McKhann, G. M., Knopman, D. S., Chertkow, H., Hyman, B. T., Jack, C. R., Kawas, C. H., Klunk, W. E., Koroshetz, W. J., Manly, J. J., Mayeux, R., Mohs, R. C., Morris, J. C., Rossor, M. N., Scheltens, P., Carrillo, M. C., Thies, B., Weintraub, S., and Phelps, C. H. (2011). The diagnosis of dementia due to Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's & Dementia*, 7(3):263–269.

Meilán, J. J. G., Martínez-Sánchez, F., Carro, J., López, D. E., Millian-Morell, L., and Arana, J. M. (2014). Speech in Alzheimer's Disease: Can Temporal and Acoustic Parameters Discriminate Dementia? *Dementia and Geriatric Cognitive Disorders*, 37(5-6):327–334.

Metz, C. E. (1978). Basic principles of ROC analysis. In *Seminars in Nuclear Medicine*, volume 8, pages 283–298. Elsevier.

Miikkulainen, R. (2010). Topology of a Neural Network. In Sammut, C. and Webb, G. I., editors, *Encyclopedia of Machine Learning*, pages 988–989. Springer.

Ming, G.-l. and Song, H. (2011). Adult Neurogenesis in the Mammalian Brain: Significant Answers and Significant Questions. *Neuron*, 70(4):687–702.

Mirzaei, G. and Adeli, H. (2022). Machine learning techniques for diagnosis of alzheimer disease, mild cognitive disorder, and other types of dementia. *Biomedical Signal Processing and Control*, 72(103293):1–13.

Mitchell, A. J. (2008). The clinical significance of subjective memory complaints in the diagnosis of mild cognitive impairment and dementia: A meta-analysis. *International Journal of Geriatric Psychiatry*, 23(11):1191–1202.

Mitchell, A. J. (2009). A meta-analysis of the accuracy of the mini-mental state examination in the detection of dementia and mild cognitive impairment. *Journal of Psychiatric Research*, 43(4):411–431.

Mitchell, A. J., McGlinchey, J. B., Young, D., Chelminski, I., and Zimmerman, M. (2009). Accuracy of specific symptoms in the diagnosis of major depressive disorder in psychiatric out-patients: Data from the MIDAS project. *Psychological Medicine*, 39(7):1107–1116.

Mitchell, A. J. and Shiri-Feshki, M. (2009). Rate of progression of mild cognitive impairment to dementia - meta-analysis of 41 robust inception cohort studies. *Acta Psychiatrica Scandinavica*, 119(4):252–265.

Mittal, S. (2020). A survey of FPGA-based accelerators for convolutional neural networks. *Neural Computing and Applications*, 32(4):1109–1139.

Mizushima, N. and Komatsu, M. (2011). Autophagy: Renovation of Cells and Tissues. *Cell*, 147(4):728–741.

Mosca, A., Sperduti, S., Pop, V., Ciavardelli, D., Granzotto, A., Punzi, M., Stuppia, L., Gatta, V., Assogna, F., Banaj, N., Piras, F., Piras, F., Caltagirone, C., Spalletta, G., and Sensi, S. L. (2018). Influence of APOE and RNF219 on Behavioral and Cognitive Features of Female Patients Affected by Mild Cognitive Impairment or Alzheimer's Disease. *Frontiers in Aging Neuroscience*, 10.

Motta, C., Lorenzo, F. D., Ponzo, V., Pellicciari, M. C., Bonnì, S., Picazio, S., Mercuri, N. B., Caltagirone, C., Martorana, A., and Koch, G. (2018). Transcranial magnetic stimulation predicts cognitive decline in patients with Alzheimer's disease. *J Neurol Neurosurg Psychiatry*, pages jnnp–2017–317879.

Mudher, A. and Lovestone, S. (2002). Alzheimer's disease – do tauists and baptists finally shake hands? *Trends in Neurosciences*, 25(1):22–26.

Mulsant, B. H. and Servan-Schreiber, E. (1988). A Connectionist Approach to the Diagnosis of Dementia. *Proceedings of the Annual Symposium on Computer Application in Medical Care*, pages 245–250.

Nakagaki, T., Yamada, H., and Tóth, Á. (2000). Maze-solving by an amoeboid organism. *Nature*, 407(6803):470–470.

Nakamura, A., Cuesta, P., Fernández, A., Arahata, Y., Iwata, K., Kuratsubo, I., Bundo, M., Hattori, H., Sakurai, T., Fukuda, K., Washimi, Y., Endo, H., Takeda, A., Diers, K., Bajo, R., Maestú, F., Ito, K., and Kato, T. (2018a). Electromagnetic signatures of the preclinical and prodromal stages of Alzheimer's disease. *Brain*, 141(5):1470–1485.

Nakamura, A., Kaneko, N., Villemagne, V. L., Kato, T., Doecke, J., Doré, V., Fowler, C., Li, Q.-X., Martins, R., Rowe, C., Tomita, T., Matsuzaki, K., Ishii, K., Ishii, K., Arahata, Y., Iwamoto, S., Ito, K., Tanaka, K., Masters, C. L., and Yanagisawa, K. (2018b). High performance plasma amyloid-β biomarkers for Alzheimer's disease. *Nature*, 554:249–254.

Nasreddine, Z. S., Phillips, N. A., Bédirian, V., Charbonneau, S., Whitehead, V., Collin, I., Cummings, J. L., and Chertkow, H. (2005). The Montreal Cognitive Assessment, MoCA: A Brief Screening Tool For Mild Cognitive Impairment. *Journal of the American Geriatrics Society*, 53(4):695–699.

National Institutes of Health (2025). Brain Basics: Know Your Brain | National Institute of Neurological Disorders and Stroke.

Navan (2014). Understanding ROC curves.

Nayak, B. and Hazra, A. (2011). How to choose the right statistical test? *Indian Journal of Ophthalmology*, 59(2):85.

Nelson, P. T., Dickson, D. W., Trojanowski, J. Q., Jack, C. R., Boyle, P. A., Arfanakis, K., Rademakers, R., Alafuzoff, I., Attems, J., Brayne, C., Coyle-Gilchrist, I. T. S., Chui, H. C., Fardo, D. W., Flanagan, M. E., Halliday, G., Hokkanen, S. R. K., Hunter, S., Jicha, G. A., Katsumata, Y., Kawas, C. H., Keene, C. D., Kovacs, G. G., Kukull, W. A., Levey, A. I., Makkinejad, N., Montine, T. J., Murayama, S., Murray, M. E., Nag, S., Rissman, R. A., Seeley, W. W., Sperling, R. A., White III, C. L., Yu, L., and Schneider, J. A. (2019).

Limbic-predominant age-related TDP-43 encephalopathy (LATE): Consensus working group report. *Brain*, 142(6):1503–1527.

Nguyen, M., Sun, N., Alexander, D. C., Feng, J., and Yeo, B. T. T. (2018). Modeling Alzheimer's disease progression using deep recurrent neural networks. In *2018 International Workshop on Pattern Recognition in Neuroimaging (PRNI)*, pages 1–4.

Nichols, E., Steinmetz, J. D., Vollset, S. E., Fukutaki, K., Chalek, J., Abd-Allah, F., Abdoli, A., Abualhasan, A., Abu-Gharbieh, E., Akram, T. T., Al Hamad, H., Alahdab, F., Alanezi, F. M., Alipour, V., Almustanyir, S., Amu, H., Ansari, I., Arabloo, J., Ashraf, T., Astell-Burt, T., Ayano, G., Ayuso-Mateos, J. L., Baig, A. A., Barnett, A., Barrow, A., Baune, B. T., Béjot, Y., Bezabhe, W. M. M., Bezabih, Y. M., Bhagavathula, A. S., Bhaskar, S., Bhattacharyya, K., Bijani, A., Biswas, A., Bolla, S. R., Boloor, A., Brayne, C., Brenner, H., Burkart, K., Burns, R. A., Cámera, L. A., Cao, C., Carvalho, F., Castro-de-Araujo, L. F. S., Catalá-López, F., Cerin, E., Chavan, P. P., Cherbuin, N., Chu, D.-T., Costa, V. M., Couto, R. A. S., Dadras, O., Dai, X., Dandona, L., Dandona, R., De la Cruz-Góngora, V., Dhamnetiya, D., Dias da Silva, D., Diaz, D., Douiri, A., Edvardsson, D., Ekholuenetale, M., El Sayed, I., El-Jaafary, S. I., Eskandari, K., Eskandarieh, S., Esmaeilnejad, S., Fares, J., Faro, A., Farooque, U., Feigin, V. L., Feng, X., Fereshtehnejad, S.-M., Fernandes, E., Ferrara, P., Filip, I., Fillit, H., Fischer, F., Gaidhane, S., Galluzzo, L., Ghashghaee, A., Ghith, N., Gialluisi, A., Gilani, S. A., Glavan, I.-R., Gnedovskaya, E. V., Golechha, M., Gupta, R., Gupta, V. B., Gupta, V. K., Haider, M. R., Hall, B. J., Hamidi, S., Hanif, A., Hankey, G. J., Haque, S., Hartono, R. K., Hasaballah, A. I., Hasan, M. T., Hassan, A., Hay, S. I., Hayat, K., Hegazy, M. I., Heidari, G., Heidari-Soureshjani, R., Herteliu, C., Househ, M., Hussain, R., Hwang, B.-F., Iacoviello, L., Iavicoli, I., Ilesanmi, O. S., Ilic, I. M., Ilic, M. D., Irvani, S. S. N., Iso, H., Iwagami, M., Jabbarinejad, R., Jacob, L., Jain, V., Jayapal, S. K., Jayawardena, R., Jha, R. P., Jonas, J. B., Joseph, N., Kalani, R., Kandel, A., Kandel, H., Karch, A., Kasa, A. S., Kassie, G. M., Keshavarz, P., Khan, M. A., Khatib, M. N., Khoja, T. A. M., Khubchandani, J., Kim, M. S., Kim, Y. J., Kisa, A., Kisa, S., Kivimäki, M., Koroshetz, W. J., Koyanagi, A., Kumar, G. A., Kumar, M., Lak, H. M., Leonardi, M., Li, B., Lim, S. S., Liu, X., Liu, Y., Logroscino, G., Lorkowski, S., Lucchetti, G., Lutzky Saute, R., Magnani, F. G., Malik, A. A., Massano, J., Mehndiratta, M. M., Menezes, R. G., Meretoja, A., Mohajer, B., Mohamed Ibrahim, N., Mohammad, Y., Mohammed, A., Mokdad, A. H., Mondello, S., Moni, M. A. A., Moniruzzaman, M., Mossie, T. B., Nagel, G., Naveed, M., Nayak, V. C., Neupane Kandel, S., Nguyen, T. H., Oancea, B., Otstavnov, N., Otstavnov, S. S., Owolabi, M. O., Panda-Jonas, S., Pashazadeh Kan, F., Pasovic, M., Patel, U. K., Pathak, M., Peres, M. F. P., Perianayagam, A., Peterson, C. B., Phillips, M. R., Pinheiro, M., Piradov, M. A., Pond, C. D., Potashman, M. H., Pottoo, F. H., Prada, S. I., Radfar, A., Raggi, A., Rahim, F., Rahman, M., Ram, P., Ranasinghe, P., Rawaf, D. L., Rawaf, S., Rezaei, N., Rezapour, A., Robinson, S. R., Romoli, M., Roshandel, G., Sahathevan, R., Sahebkar, A., Sahraian, M. A., Sathian, B., Sattin, D., Sawhney, M., Saylan, M., Schiavolin, S., Seylani, A., Sha, F., Shaikh, M. A., Shaji, K., Shannawaz, M., Shetty, J. K., Shigematsu, M., Shin, J. I., Shiri, R., Silva, D. A. S., Silva, J. P., Silva, R., Singh, J. A., Skryabin, V. Y., Skryabina, A. A., Smith, A. E., Soshnikov, S., Spurlock, E. E., Stein, D. J., Sun, J., Tabarés-Seisdedos, R., Thakur, B., Timalsina, B., Tovani-Palone, M. R., Tran, B. X., Tsegaye, G. W., Valadan Tahbaz, S., Valdez, P. R., Venketasubramanian, N., Vlassov, V., Vu, G. T., Vu, L. G., Wang, Y.-P., Wimo, A., Winkler, A. S., Yadav, L., Yahyazadeh Jabbari, S. H., Yamagishi, K., Yang, L.,

Yano, Y., Yonemoto, N., Yu, C., Yunusa, I., Zadey, S., Zastrozhin, M. S., Zastrozhina, A., Zhang, Z.-J., Murray, C. J. L., and Vos, T. (2022). Estimation of the global prevalence of dementia in 2019 and forecasted prevalence in 2050: An analysis for the Global Burden of Disease Study 2019. *The Lancet Public Health*, 7(2):e105–e125.

Niemantsverdriet, E., Feyen, B. F., Le Bastard, N., Martin, J.-J., Goeman, J., De Deyn, P. P., Bjerke, M., and Engelborghs, S. (2018). Added Diagnostic Value of Cerebrospinal Fluid Biomarkers for Differential Dementia Diagnosis in an Autopsy-Confirmed Cohort. *Journal of Alzheimer's Disease*, 63(1):373–381.

Ning, K., Chen, B., Sun, F., Hobel, Z., Zhao, L., Matloff, W., and Toga, A. W. (2018). Classifying Alzheimer's disease with brain imaging and genetic data using a neural network framework. *Neurobiology of Aging*, 68:151–158.

Niu, H., Álvarez-Álvarez, I., Guillén-Grima, F., and Aguinaga-Ontoso, I. (2017). Prevalencia e incidencia de la enfermedad de Alzheimer en Europa: metaanálisis. *Neurología*, 32(8):523–532.

Noor, M. B. T., Zenia, N. Z., Kaiser, M. S., Mamun, S. A., and Mahmud, M. (2020). Application of deep learning in detecting neurological disorders from magnetic resonance images: A survey on the detection of Alzheimer's disease, Parkinson's disease and schizophrenia. *Brain Informatics*, 7(11):1–21.

Ozkan, D., Katar, O., Ak, M., Al-antari, M. A., Ak, N. Y., Yildirim, O., Mir, H. S., Tan, R.-S., and Acharya, U. R. (2024). Deep Learning Techniques for Automated Dementia Diagnosis Using Neuroimaging Modalities: A Systematic Review (2012-2023). *IEEE Access*, pages 1–1.

Palmqvist, S., Zetterberg, H., Mattsson, N., Johansson, P., Minthon, L., Blennow, K., Olsson, M., and Hansson, O. (2015). Detailed comparison of amyloid PET and CSF biomarkers for identifying early Alzheimer disease. *Neurology*, 85(14):1240–1249.

Park, J. H. (2024). Classification of mild cognitive impairment using functional near-infrared spectroscopy-derived biomarkers with convolutional neural networks. *Psychiatry Investigation*, 21(3):294–299.

Parsapoor, M. (2023). AI-based assessments of speech and language impairments in dementia. *Alzheimer's & dementia : the journal of the Alzheimer's Association*, 19(10):4675–4687.

Patino-Ramirez, F., Boussard, A., Arson, C., and Dussutour, A. (2019). Substrate composition directs slime molds behavior. *Scientific Reports*, 9:15444.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., and Cournapeau, D. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12:2825–2830.

Pellegrini, E., Ballerini, L., Hernandez, M. d. C. V., Chappell, F. M., González-Castro, V., Anblagan, D., Danso, S., Muñoz-Maniega, S., Job, D., Pernet, C., Mair, G., MacGillivray, T. J., Trucco, E., and Wardlaw, J. M. (2018). Machine learning of neuroimaging for

assisted diagnosis of cognitive impairment and dementia: A systematic review. *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring*, 10:519–535.

Pellicer Sarmiento, F. (2024). *Análisis comparativo de la Red Ontogénica SupeRGNG en el diagnóstico temprano de la Enfermedad de Alzheimer*. Final Project, Universidad de Las Palmas de Gran Canaria, Las Palmas de Gran Canaria, Spain.

Pena, D., Barman, A., Suescun, J., Jiang, X., Schiess, M. C., Giancardo, L., and the Alzheimer's Disease Neuroimaging Initiative (2019). Quantifying Neurodegenerative Progression With DeepSymNet, an End-to-End Data-Driven Approach. *Frontiers in Neuroscience*, 13.

Pérez Del Pino, M. Á. (2015). *EDEVITALZH : entorno clínico virtual para ayuda al diagnóstico temprano de la enfermedad de Alzheimer y otras demencias. Uso en Telemedicina*. Final Project, Universidad de Las Palmas de Gran Canaria, Las Palmas de Gran Canaria.

Pérez del Pino, M. Á., García Báez, P., Fernández López, P., and Suárez Araujo, C. P. (2012). Self-Organizing Maps for Early Detection of Denial of Service Attacks. In *Recent Advances in Intelligent Engineering Systems*, volume 378 of *Studies in Computational Intelligence*, pages 195–219. Springer Berlin Heidelberg.

Pérez-del-Pino, M. A., García-Báez, P., Martínez-García, J. M., and Suárez-Araujo, C. P. (2014). Technological Infrastructure and Business Intelligence Strategies for the EDEVITALZH eHealth Delivery System. In Klempous, R., Nikodem, J., Jacak, W., and Chaczko, Z., editors, *Advanced Methods and Applications in Computational Intelligence*, number 6 in Topics in Intelligent Engineering and Informatics, pages 145–164. Springer International Publishing, Heidelberg.

Petersen, R. C. (2004). Mild cognitive impairment as a diagnostic entity. *Journal of Internal Medicine*, 256(3):183–194.

Pfeffer, R. I., Kurosaki, T. T., Harrah, C. H., Chance, J. M., and Filos, S. (1982). Measurement of Functional Activities in Older Adults in the Community. *Journal of Gerontology*, 37(3):323–329.

Pichot, P. and López-Ibor Aliño, J. J. (1998). *DSM-IV: manual diagnóstico y estadístico de los trastornos mentales*. Masson, Barcelona.

Polikar, R. (2006). Ensemble based systems in decision making. *IEEE Circuits and Systems Magazine*, 6(3):21–45.

Powers, D. M. (2011). Evaluation: From precision, recall and F-measure to ROC, informedness, markedness and correlation. *Journal of Machine Learning Technologies*.

Prieto Jurczynska, C., Eimil Ortiz, M., López de Silanes de Miguel, C., and Llanero Luque, M. (2011). Impacto Social de la Enfermedad de Alzheimer y otras demencias 2011.

Provost, F., Fawcett, T., and Kohavi, R. (1997). The Case Against Accuracy Estimation for Comparing Induction Algorithms. In *Proceedings of the Fifteenth International Conference on Machine Learning*.

Pulido, M. L. B., Hernández, J. B. A., Ballester, M. Á. F., González, C. M. T., Mekyska, J., and Smékal, Z. (2020). Alzheimer's disease and automatic speech analysis: A review. *Expert Systems with Applications*, 150(113213):1–19.

Qi, X., Zhou, Q., Dong, J., and Bao, W. (2023). Noninvasive automatic detection of Alzheimer's disease from spontaneous speech: A review. *Frontiers in Aging Neuroscience*, 15:1224723.

Qu, C., Zou, Y., Ma, Y., Chen, Q., Luo, J., Fan, H., Jia, Z., Gong, Q., and Chen, T. (2022). Diagnostic Performance of Generative Adversarial Network-Based Deep Learning Methods for Alzheimer's Disease: A Systematic Review and Meta-Analysis. *Frontiers in Aging Neuroscience*, 14:1–18.

Rathore, S., Habes, M., Iftikhar, M. A., Shacklett, A., and Davatzikos, C. (2017). A review on neuroimaging-based classification studies and associated feature extraction methods for Alzheimer's disease and its prodromal stages. *NeuroImage*, 155:530–548.

Rattanabannakit, C., Risacher, S. L., Gao, S., Lane, K. A., Brown, S. A., McDonald, B. C., Unverzagt, F. W., Apostolova, L. G., Saykin, A. J., and Farlow, M. R. (2016). The Cognitive Change Index as a Measure of Self and Informant Perception of Cognitive Decline: Relation to Neuropsychological Tests. *Journal of Alzheimer's Disease*, 51(4):1145–1155.

Raut, A. and Dalal, V. (2017). A Machine Learning Based Approach for Early Detection of Alzheimer's Disease by Extracting Texture and Shape Features of the Hippocampus Region from MRI Scans. *IJARCCE*, 6(6):320–325.

Readhead, B., Haure-Mirande, J.-V., Funk, C. C., Richards, M. A., Shannon, P., Haroutunian, V., Sano, M., Liang, W. S., Beckmann, N. D., Price, N. D., Reiman, E. M., Schadt, E. E., Ehrlich, M. E., Gandy, S., and Dudley, J. T. (2018). Multiscale Analysis of Independent Alzheimer's Cohorts Finds Disruption of Molecular, Genetic, and Clinical Networks by Human Herpesvirus. *Neuron*, 99(1):64–82.e7.

Reiss, A. B., de Levante Raphael, D., Chin, N. A., and Sinha, V. (2022). The physician's Alzheimer's disease management guide: Early detection and diagnosis of cognitive impairment, Alzheimer's disease and related dementia. *AIMS Public Health*, 9(4):661–689.

Reitz, C. and Mayeux, R. (2014). Alzheimer disease: Epidemiology, Diagnostic Criteria, Risk Factors and Biomarkers. *Biochemical Pharmacology*, 88(4):640–651.

Rish, I. (2001). An empirical study of the naive Bayes classifier. Technical Report RC 22230 (W0111-014), IBM.

Rojas, R. (1996). Kohonen Networks. In *Neural Networks: A Systematic Introduction*, pages 391–412. Springer-Verlag, Berlin.

Ronson, C. E., editor (2011). *Alzheimer's Diagnosis*. Neuroscience Research Progress. Nova Science Publishers.

Rosen, K. H. (2004). *Matemática discreta y sus aplicaciones*. McGraw-Hill.

Rosen, W. G., Mohs, R. C., and Davis, K. L. (1984). A new rating scale for Alzheimer's disease. *The American Journal of Psychiatry*, 141(11):1356–1364.

Rosenberg, A. and Hirschberg, J. (2007). V-Measure: A Conditional Entropy-Based External Cluster Evaluation Measure. In *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, pages 410–420, Prague.

Rosenberg, S. J., Ryan, J. J., and Prifitera, A. (1984). Rey Auditory-Verbal Learning Test performance of patients with and without memory impairment. *Journal of Clinical Psychology*, 40(3):785–787.

Rosenblatt, F. (1961). Principles of Neurodynamics. Perceptrons and the Theory of Brain Mechanisms. Technical Report VG-1196-G-8, Cornell Aeronautical Lab Inc, Buffalo, New York.

Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20:53–65.

Ruiz-Gómez, S., Gómez, C., Poza, J., Gutiérrez-Tobal, G., Tola-Arribas, M., Cano, M., and Hornero, R. (2018). Automated Multiclass Classification of Spontaneous EEG Activity in Alzheimer's Disease and Mild Cognitive Impairment. *Entropy*, 20(1):35.

Rumelhart, D. E., Hinton, G. E., and McClelland, J. L. (1986). A General Framework for Parallel Distributed Processing. In *Parallel Distributed Processing*, volume 1, pages 45–76. MIT Press, Cambridge, MA, USA.

Salazar, D. R. C., Maldonado, C. B. G., Alvarado, H. F. G., and Salazar, D. X. C. (2018). Alzheimer detection model applying artificial intelligence techniques. In *2018 13th Iberian Conference on Information Systems and Technologies (CISTI)*, pages 1–8.

Saleem, T. J., Zahra, S. R., Wu, F., Alwakeel, A., Alwakeel, M., Jeribi, F., and Hijji, M. (2022). Deep Learning-Based Diagnosis of Alzheimer's Disease. *Journal of Personalized Medicine*, 12(5):1–29.

Santabárbara, J., Sevil-Pérez, A., Olaya, B., Gracia-García, P., and López-Antón, R. (2019). Depresión tardía clínicamente relevante y riesgo de demencia: Revisión sistemática y metaanálisis de estudios prospectivos de cohortes. *Revista de Neurología*, 68(12):493–502.

Santiago Ramón y Cajal (1917). Recuerdos de mi vida. Láminas ilustrativas.

Schatz, C. J. (1992). The Developing Brain. *Scientific American*, 267(3):60–67.

Schölkopf, B., Smola, A., and Müller, K.-R. (1998a). Nonlinear Component Analysis as a Kernel Eigenvalue Problem. *Neural Computation*, 10(5):1299–1319.

Schölkopf, B., Smola, A. J., and Müller, K.-R. (1998b). Kernel Principal Component Analysis. In *Advances in Kernel Methods: Support Vector Learning*, CogNet. The MIT Press.

Schott, J. M. and Petersen, R. C. (2015). New criteria for Alzheimer's disease: Which, when and why? *Brain*, 138(5):1134–1137.

Scinto, L. F. M. and Daffner, K. R., editors (2000). *Early Diagnosis of Alzheimer's Disease*. Current Clinical Neurology. Humana Press.

Senanayake, U., Sowmya, A., and Dawes, L. (2018). Deep fusion pipeline for mild cognitive impairment diagnosis. In *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, pages 1394–1997.

Senanayake, U., Sowmya, A., Dawes, L., Kochan, N. A., Wen, W., and Sachdev, P. (2017). Deep Learning Approach for Classification of Mild Cognitive Impairment Subtypes. In *Proceedings of the 6th International Conference on Pattern Recognition Applications and Methods (ICPRAM 2017)*, pages 655–662. SCITEPRESS - Science and Technology Publications.

Shalev-Shwartz, S. and Ben-David, S. (2014). *Understanding Machine Learning: From Theory to Algorithms*. Cambridge University Press, Cambridge.

Shankar, S. and Singh, R. (2014). Demystifying statistics: How to choose a statistical test? *Indian Journal of Rheumatology*, 9(2):77–81.

Sharma, A., Vans, E., Shigemizu, D., Boroevich, K. A., and Tsunoda, T. (2019). DeepInsight: A methodology to transform a non-image data to an image for convolution neural network architecture. *Scientific Reports*, 9(1):11399.

Shastry, K. A., Vijayakumar, V., V, M. K. M., B A, M., and B N, C. (2022). Deep Learning Techniques for the Effective Prediction of Alzheimer's Disease: A Comprehensive Review. *Healthcare*, 10(10):1–22.

Shokri-Kojori, E., Wang, G.-J., Wiers, C. E., Demiral, S. B., Guo, M., Kim, S. W., Lindgren, E., Ramirez, V., Zehra, A., Freeman, C., Miller, G., Manza, P., Srivastava, T., Santi, S. D., Tomasi, D., Benveniste, H., and Volkow, N. D. (2018). β-Amyloid accumulation in the human brain after one night of sleep deprivation. *Proceedings of the National Academy of Sciences*, 115(17):4483–4488.

Siegel, L. J. (2008). Are Telomeres the Key to Aging and Cancer.

Slaby, A. (2007a). ROC Analysis with Matlab. In *Information Technology Interfaces, 2007. ITI 2007. 29th International Conference On*, pages 191–196. IEEE.

Slaby, A. (2007b). ROC Matlab: Paper and Matlab code and toolboxes.

SmartVision Europe (2017). How to choose the correct statistical test - a rough guide.

Sokolova, M., Japkowicz, N., and Szpakowicz, S. (2006). Beyond Accuracy, F-Score and ROC: A Family of Discriminant Measures for Performance Evaluation. In *AI 2006: Advances in Artificial Intelligence*, volume 4304 of *Lecture Notes in Computer Science*, pages 1015–1021. Springer.

Sosa-Marrero, A., Cabrera-León, Y., Fernández-López, P., García-Báez, P., Navarro-Mesa, J. L., and Suárez-Araujo, C. P. (2021). Detection of Alzheimer's Disease Versus Mild Cognitive Impairment Using a New Modular Hybrid Neural Network. In Rojas, I., Joya, G., and Catala, A., editors, *Advances in Computational Intelligence*, Lecture Notes in Computer Science, pages 223–235, Cham. Springer International Publishing.

Sperling, R. A., Aisen, P. S., Beckett, L. A., Bennett, D. A., Craft, S., Fagan, A. M., Iwatsubo, T., Jack Jr., C. R., Kaye, J., Montine, T. J., Park, D. C., Reiman, E. M., Rowe, C. C., Siemers, E., Stern, Y., Yaffe, K., Carrillo, M. C., Thies, B., Morrison-Bogorad, M., Wagster, M. V., and Phelps, C. H. (2011). Toward defining the preclinical stages of Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's & Dementia*, 7(3):280–292.

Suárez-Araujo, C. P., Cabrera-León, Y., Fernández-López, P., and Báez, P. G. (2024). Current trends on the early diagnosis of Alzheimer's Disease by means of neural computation methods. *International Journal of Electronics and Telecommunications*, 70(2):277–283.

Suárez Araujo, C. P., García Báez, P., Procházka, A., Rodríguez Espinosa, N., and Fernández Viadero, C. (2017). Mild Cognitive Impairment detection by Simultaneous Use of Scales: A Neural Computing Solution. *Innovation in Aging*, 1(S1):5.

Suárez-Araujo, C. P., Pérez-del-Pino, M. A., Báez, P. G., and Fernández-López, P. (2004). Clinical web environment to assist the diagnosis of Alzheimer's disease and other dementias. In *AIC'04: Proceedings of the 4th WSEAS International Conference on Applied Informatics and Communications*, volume 3, pages 2083–2088, Tenerife, Canary Islands, Spain.

Suárez Araujo, C. P., Pérez del Pino, M. Á., García Báez, P., and Fernández López, P. (2012). EDEVITALZH: Predictive, Preventive, Participatory and Personalized e-Health Platform to Assist in the Geriatrics and Neurology Clinical Scopes. In *International Conference on Computer Aided Systems Theory (EUROCAST 2011)*, number 6928 in Lecture Notes in Computer Science, pages 264–271. Springer Berlin Heidelberg.

Suk, H.-I. and Shen, D. (2013). Deep Learning-Based Feature Representation for AD/MCI Classification. In *MICCAI 2013*, volume 16, pages 583–590.

Sulaimany, S. and Safahi, Z. (2023). Visibility graph analysis for brain: Scoping review.

Swets, J. A. (1988). Measuring the Accuracy of Diagnostic Systems. *Science*, 240(4857):1285–1293.

Systat Software (2017). ROC Curves Analysis with SigmaPlot. White paper, SYSTAT Software, Inc.

Takeda, K., Okazaki, S., and Ushiyama, J. (2015). Noninvasive Techniques for Measuring Brain Activity: EEG, fMRI, and fNIRS. *The Japanese Journal of Ergonomics*, 51(6):411–419.

Tang, J., Wu, L., Huang, H., Feng, J., Yuan, Y., Zhou, Y., Huang, P., Xu, Y., and Yu, C. (2013). Back propagation artificial neural network for community Alzheimer's disease screening in China. *Neural Regeneration Research*, 8(3):270–276.

Tape, T. G. and University of Nebraska Medical Center (2014). The Area Under a ROC Curve.

Teng, E., Becker, B. W., Woo, E., Knopman, D. S., Cummings, J. L., and Lu, P. H. (2010). Utility of the Functional Activities Questionnaire for Distinguishing Mild Cognitive Impairment from Very Mild Alzheimer's Disease. *Alzheimer disease and associated disorders*, 24(4):348–353.

Terry, A. V. and Buccafusco, J. J. (2003). The Cholinergic Hypothesis of Age and Alzheimer's Disease-Related Cognitive Deficits: Recent Challenges and Their Implications for Novel Drug Development. *Journal of Pharmacology and Experimental Therapeutics*, 306(3):821–827.

Tronstad, C. and Pripp, A. H. (2014). Statistical methods for bioimpedance analysis. *Journal of Electrical Bioimpedance*, 5(1):14–27.

Tsoi, K. K. F., Jia, P., Dowling, N. M., Titiner, J. R., Wagner, M., Capuano, A. W., and Donohue, M. C. (2023). Applications of artificial intelligence in dementia research. *Cambridge Prisms: Precision Medicine*, 1:e9.

University of Minnesota (2018). Types of Statistical Tests.

University of Southern California, L. o. N. I. (2004). Alzheimer's Disease Neuroimaging Initiative (ADNI).

van der Maaten, L. and Hinton, G. (2008). Visualizing Data using t-SNE. *Journal of Machine Learning Research*, 9(86):2579–2605.

Vapnik, V. N. (1999). An overview of statistical learning theory. *IEEE Transactions on Neural Networks*, 10(5):988–999.

Vemuri, P., Gunter, J. L., Senjem, M. L., Whitwell, J. L., Kantarci, K., Knopman, D. S., Boeve, B. F., Petersen, R. C., and Jack, C. R. (2008). Alzheimer's disease diagnosis in individual subjects using structural MR images: Validation studies. *NeuroImage*, 39(3):1186–1197.

Vimbi, V., Shaffi, N., and Mahmud, M. (2024). Interpreting artificial intelligence models: A systematic review on the application of LIME and SHAP in Alzheimer's disease detection. *Brain informatics*, 11(1):10.

Voevodskaya, O., Simmons, A., Nordenskjöld, R., Kullberg, J., Ahlström, H., Lind, L., Wahlund, L.-O., Larsson, E.-M., Westman, E., and Initiative, A. D. N. (2014). The effects of intracranial volume adjustment approaches on multiple regional MRI volumes in healthy aging and Alzheimer's disease. *Frontiers in Aging Neuroscience*, 6.

Vos, S. J. B., Verhey, F., Frölich, L., Kornhuber, J., Wiltfang, J., Maier, W., Peters, O., Rüther, E., Nobili, F., Morbelli, S., Frisoni, G. B., Drzezga, A., Didic, M., van Berckel, B. N. M., Simmons, A., Soininen, H., Kłoszewska, I., Mecocci, P., Tsolaki, M., Vellas, B., Lovestone, S., Muscio, C., Herukka, S.-K., Salmon, E., Bastin, C., Wallin, A., Nordlund, A., de Mendonça, A., Silva, D., Santana, I., Lemos, R., Engelborghs, S., Van der Mussele, S., The Alzheimer's Disease Neuroimaging Initiative, Freund-Levi, Y., Wallin, Å. K., Hampel, H., van der Flier, W., Scheltens, P., and Visser, P. J. (2015). Prevalence and prognosis of Alzheimer's disease at the mild cognitive impairment stage. *Brain*, 138(5):1327–1338.

Waskom, M. L. (2021). Seaborn: Statistical data visualization. *Journal of Open Source Software*, 6(60):3021.

Wechsler, D. (1945). A Standardized Memory Scale for Clinical Use. *The Journal of Psychology*, 19(1):87–95.

Westerlund, M. L. (2005). Classification with Kohonen Self-Organizing Maps. *Soft Computing, Haskoli Islands*.

Widrow, B. and Lehr, M. (1990). 30 years of adaptive neural networks: Perceptron, Madaline, and backpropagation. *Proceedings of the IEEE*, 78(9):1415–1442.

Wimo, A., Guerchet, M., Ali, G.-C., Wu, Y.-T., Prina, A. M., Winblad, B., Jönsson, L., Liu, Z., and Prince, M. (2017). The worldwide costs of dementia 2015 and comparisons with 2010. *Alzheimer's & Dementia*, 13(1):1–7.

Wingo, T. S., Cutler, D. J., Wingo, A. P., Le, N.-A., Rabinovici, G. D., Miller, B. L., Lah, J. J., and Levey, A. I. (2019). Association of Early-Onset Alzheimer Disease With Elevated Low-Density Lipoprotein Cholesterol Levels and Rare Genetic Coding Variants of APOB. *JAMA Neurology*, 76(7):809–817.

Wong, W. Y., Wong, M. S., and Lo, K. H. (2007). Clinical Applications of Sensors for Human Posture and Movement Analysis: A Review. *Prosthetics and Orthotics International*, 31(1):62–75.

Wu, C. C., Su, C. H., Islam, M. M., and Liao, M. H. (2023). Artificial intelligence in dementia: A bibliometric study. *Diagnostics*, 13(12):2109.

Wu, L. and Zhao, L. (2016). ApoE2 and Alzheimer's disease: Time to take a closer look. *Neural Regeneration Research*, 11(3):412–413.

Yang, D., Hong, K.-S., Yoo, S.-H., and Kim, C.-S. (2019). Evaluation of Neural Degeneration Biomarkers in the Prefrontal Cortex for Early Identification of Patients With Mild Cognitive Impairment: An fNIRS Study. *Frontiers in Human Neuroscience*, 13.

Yang, D., Huang, R., Yoo, S.-H., Shin, M.-J., Yoon, J. A., Shin, Y.-I., and Hong, K.-S. (2020). Detection of Mild Cognitive Impairment Using Convolutional Neural Network: Temporal-Feature Maps of Functional Near-Infrared Spectroscopy. *Frontiers in Aging Neuroscience*, 12.

Yao, H., Zhang, N., Zhang, R., Duan, M., Xie, T., Pan, J., Peng, E., Huang, J., Zhang, Y., Xu, X., Xu, H., Zhou, F., and Wang, G. (2020). Severity Detection for the Coronavirus Disease 2019 (COVID-19) Patients Using a Machine Learning Model Based on the Blood and Urine Tests. *Frontiers in Cell and Developmental Biology*, 8:683.

Yao, X., Yan, J., Ginda, M., Börner, K., Saykin, A. J., Shen, L., and Initiative, f. t. A. D. N. (2017). Mapping longitudinal scientific progress, collaboration and impact of the Alzheimer's disease neuroimaging initiative. *PLOS ONE*, 12(11):e0186095.

Yesavage, J. A. and Sheikh, J. I. (1986). Geriatric Depression Scale (GDS): Recent Evidence and Development of a Shorter Version. *Clinical Gerontologist*, 5(1-2):165–173.

Yu, J.-T., Xu, W., Tan, C.-C., Andrieu, S., Suckling, J., Evangelou, E., Pan, A., Zhang, C., Jia, J., Feng, L., Kua, E.-H., Wang, Y.-J., Wang, H.-F., Tan, M.-S., Li, J.-Q., Hou, X.-H., Wan, Y., Tan, L., Mok, V., Tan, L., Dong, Q., Touchon, J., Gauthier, S., Aisen, P. S., and Vellas, B. (2020). Evidence-based prevention of Alzheimer's disease: Systematic review and meta-analysis of 243 observational prospective studies and 153 randomised controlled trials. *Journal of Neurology, Neurosurgery & Psychiatry*, pages jnnp–2019–321913.

Yu, L. and Liu, H. (2003). Feature Selection for High-Dimensional Data: A Fast Correlation-Based Filter Solution. In *Proceedings of the Twentieth International Conference on Machine Learning (ICML 2003)*, pages 856–863, Washington DC.

Zeng, H.-M., Han, H.-B., Zhang, Q.-F., and Bai, H. (2021). Application of modern neuroimaging technology in the diagnosis and study of Alzheimer's disease. *Neural Regeneration Research*, 16(1):73–79.

Zhang, C., Yang, H., Fan, C. C., Chen, S., Fan, C., Hou, Z. G., Chen, J., Peng, L., Xiang, K., Wu, Y., and Xie, H. (2023a). Comparing multi-dimensional fNIRS features using bayesian optimization-based neural networks for mild cognitive impairment (MCI) detection. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 31:1019–1029.

Zhang, L., Wang, M., Liu, M., and Zhang, D. (2020). A Survey on Deep Learning for Neuroimaging-Based Brain Disorder Analysis. *Frontiers in Neuroscience*, 14(779):1–19.

Zhang, S., Yang, J., Zhang, Y., Zhong, J., Hu, W., Li, C., and Jiang, J. (2023b). The combination of a graph neural network technique and brain imaging to diagnose neurological disorders: A review and outlook. *Brain Sciences 2023, Vol. 13, Page 1462*, 13(10):1462.

Zhang, T. and Yu, B. (2005). Boosting with early stopping: Convergence and consistency. *The Annals of Statistics*, 33(4).

Zhang, W., Li, Y., Ren, W., and Liu, B. (2023c). Artificial intelligence technology in Alzheimer's disease research. *Intractable and Rare Diseases Research*, 12(4):208–212.

Zhang, Z., Li, G., Xu, Y., and Tang, X. (2021). Application of Artificial Intelligence in the MRI Classification Task of Human Brain Neurological and Psychiatric Diseases: A Scoping Review. *Diagnostics*, 11(8):1–27.

Zhao, X., Ang, C. K. E., Acharya, U. R., and Cheong, K. H. (2021). Application of Artificial Intelligence techniques for the detection of Alzheimer's disease using structural MRI images. *Biocybernetics and Biomedical Engineering*, 41(2):456–473.

Zhao, Z., Chuah, J. H., Lai, K. W., Chow, C. O., Gochoo, M., Dhanalakshmi, S., Wang, N., Bao, W., and Wu, X. (2023). Conventional machine learning and deep learning in Alzheimer's disease diagnosis using neuroimaging: A review. *Frontiers in Computational Neuroscience*, 17:1038636.

Zheng, C., Xia, Y., Pan, Y., and Chen, J. (2015). Automated identification of dementia using medical imaging: A survey from a pattern classification perspective. *Brain Informatics*, 3(1):17–27.

Zhou, Q., Goryawala, M., Cabrerizo, M., Barker, W., Loewenstein, D., Duara, R., and Adjouadi, M. (2014). Multivariate Analysis of structural MRI and PET (FDG and 18F-AV-45) for Alzheimer's disease and its prodromal stages. In *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 1051–1054.

Zhu, C. W. and Sano, M. (2006). Economic considerations in the management of Alzheimer's disease. *Clinical Interventions in Aging*, 1(2):143–154.

Zhu, X. (2005). Semi-supervised learning literature survey. *Citeseer*, pages 1–60.

Zhu, X. (2011). Semi-supervised learning. In *Encyclopedia of Machine Learning*, pages 892–897. Springer US.

Zweig, M. H. and Campbell, G. (1993). Receiver-operating characteristic (ROC) plots: A fundamental evaluation tool in clinical medicine. *Clinical Chemistry*, 39(4):561–577.

Publications and communications

In this chapter the lists of publications and communications derived from the developments carried out in this PhD thesis are shown.

Publications

Ylermi Cabrera-León, Patricio García Báez, Pablo Fernández-López, and Carmen Paz Suárez-Araujo. Neural computation methods for the early diagnosis and prognosis of Alzheimer's Disease based on neuroimaging. A Systematic Review. (UNDER REVIEW) 2025.

Ylermi Cabrera-León, Pablo Fernández-López, Patricio García Báez, Konrad Kluwak, Juan Luis Navarro-Mesa, Carmen Paz Suárez-Araujo, and for the Alzheimer's Disease Neuroimaging Initiative. Toward an intelligent computing system for the early diagnosis of Alzheimer's disease based on the modular hybrid growing neural gas. *Digital Health*, 10: 1–16, October 2024a. ISSN 2055-2076. doi: 10.1177/20552076241284349.

Carmen Paz Suárez-Araujo, Ylermi Cabrera-León, Pablo Fernández-López, and Patricio García Báez. Current trends on the early diagnosis of Alzheimer's Disease by means of neural computation methods. *International Journal of Electronics and Telecommunications*, 70(2):277–283, June 2024. ISSN 2300-1933. doi: 10.24425/ijet.2024.149542.

Ylermi Cabrera-León, Patricio García Báez, Pablo Fernández-López, and Carmen Paz Suárez-Araujo. Neural Computation-Based Methods for the Early Diagnosis and Prognosis of Alzheimer's Disease Not Using Neuroimaging Biomarkers: A Systematic Review. *Journal of Alzheimer's Disease*, 98(3):793–823, April 2024b. ISSN 18758908. doi: 10.3233/JAD-231271.

Carmen Paz Suárez-Araujo, Patricio García Báez, Ylermi Cabrera-León, Ales Prochazka, Norberto Rodríguez Espinosa, Carlos Fernández Viadero, and for the Alzheimer's Disease Neuroimaging Initiative. A Real-Time Clinical Decision Support System, for Mild Cognitive Impairment Detection, Based on a Hybrid Neural Architecture. *Computational and Mathematical Methods in Medicine*, 2021(e5545297):1–9, June 2021. ISSN 1748-670X. doi: 10.1155/2021/5545297.

C. P. Suárez-Araujo, Y. Cabrera-León, C. Fernández-Viadero, N. Rodríguez-Espinosa, and P. García Báez. Solución computacional para la detección del DCL y del DCL temprano basada en el aprendizaje automático. *Revista Española de Geriatría y Gerontología*, 53 (Supplement 1):57–57, June 2018. ISSN 0211-139X. doi: 10.1016/j.regg.2018.04.138.

Communications

Ylermi Cabrera-León, Pablo Fernández-López, Patricio García Báez, and Carmen Paz Suárez-Araujo. Early diagnosis of Alzheimer's disease based on the Supervised Reconfigurable Growing Neural Gas. Towards an eHealth solution. In *1st Congress Bridge to Africa*, pages 171–176, Las Palmas de Gran Canaria, May 2024. Servicio de Publicaciones y Difusión Científica de la Universidad de Las Palmas de Gran Canaria. ISBN 978-84-9042-527-5.

Ylermi Cabrera-León, Patricio García Báez, Pablo Fernández-López, and Carmen Paz Suárez-Araujo. Study on Mild Cognitive Impairment and Alzheimer's Disease Classification using a New Ontogenic Neural Architecture, the Supervised Reconfigurable Growing Neural Gas. In *2023 Annual Modeling and Simulation Conference (ANNSIM 2023)*, pages 425–436, Mohawk College, ON, Canada, May 2023. IEEE. ISBN 978-1-7138-7328-0.

Alberto Sosa-Marrero, Ylermi Cabrera-León, Pablo Fernández-López, Patricio García-Báez, Juan Luis Navarro-Mesa, and Carmen Paz Suárez-Araujo. Detection of Alzheimer's Disease Versus Mild Cognitive Impairment Using a New Modular Hybrid Neural Network. In Ignacio Rojas, Gonzalo Joya, and Andreu Catala, editors, *Advances in Computational Intelligence*, Lecture Notes in Computer Science, pages 223–235, Cham, August 2021. Springer International Publishing. ISBN 978-3-030-85099-9. doi: 10.1007/978-3-030-85099-9_18.

Ylermi Cabrera-León, Patricio García Báez, and Carmen Paz Suárez-Araujo. Neural computing and Deep Learning solutions for early diagnosis of Alzheimer's Disease. Trends in diagnostic methods. In *EUROCAST 2019*, Las Palmas de Gran Canaria, Spain, February 2019. ISBN 978-84-09-09208-6.

Ylermi Cabrera-León, Patricio García Báez, Juan Ruiz-Alzola, and Carmen Paz Suárez-Araujo. Classification of Mild Cognitive Impairment Stages Using Machine Learning Methods. In *2018 IEEE 22nd International Conference on Intelligent Engineering Systems (INES)*, pages 67–72, Las Palmas de Gran Canaria, Spain, June 2018. IEEE. ISBN 978-1-5386-1122-7. doi: 10.1109/INES.2018.8523858.