



Universidad de Córdoba

**Máster en Inteligencia Computacional e Internet de las
Cosas**

Trabajo Fin de Máster

**Análisis Exploratorio del Desarrollo Infantil en la Primera
Infancia con Técnicas de Aprendizaje Automático**

Alberto Fernández Merchán

**Tutores: Dr. Enrique Yeguas Bolívar y
José Manuel Alcalde Llergo**

Convocatoria de Junio, Curso 2024/25



**MÁSTER UNIVERSITARIO EN INTELIGENCIA
COMPUTACIONAL E INTERNET DE LAS COSAS**



TRABAJO FIN DE MÁSTER

AUTORIZACIÓN DE LA PRESENTACIÓN DEL TRABAJO FIN DE MÁSTER

ENRIQUE YEGUAS , Profesor/a TITULAR , del Departamento CCIA de la Universidad de Córdoba,

JOSE MANUEL ALCALDE LLERGO , Profesor/a , del Departamento CCIA de la Universidad de Córdoba,

autoriza(n) a

ALBERTO FERNÁNDEZ
D./Dña. MERCHÁN , con documento de identificación n.º 49114865K

"Análisis Exploratorio del Desarrollo Infantil en la Primera Infancia con Técnicas de Aprendizaje Automático"

a presentar el trabajo fin de máster , como Trabajo fin de Máster del Máster en Inteligencia Computacional e Internet de las Cosas de la Universidad de Córdoba

Solicita(n) se admitido para su defensa en la convocatoria:

Curso: 2024/2025

Ordinaria: Junio [X] Septiembre []

Extraordinaria: Diciembre [] Abril []

Rdo.: ENRIQUE YEGUAS BOLÍVAR y JOSÉ MANUEL ALCALDE LLERGO

Fecha: 11 de JULIO de 2025

(El tutor/a/es/as del Trabajo fin de Máster)

ANEXO 3**TRABAJO FIN DE MÁSTER****INFORME DEL TUTOR/A DEL TRABAJO FIN DE MÁSTER****Informe de los principales logros del Trabajo Fin de Máster**

Este TFM realiza un análisis exploratorio de los datos de niños y niñas normotípicos recogidos por el Centro de Atención Infantil Temprana. Los datos integran variables fundamentales del desarrollo infantil como son el lenguaje, motricidad fina, motricidad gruesa y las dimensiones social, cognitiva desde los 0 meses hasta los siete años de edad, agrupándose según la edad por meses o grupos de meses.

La detección temprana de los Trastornos del Neurodesarrollo permite intervenciones que pueden influir en el desarrollo infantil. Este estudio aplica técnicas de aprendizaje automático para analizar patrones de crecimiento e identificar desviaciones. Utilizando datos de 638 niños españoles con desarrollo neurotípico (de entre 1 y 84 meses de edad), segmentamos la población por edad y aplicamos agrupamiento para explorar tendencias del desarrollo. Además, evaluamos modelos de aprendizaje automático para refinar la selección de variables y mejorar la clasificación.

Los resultados contribuyen a la detección y el cribado de trastornos neuronales del desarrollo, ofreciendo un enfoque basado en datos para apoyar la toma de decisiones en la atención a la primera infancia.

Se utiliza análisis de clústeres, reducción de dimensionalidad y análisis de correlación. Se ha logrado reducir la dimensionalidad del conjunto de datos, manteniendo la varianza más relevante, lo que mejoró la visualización e interpretación de la estructura de los datos. Sin embargo, los métodos de agrupamiento no produjeron grupos bien definidos, lo que podría explicarse en parte por el tamaño limitado del conjunto de datos actual, especialmente en ciertos rangos de edad donde la escasez de datos limita la capacidad para identificar patrones estables y generalizables.

No obstante, los análisis de correlación revelaron asociaciones significativas entre los dominios del desarrollo, particularmente en el grupo de edad de 60 a 71 meses. Estos resultados destacan la interconexión entre las habilidades adaptativas, cognitivas, motoras finas y del lenguaje, respaldando la noción de interacciones entre dominios durante el desarrollo temprano.

Los resultados están siendo de gran utilidad para el Centro de Atención Infantil Temprana y de ahí la publicación conjunta de un artículo para el Congreso MetroXRaine 2025.

Fdo.: Enrique Yeguas Bolívar**Fecha:** 11 de Julio de 2025

(El tutor/a del Trabajo fin de Máster)

Fdo.: José Manuel Alcalde Llergo**Fecha:** 11 de Julio de 2025

(El tutor/a del Trabajo fin de Máster)

AGRADECIMIENTOS

Me gustaría agradecer a mis tutores, Enrique y Pepe, por su apoyo, dedicación y orientación a lo largo de mi Trabajo Fin de Máster. Gracias por haberme tratado con tanta cercanía y compromiso durante todo el curso. Su disponibilidad constante y su forma de implicarse con cada detalle han sido fundamentales para poder llevar a cabo este proyecto.

También quiero agradecer a mis compañeros del máster, en especial a David y a Marcos, por haber hecho este camino más llevadero y enriquecedor. Las conversaciones, las dudas compartidas y los ánimos en los momentos difíciles han marcado la diferencia. Ha sido un regalo recorrer esta etapa junto a ellos.

RESUMEN

La detección temprana de los Trastornos del Neurodesarrollo (TND) es fundamental para implementar intervenciones que favorezcan el desarrollo infantil y mejoren el pronóstico a largo plazo. Este estudio aplica técnicas de aprendizaje automático para analizar patrones de crecimiento y detectar posibles desviaciones evolutivas en niños normotípicos. Para ello, se emplea un conjunto de datos de 639 niños españoles, con edades comprendidas entre 1 y 84 meses, evaluados mediante el cuestionario DAUCO en áreas clave del desarrollo: lenguaje, cognitiva, motricidad fina, motricidad gruesa, adaptativa y social.

Con el fin de optimizar el análisis, la muestra se segmentó por rangos de edad, identificando que la franja de 60 a 71 meses era la más representativa. Sobre este grupo, se aplicaron técnicas de reducción de dimensionalidad y algoritmos de *clustering* no supervisado para explorar trayectorias del desarrollo. Además, se evaluó la estabilidad y la calidad de los agrupamientos utilizando métricas internas como el coeficiente de *Silhouette*, el índice de *Davies-Bouldin* y el índice de *Calinski-Harabasz*, así como métricas externas (*Adjusted Random Index*, *Normalized Mutual Information*).

Los resultados muestran correlaciones significativas entre ítems de distintas áreas del desarrollo y permiten identificar configuraciones óptimas para los algoritmos de agrupamiento. Este enfoque basado en datos contribuye al cribado y detección temprana de TND, proporcionando una herramienta complementaria para la toma de decisiones clínicas en el ámbito de la Atención Temprana.

Palabras clave: Aprendizaje automático, Agrupamiento, Atención temprana, Desarrollo infantil, Trastornos del neurodesarrollo.

ABSTRACT

Early identification of Neurodevelopmental Disorders (NDDs) is essential for implementing interventions that support child development and improve long-term outcomes. This study applies machine learning techniques to analyze developmental patterns and explore potential deviations in normotypical children. The dataset includes 639 Spanish children aged between 1 and 84 months, assessed using the DAUCO questionnaire across key developmental domains: language, cognition, fine motor, gross motor, adaptive, and social skills.

To optimize the analysis, the sample was segmented by age groups, identifying the 60–71 months range as the most representative. Dimensionality reduction techniques and unsupervised clustering algorithms were applied to this group to explore developmental trajectories. The stability and quality of the clusters were evaluated using internal metrics such as the Silhouette coefficient, Davies-Bouldin index, and Calinski-Harabasz index, as well as external metrics like Adjusted Rand Index and Normalized Mutual Information.

The results reveal significant correlations between items from different developmental domains and help identify optimal configurations for clustering algorithms. This data-driven approach contributes to the exploration of developmental patterns and may support future clinical decision-making in Early Childhood Intervention contexts.

Keywords: Machine learning, Clustering, Early intervention, Child development, Neurodevelopmental disorders.

ÍNDICE GENERAL

1. Introducción	1
1.1. Contexto	1
1.2. Estado del arte	2
1.3. Limitaciones de los métodos tradicionales	3
1.4. Planteamiento del problema	4
1.5. Objetivos	4
2. Metodología	7
2.1. Recogida de datos mediante la aplicación DAUCO	7
2.2. Descripción del conjunto de datos	11
2.3. Preprocesamiento	11
2.4. Ejemplos de ítems del grupo de edad 60–71 meses	16
2.5. Técnicas de análisis	19
2.6. Implementación	20
2.7. Métricas de evaluación	20
3. Resultados y Análisis	23
3.1. Exploración de los datos	23
3.2. Resultados del agrupamiento	25
3.3. Impacto del preprocesamiento con PCA	31
3.4. Interpretación exploratoria de los resultados	32
4. Conclusiones y trabajo futuro	35
4.1. Comparación entre algoritmos	35

4.2. Limitaciones del enfoque	36
4.3. Implicaciones prácticas y aplicaciones futuras	36
5. Contribuciones	39
Referencias bibliográficas	40
I Apéndices	45
A. Artículo MetroXRaine	47

ÍNDICE DE FIGURAS

2.1. Diagrama de flujo del proceso de análisis en el DAUCO Test	10
2.2. Distribución de individuos por grupo de edad en el conjunto de datos. .	13
2.3. Distribución de valores «no se puede evaluar» en el grupo de edad de 60 a 71 meses	14
3.1. Matriz de correlación entre ítems del test DAUCO en el grupo de edad de 60-71 meses.	24
3.2. Estudio de la evolución del coeficiente de silueta en Spectral Clustering.	26
3.3. Matrices de estabilidad en Spectral Clustering	27
3.4. Estudio del coeficiente de silueta en Clustering Aglomerativo.	27
3.5. Matrices de estabilidad en Clustering Aglomerativo	28
3.6. Estudio de la evolución del coeficiente de silueta en Gaussian Mixture Models.	29
3.7. Matrices de estabilidad en Gaussian Mixture Models	30

ÍNDICE DE TABLAS

2.1. Número de ítems únicos por área del desarrollo en el test DAUCO . . .	8
2.2. Codificación de los grupos de edad utilizados en el análisis.	12
2.3. Número de ítems por área funcional en cada grupo de edad	15
2.4. Ejemplos de ítems del área de Lenguaje (60–71 meses)	16
2.5. Ejemplos de ítems del área Social (60–71 meses)	17
2.6. Ejemplos de ítems del área Cognitiva (60–71 meses)	18
2.7. Ejemplos de ítems del área Adaptativa (60–71 meses)	18
2.8. Ejemplos de ítems del área de Motricidad Gruesa (60–71 meses)	19
2.9. Ejemplos de ítems del área de Motricidad Fina (60–71 meses)	19
3.1. Tabla de métricas promedio en rendimiento y estabilidad.	30
3.2. Tabla comparativa de resultados del clustering con PCA y sin PCA . . .	32

INTRODUCCIÓN

Este capítulo contextualiza el problema de investigación en el ámbito de los TNDs, exponiendo la importancia de su detección temprana y las limitaciones de los enfoques diagnósticos tradicionales. Se revisa el estado del arte en el uso de técnicas de inteligencia artificial aplicadas al análisis del desarrollo infantil y se plantea el enfoque metodológico de este trabajo, así como los objetivos generales y específicos que guían el estudio.

1.1 CONTEXTO

Los Trastornos del Neurodesarrollo (TND) constituyen un conjunto de condiciones de origen temprano que afectan de forma persistente a la adquisición y funcionamiento de diversas habilidades, interfiriendo significativamente en el desarrollo personal, social y académico de los niños. Estas alteraciones impactan en áreas clave como el lenguaje, la cognición, la atención, la coordinación motora, la autorregulación emocional o la interacción social, dando lugar a trayectorias evolutivas atípicas que pueden mantenerse o amplificarse con el tiempo si no se detectan de forma precoz.

Según el *Manual Diagnóstico y Estadístico de los Trastornos Mentales* (DSM-5) y la *Clasificación Internacional de Enfermedades* (CIE-11), los TND comprenden entidades como el trastorno del espectro autista (TEA), el trastorno por déficit de atención con hiperactividad (TDAH), la discapacidad intelectual, los trastornos del lenguaje, los trastornos motores y del aprendizaje, entre otros ([American Psychiatric Association, 2013](#)). Estas condiciones presentan una elevada heterogeneidad clínica y suelen coexistir entre sí, dificultando su identificación mediante modelos diagnósticos tradicionales.

En respuesta a esta complejidad, se ha propuesto el abandono de enfoques categoriales en favor de modelos dimensionales, que interpretan el desarrollo como un continuo y permiten representar con mayor precisión la variabilidad y superposición sintomática de los TND ([Morris-Rosendahl and Crocq, 2020](#)). No obstante, la detección temprana sigue representando un reto clínico debido a la falta de consenso en los criterios diagnósticos y a las limitaciones de las herramientas disponibles ([Reilly et al., 2014](#)).

La atención temprana, centrada en la etapa de 0 a 6 años, busca responder de forma ajustada a las señales de alerta que puedan anticipar una alteración en el desarrollo ([Temprana, 2000](#)). Sin embargo, las herramientas de evaluación tradicionales, como las

escalas estandarizadas, presentan limitaciones que afectan a su sensibilidad y validez: segmentación arbitraria por edad, débil adaptación transcultural, y escasa fundamentación en modelos teóricos sólidos (Muñiz, 2010; Silva et al., 2018; Visser et al., 2012; Luque de Dios et al., 2023).

En este contexto, ha crecido el interés por aplicar enfoques alternativos basados en inteligencia artificial. En particular, las técnicas no supervisadas, como los algoritmos de *clustering*, permiten identificar agrupaciones naturales en los datos sin imponer categorías clínicas predefinidas. Cuando se aplican a herramientas como el test DAUCO (Luque de Dios et al., 2025), que recoge información estructurada sobre el desarrollo infantil entre los 1 y 84 meses, pueden facilitar la caracterización de perfiles evolutivos diferenciados. Estudios recientes respaldan esta aproximación como vía para enriquecer los procesos de cribado y orientación diagnóstica (Cuppens et al., 2023).

1.2 ESTADO DEL ARTE

Los TND constituyen un área de alta complejidad clínica, lo que ha impulsado un cambio de paradigma desde modelos categoriales hacia enfoques dimensionales más flexibles (Morris-Rosendahl and Crocq, 2020; Reilly et al., 2014). Esta transición ha puesto en cuestión la utilidad de las escalas clásicas de evaluación del desarrollo, señalando limitaciones como la escasa sensibilidad ante perfiles intermedios, la segmentación rígida por edades, o la baja representatividad cultural (Luque de Dios et al., 2023; Silva et al., 2018; Visser et al., 2012). Incluso instrumentos recientes como el test DAUCO requieren estudios adicionales de validación psicométrica (Luque de Dios et al., 2025).

Frente a estos desafíos, la Inteligencia Artificial ofrece nuevas posibilidades para el análisis de datos del desarrollo infantil. El aprendizaje automático supervisado ha sido utilizado para entrenar modelos diagnósticos a partir de bases de datos etiquetadas, pero su aplicabilidad en edades tempranas es limitada por la escasez de diagnósticos fiables y la opacidad de muchos modelos (Vincent et al., 2024). En cambio, los métodos no supervisados, como los algoritmos de *clustering*, permiten explorar patrones intrínsecos en los datos sin necesidad de etiquetas previas. Estos enfoques han demostrado su utilidad para detectar perfiles evolutivos diferenciados en muestras tanto clínicas como en riesgo (Toki et al., 2024; Song et al., 2022; Cuppens et al., 2023; Rivard et al., 2023; Mayer et al., 2022), así como para mejorar la generalización de modelos diagnósticos mediante técnicas de adaptación de dominio y análisis de neuroimagen (Mengi

and Malhotra, 2024; Wang et al., 2020; Chavas et al., 2022).

Por otro lado, el *clustering* ha permitido detectar patrones ocultos en poblaciones aparentemente homogéneas, agrupando niños según perfiles fenotípicos o de desarrollo. En esta línea, (Cuppens et al., 2023) utilizaron técnicas de *clustering* jerárquico y *k-means* sobre el conjunto de pacientes *Deciphering Developmental Disorders* (DDD) para identificar subgrupos con características clínicas compartidas, con implicaciones importantes en el diseño de ensayos clínicos tipo «basket trials». (Rivard et al., 2023) también aplicaron análisis no supervisado a niños con TEA para extraer perfiles de desarrollo diferencial en las dimensiones de comunicación y socialización. Más allá de los límites diagnósticos tradicionales, algunos trabajos adoptan un enfoque trans-diagnóstico para explorar la heterogeneidad cognitiva compartida entre TEA, TDAH y esquizofrenia (Al-Saoud et al., 2024; Chien et al., 2022).

Respecto a la reducción de dimensionalidad y selección de características, estos métodos han resultado fundamentales para filtrar los ítems más relevantes en escalas extensas de evaluación del desarrollo infantil. Algunos trabajos emplean técnicas clásicas como PCA o LASSO, mientras que otros optan por enfoques más recientes como *autoencoders* o métodos basados en importancia de variables. En este sentido, (Rahman et al., 2020) revisan múltiples estrategias de selección de características aplicadas a TEA, destacando su utilidad para reducir la dimensionalidad y mejorar el rendimiento de los modelos.

En conjunto, estos trabajos reflejan un creciente interés por integrar la inteligencia artificial en la detección precoz de los TND, aportando eficiencia, escalabilidad y objetividad a procesos tradicionalmente manuales y dependientes del juicio clínico. No obstante, aún existen desafíos relacionados con la generalización de los modelos, la calidad de los datos disponibles y la necesidad de validaciones clínicas en contextos reales (Bastiaansen et al., 2022).

1.3 LIMITACIONES DE LOS MÉTODOS TRADICIONALES

A pesar de su amplia utilización, las escalas estandarizadas presentan carencias importantes. Muchos instrumentos no se ajustan adecuadamente a la diversidad cultural, lingüística o socioeconómica de la población a la que se aplican, y tienden a basarse en divisiones arbitrarias por edad que no reflejan la continuidad real del desarrollo (Visser et al., 2012; Luque de Dios et al., 2025). Además, dependen en exceso del juicio clínico, lo que introduce una elevada carga subjetiva y puede dificultar la detección

de casos leves o atípicos (Reilly et al., 2014). Algunas escalas también muestran debilidades estructurales que limitan su capacidad explicativa y su alineación con modelos contemporáneos del desarrollo.

Estas limitaciones justifican la exploración de métodos complementarios como los análisis no supervisados, que permiten generar agrupaciones de forma empírica y sin necesidad de aplicar categorías diagnósticas predefinidas (Cuppens et al., 2023).

1.4 PLANTEAMIENTO DEL PROBLEMA

La identificación temprana de los TND es fundamental para activar intervenciones eficaces en los periodos más sensibles del desarrollo. Sin embargo, los procedimientos actuales presentan limitaciones considerables: ausencia de consenso en los criterios diagnósticos, alta variabilidad en la manifestación de síntomas, y una fuerte dependencia de observaciones clínicas subjetivas (Reilly et al., 2014).

Las escalas tradicionales, aunque estructuradas, no siempre permiten detectar señales tempranas de desviación evolutiva, especialmente cuando los perfiles no encajan claramente en categorías clínicas convencionales (Luque de Dios et al., 2025). Esto restringe su utilidad como herramienta de cribado temprano.

En este contexto, se plantea la posibilidad de aplicar técnicas de aprendizaje automático no supervisado a datos derivados de instrumentos estructurados como el test DAUCO. El objetivo es explorar si es posible identificar agrupaciones de niños con trayectorias evolutivas diferenciadas sin recurrir a etiquetas diagnósticas. Además, se propone el uso de PCA como técnica de reducción de dimensionalidad para simplificar el espacio de variables, reducir la redundancia y facilitar el proceso de agrupamiento posterior.

Esta aproximación busca complementar la evaluación clínica tradicional mediante un enfoque exploratorio basado en datos, que permita avanzar hacia sistemas de cribado más sensibles, empíricamente fundamentados y clínicamente útiles.

1.5 OBJETIVOS

El objetivo general de este trabajo es aplicar técnicas de aprendizaje automático no supervisado al análisis de datos procedentes de evaluaciones del desarrollo infantil, con el fin de explorar si es posible identificar agrupaciones de niños que reflejen tra-

yectorias evolutivas diferenciadas. Esta aproximación busca contribuir al desarrollo de herramientas analíticas que apoyen la detección temprana de posibles alteraciones en el desarrollo, complementando la información aportada por las escalas tradicionales.

Para alcanzar este propósito, se plantean los siguientes objetivos específicos:

- En primer lugar, se realizará un análisis exploratorio de la estructura interna del conjunto de datos, empleando medidas de correlación entre ítems, con el fin de identificar relaciones significativas entre distintas áreas del desarrollo.
- A continuación, se aplicarán técnicas de reducción de dimensionalidad, como el PCA, con el objetivo de simplificar el espacio de características, minimizar la redundancia y facilitar el proceso de agrupamiento.
- Posteriormente, se implementarán diferentes algoritmos de *clustering* no supervisado sobre los datos preprocesados, con el fin de evaluar su capacidad para generar agrupaciones consistentes y potencialmente interpretables.
- La calidad y estabilidad de las agrupaciones se valorará mediante el uso de métricas internas, como el coeficiente de silueta (*Silhouette*), el índice de Davies-Bouldin o el de Calinski-Harabasz, así como métricas externas como el *Adjusted Random Index* (ARI) o el *Normalized Mutual Information* (NMI).
- Finalmente, se interpretarán los grupos obtenidos en función de las áreas del desarrollo evaluadas, con el objetivo de identificar perfiles que puedan representar desviaciones respecto a las trayectorias esperadas para la edad, y analizar su posible utilidad como apoyo en el proceso de detección temprana.

METODOLOGÍA

Este capítulo describe la metodología empleada en el estudio, incluyendo el proceso de recogida de datos mediante la aplicación DAUCO, el preprocesamiento de la información y las técnicas de análisis utilizadas. Se justifica la elección de herramientas estadísticas y computacionales, y se presenta el diseño experimental adoptado para identificar patrones relevantes en el desarrollo infantil. Todo ello con el objetivo de garantizar la validez y replicabilidad del estudio.

2.1 RECOGIDA DE DATOS MEDIANTE LA APLICACIÓN DAUCO

El cuestionario DAUCO Test (Luque de Dios et al., 2025) fue desarrollado en el marco del proyecto de investigación financiado por la Consejería de Salud de la Junta de Andalucía, con el objetivo de facilitar la detección precoz de trastornos del desarrollo en niños de 0 a 6 años. La aplicación ha sido elaborada por un equipo multidisciplinar de profesionales de la Universidad de Córdoba y del Centro de Atención Temprana Infantil de la Universidad de Córdoba (CAIT-UCO), en colaboración con el Instituto Maimónides de Investigación Biomédica de Córdoba (IMIBIC) y la Fundación Cajasur.

La recogida de datos se realiza a través de una aplicación multiplataforma diseñada para su uso por parte de profesionales clínicos y educativos. En ella, los evaluadores dan de alta a cada menor introduciendo datos sociodemográficos básicos, antecedentes médicos y datos del contexto familiar. Posteriormente, se selecciona el tramo de edad correspondiente y se procede a la administración de las escalas de desarrollo.

La evaluación se basa en la observación directa del profesional o, en determinados contextos, en la entrevista estructurada con la familia del menor. Esta flexibilidad permite adaptar la recogida de información a distintas realidades clínicas o educativas.

Cada ítem del test se responde con una de tres opciones: *Sí*, *No* o *No se puede evaluar*. Además, existe la posibilidad de añadir comentarios cualitativos en cada respuesta. Las áreas evaluadas incluyen lenguaje, cognitiva, adaptativa, social, motricidad fina y motricidad gruesa, seleccionadas en función de la edad cronológica o ajustadas manualmente a criterio del profesional.

Al finalizar la evaluación, la aplicación genera un informe digital para el evaluador. Los informes contienen los resultados obtenidos en cada escala y están diseñados para

facilitar la interpretación clínica o su uso con fines de investigación.

La Tabla 2.1 muestra un resumen del número de ítems únicos incluidos en el test DAUCO según el área del desarrollo evaluada.

Tabla 2.1: Número de ítems únicos por área del desarrollo en el test DAUCO

Área del desarrollo	Número de ítems únicos
Cognitiva	570
Lenguaje	444
Social	382
Adaptativa	320
Motricidad gruesa	306
Motricidad fina	264

Todos los datos exportados se encuentran debidamente anonimizados, eliminando cualquier información identificativa. Solo se conservan las respuestas a los ítems, el tramo de edad correspondiente y variables generales como sexo y país de procedencia.

En este trabajo, se han utilizado exclusivamente los datos anonimizados exportados desde la aplicación de DAUCO proporcionados por María Araceli Sánchez Raya, directora del CAIT-UCO y experta en TND. No se participó en la fase de recogida directa de los datos, por lo que no se dispone de información adicional más allá de la incluida en los archivos suministrados.

Todos los participantes proporcionaron su consentimiento informado antes de participar en el estudio, asegurando que estaban completamente al tanto de los objetivos de la investigación y del tratamiento de sus datos. El estudio cumplió con todas las directrices éticas, incluida la Ley Orgánica 3/2018 de Protección de Datos Personales y los principios éticos establecidos en la Declaración de Helsinki. La aprobación ética del estudio fue concedida por el comité de ética institucional (número de aprobación 284, ref. 4153). Además, para garantizar la privacidad de los datos, toda la información obtenida fue tratada con estricta confidencialidad. El cuestionario fue diseñado para ser anónimo e incluía una sección introductoria explicando los objetivos del estudio, seguida de un formulario de consentimiento informado. Este procedimiento se ha llevado a cabo conforme a lo dispuesto en la legislación española sobre protección de datos personales [Government of Spain \(2018\)](#) y en los principios éticos internacionales para la investigación médica con seres humanos [World Medical Association \(2013\)](#).

Para facilitar la comprensión del procedimiento seguido en este trabajo, la Figu-

ra 2.1 muestra un diagrama de flujo con las principales etapas del análisis. El proceso comienza con la recogida de datos mediante la aplicación DAUCO, seguida del preprocesado y la reducción de dimensionalidad mediante PCA. A continuación, se aplican algoritmos de agrupamiento no supervisado y se evalúan los resultados con diversas métricas. Finalmente, se interpretan los grupos obtenidos en función de las áreas del desarrollo infantil.

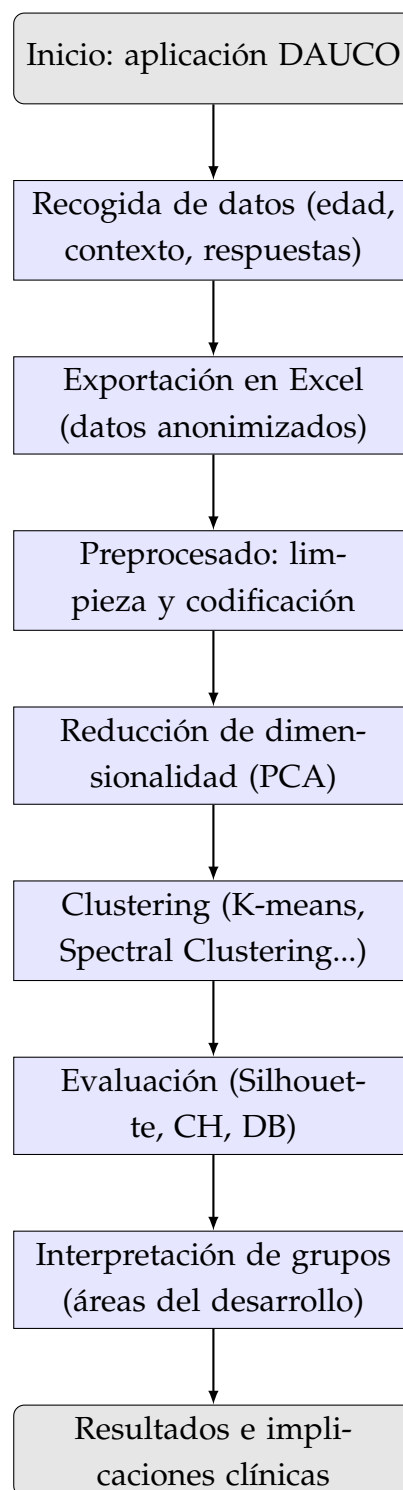


Figura 2.1: Diagrama de flujo del proceso de análisis aplicado a los datos del test DAU-CO

2.2 DESCRIPCIÓN DEL CONJUNTO DE DATOS

El conjunto de datos empleado en este estudio proviene de las respuestas recogidas mediante el cuestionario DAUCO Test, desarrollado por (Luque de Dios et al., 2025). Se trata de una escala diseñada para evaluar el desarrollo infantil en niños normotípicos de entre 1 y 84 meses. Este instrumento incluye un total de 1783 ítems codificados de forma binaria (0 = no adquirido; 1 = adquirido), organizados por áreas del desarrollo: lenguaje (L), cognitiva (C), adaptativa (A), motricidad fina (MF), motricidad gruesa (MG) y social (S), así como preguntas que integran múltiples dimensiones de manera simultánea.

Cada fila del conjunto representa la evaluación de un menor, mientras que las columnas corresponden a los distintos ítems del test. Además de los valores 0 «No» y 1 «Sí», algunos ítems incluyen el valor 99, que indica una respuesta del tipo «no se puede evaluar». También existen celdas vacías, ya que no todas las preguntas son aplicables a todos los rangos de edad.

Con el objetivo de adaptar los análisis a la estructura evolutiva del desarrollo, se dividió la muestra en 22 grupos codificados del 0 al 21, siguiendo los tramos definidos por la propia escala (desde el primer mes hasta los 72–84 meses). Esta segmentación, basada en hitos del desarrollo, permitió un análisis más ajustado a las trayectorias típicas: durante el primer año de vida, los niños fueron agrupados mes a mes; entre los 12 y 24 meses, en intervalos trimestrales; entre los 24 y 36 meses, en semestrales; y, a partir de los 36 meses, en tramos anuales. La codificación utilizada para simplificar estos grupos se presenta en la Tabla 2.2, y la distribución de sujetos por grupo puede consultarse en la Figura 2.2.

Este trabajo se centra en la franja de edad de 60–71 meses, al tratarse del grupo más representativo de la muestra, con un 17% del total de individuos.

2.3 PREPROCESAMIENTO

Durante el preprocesamiento se llevaron a cabo varias etapas de limpieza y filtrado para garantizar la calidad de los datos empleados en los análisis.

En primer lugar, se eliminaron los registros con un porcentaje elevado de datos ausentes. En los casos con ausencias puntuales, se aplicó imputación por moda dentro del mismo grupo de edad, con el objetivo de preservar los patrones característicos de

Tabla 2.2: Codificación de los grupos de edad utilizados en el análisis.

Rango de edad	Código
Primer mes	0
Segundo mes	1
Tercer mes	2
Cuarto mes	3
Quinto mes	4
Sexto mes	5
Séptimo mes	6
Octavo mes	7
Noveno mes	8
Décimo mes	9
Undécimo mes	10
Doce meses	11
12–15 meses	12
15–18 meses	13
18–21 meses	14
21–24 meses	15
24–30 meses	16
30–36 meses	17
36–48 meses	18
48–59 meses	19
60–71 meses	20
72–84 meses	21

cada etapa evolutiva.

Una fuente particular de valores ausentes fueron las respuestas codificadas como 99, que en el cuestionario DAUCO indican una opción del tipo «no se puede evaluar». En la Figura 2.3 se muestra la frecuencia de estos valores en cada ítem dentro del grupo de edad de 60 a 71 meses. Algunos ítems presentan tasas elevadas de respuesta tipo 99, lo que compromete su utilidad para el análisis.

Para asegurar que los modelos trabajaran con información suficiente y fiable, se aplicaron dos criterios adicionales de filtrado de variables:

- **Eliminación de columnas con menos del 80 % de respuestas válidas**, excluyendo tanto celdas vacías como respuestas tipo 99. Este umbral garantiza una repre-

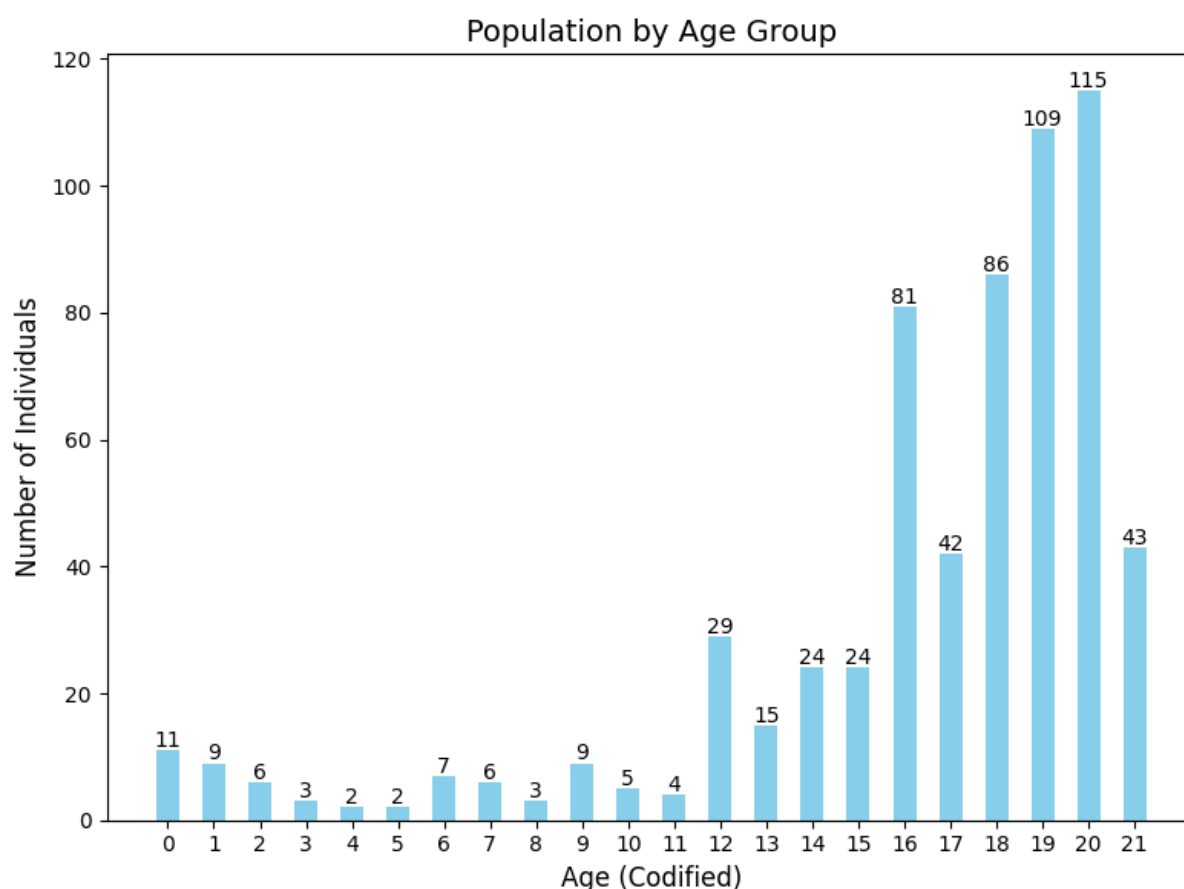


Figura 2.2: Distribución de individuos por grupo de edad en el conjunto de datos.

sentatividad mínima en cada variable.

- **Eliminación de columnas con una única respuesta válida.** Las variables sin variabilidad (por ejemplo, aquellas en las que todos los niños responden «sí» o «no») no aportan valor al proceso de agrupamiento y fueron descartadas.

Estas transformaciones redujeron de forma controlada la dimensionalidad del conjunto de datos, conservando únicamente aquellas variables con un mínimo de representatividad y capacidad discriminativa. En total, se eliminaron 119 ítems del cuestionario DAUCO correspondientes a las áreas funcionales (L, C, A, MF, MG, S), por no cumplir los criterios de completitud o variabilidad.

El conjunto final resultante incluyó 639 registros, listos para su posterior análisis mediante técnicas de reducción de dimensionalidad y agrupamiento no supervisado.

Con el objetivo de analizar la distribución y evolución de los ítems a lo largo del desarrollo, se ha contabilizado el número de preguntas disponibles por área funcional

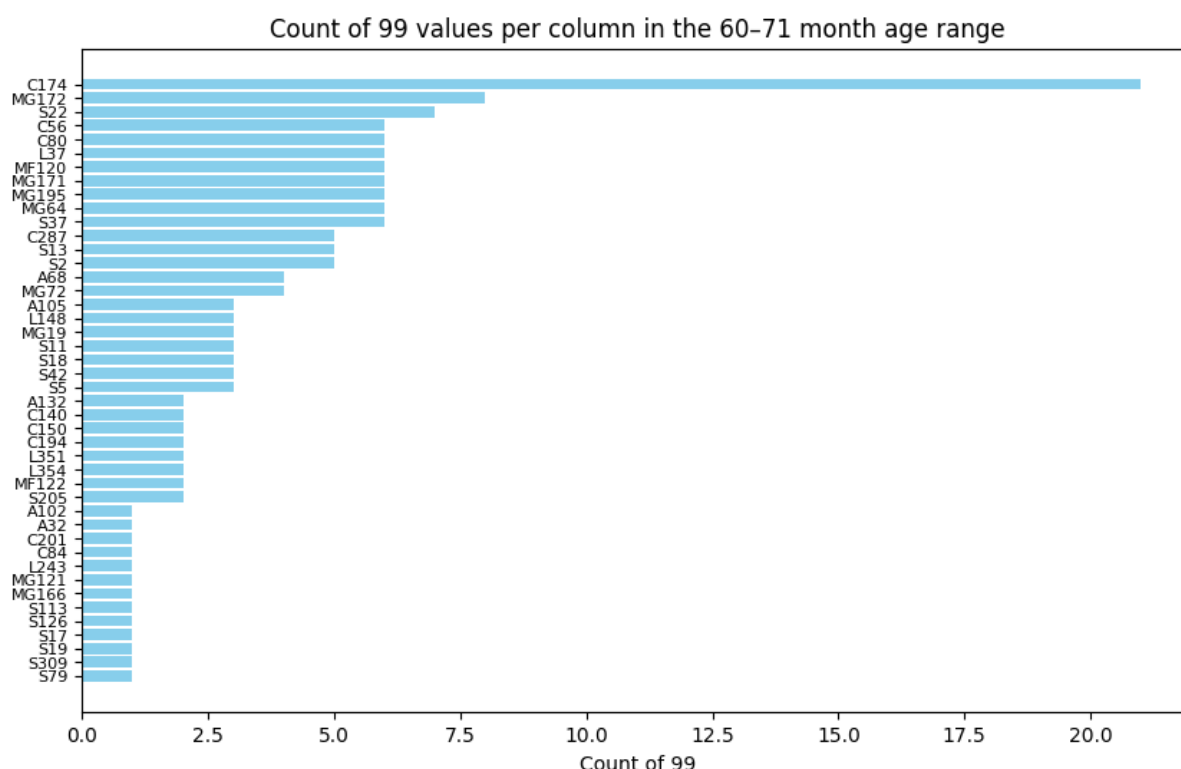


Figura 2.3: Cantidad de valores 99 («no se puede evaluar») por ítem del DAUCO Test en el grupo de edad 60–71 meses.

en cada uno de los grupos de edad definidos en el cuestionario DAUCO. La Tabla 2.3 resume estos resultados, considerando tanto los ítems específicos de un área como aquellos que integran varias dimensiones simultáneamente (por ejemplo, preguntas con códigos como SA o CMF, que se computan dentro de todas las áreas implicadas). Esta tabla permite visualizar cómo varía la carga de evaluación en función del tramo evolutivo, así como identificar posibles desequilibrios en la cobertura entre dominios del desarrollo.

Tabla 2.3: Número de ítems por área funcional en cada grupo de edad

Rango de edad	Lenguaje (L)	Cognitiva (C)	Adaptativa (A)	Mot. fina (MF)	Mot. gruesa (MG)	Social (S)
Primer mes	20	25	6	5	14	17
Segundo mes	37	33	41	22	19	49
Tercer mes	14	16	25	12	17	36
Cuarto mes	18	23	14	20	18	22
Quinto mes	12	38	19	14	17	21
Sexto mes	14	1	3	0	0	3
Séptimo mes	41	36	18	17	8	20
Octavo mes	38	51	11	16	15	21
Noveno mes	51	40	13	17	8	17
Décimo mes	23	45	15	25	15	26
Undécimo mes	13	34	12	9	7	14
Doce meses	24	32	15	16	10	14
12-15 meses	39	30	19	19	16	17
15-18 meses	16	16	12	13	17	13
18-21 meses	48	29	19	18	11	31
21-24 meses	32	63	32	46	21	47
24-30 meses	59	81	42	62	23	45
30-36 meses	95	124	60	70	36	63
36-48 meses	39	46	49	37	25	47
48-59 meses	20	37	25	17	12	41
60-71 meses	34	51	37	17	13	46
72-84 meses	14	15	22	0	0	20

2.4 EJEMPLOS DE ÍTEMS DEL GRUPO DE EDAD 60–71 MESES

Para ejemplificar con mayor detalle el tipo de ítems considerados en el análisis, a continuación se presentan tablas con preguntas seleccionadas en el rango de edad 60–71 meses, organizadas por área funcional. Estas tablas permiten visualizar el tipo de habilidades que se evalúan en cada dominio del desarrollo.

La Tabla 2.4 muestra una selección representativa de ítems del área de lenguaje. Se incluyen habilidades relacionadas con la comprensión verbal, la expresión oral y el uso adecuado del lenguaje en distintos contextos comunicativos.

Tabla 2.4: Ejemplos de ítems del área de Lenguaje (60–71 meses)

Código	Descripción
L121	Expresa sus sentimientos
L148	Identifica palabras que riman
L167	Interviene en una conversación
L241	Recuerda hechos de una historia contada
L243	Relaciona palabras con imágenes
L351	Utiliza el comparativo
L354	Utiliza el plural terminado en «es»
L37	Comprende el pasado de los verbos «ser» y «estar»

Por su parte, la Tabla 2.5 recoge ítems del área social, centrados en la interacción con otros niños y con adultos, el respeto a las normas y la adaptación al entorno escolar.

Tabla 2.5: Ejemplos de ítems del área Social (60-71 meses)

Código	Descripción
S11	Muestra compasión cuando otro niño/a se ha hecho daño
S113	Espera su turno para conseguir la atención del adulto/a
S124	Hace caso de las indicaciones que se le hacen
S126	Hace frente a las burlas y riñas
S13	Pide permiso antes de coger cosas de otro niño/a
S14	Permanece sentado/a y concentrado/a
S142	Inicia contactos con adultos/as conocidos/as
S17	Realiza juegos de simulación con roles
S18	Conoce gustos de al menos dos amigos/as
S19	Saluda a adultos/as conocidos/as sin indicación
S191	Muestra actitud positiva hacia la escuela
S194	Obedece normas y órdenes de la clase
S2	Asume el liderazgo ante problemas
S20	Se siente orgulloso/a de sus posesiones
S205	Prefiere jugar con compañeros/as
S21	Tiene un amigo/a con vínculo fuerte
S22	Observa reacciones del cuidador/a ante extraños
S309	Utiliza a adultos/as como recurso
S37	Asimila bondad moral por ejemplos
S42	Busca alternativas para resolver problemas
S5	Juega a juegos de tablero con reglas
S6	Nombra a su mejor amigo/a
S74	Resuelve conflictos
S76	Cuenta en casa lo hecho en la escuela
S79	Explica cosas sin demasiada vergüenza
S9	Hace alarde de sus logros
S91	Distingue conductas aceptables de no aceptables

La Tabla 2.6 incluye ítems del área cognitiva, con foco en la lógica, la atención, el razonamiento y el conocimiento general.

Tabla 2.6: Ejemplos de ítems del área Cognitiva (60–71 meses)

Código	Descripción
C140	Distingue su mano derecha de la izquierda
C150	Empareja palabras sencillas
C174	Forma un rectángulo con piezas triangulares
C194	Identifica partes incompletas de un dibujo
C195	Identifica colores
C201	Identifica primero y último de una fila
C287	Reconoce errores en dibujos absurdos
C56	Cita animales que producen leche
C80	Completa puzzle de una persona
C84	Comprende función de órganos externos
C92	Conoce los días de la semana

En la Tabla 2.7 se presentan ítems del área adaptativa, que evalúan la autonomía personal, la autorregulación y la capacidad para integrarse en contextos sociales y educativos.

Tabla 2.7: Ejemplos de ítems del área Adaptativa (60–71 meses)

Código	Descripción
A102	Escribe su nombre
A105	Intenta concluir la tarea
A132	Obedece normas de la clase
A140	Pide permiso
A142	Practica juegos de roles
A32	Asimila normas y actúa acorde
A325	Controla los esfínteres diurnos
A326	Controla los esfínteres nocturnos
A68	Responde a preguntas hipotéticas
A69	Continúa tareas con mínima supervisión
A99	Entiende reglas simples como «uno tú y otro yo»
CA186	Guarda los juguetes

La Tabla 2.8 resume algunos ítems del área de motricidad gruesa, vinculados al equilibrio, la coordinación y el control postural.

Tabla 2.8: Ejemplos de ítems del área de Motricidad Gruesa (60–71 meses)

Código	Descripción
MG171	Salta hacia delante al menos 60 cm
MG172	Salta con un solo pie y aterriza igual
MG19	Avanza dando brincos 3 metros
MG195	Se mantiene sobre un pie con ojos cerrados
MG64	Camina poniendo un pie delante del otro
MG72	Coge pelota grande 1 de cada 3 intentos

Finalmente, la Tabla 2.9 muestra ítems representativos del área de motricidad fina, que evalúan la precisión manual, la coordinación visual y motora y habilidades gráficas básicas.

Tabla 2.9: Ejemplos de ítems del área de Motricidad Fina (60–71 meses)

Código	Descripción
MF120	Hace un nudo
MF122	Hace una pelota arrugando papel
MF59	Copia los números del 1 al 5
MF60	Copia palabras sencillas
MF78	Dibuja una persona con seis elementos

2.5 TÉCNICAS DE ANÁLISIS

El análisis se dividió en tres fases principales: exploración de la estructura de los datos, reducción de dimensionalidad y agrupamiento no supervisado.

En primer lugar, se analizaron las correlaciones entre ítems para identificar relaciones entre distintas áreas del desarrollo. Para ello se utilizó la correlación de Pearson entre columnas binarizadas, con el objetivo de detectar redundancias o sinergias entre dimensiones, y obtener una visión preliminar de la organización interna del test.

Posteriormente, se aplicó PCA como técnica de reducción de dimensionalidad. Esta transformación permitió proyectar los datos originales en un espacio con menor número de variables, preservando la mayor parte de la varianza explicada. El número óptimo de componentes se seleccionó evaluando la curva de varianza acumulada,

estableciendo un umbral del 90% como criterio de corte.

Sobre la representación reducida de los datos se aplicaron distintos algoritmos de *clustering* no supervisado, con el objetivo de explorar la existencia de agrupaciones naturales entre los sujetos. Se emplearon métodos como K-Means, Gaussian Mixture Models (GMM) y Spectral Clustering.

Además, se examinó la estabilidad de los agrupamientos frente a variaciones en la inicialización o el número de clústeres, y se compararon los resultados obtenidos para diferentes valores de k .

2.6 IMPLEMENTACIÓN

El análisis fue implementado en Python 3.11.5, utilizando Jupyter Notebooks para facilitar un desarrollo limpio, explicativo y reproducible. Todo el código generado se encuentra disponible públicamente en un repositorio de Github¹. Se emplearon las siguientes librerías especializadas:

- Pandas y NumPy para la manipulación y limpieza de datos.
- Scikit-learn para los modelos de PCA, clustering y métricas de evaluación.
- Matplotlib y Seaborn para la visualización de datos y correlaciones.

Todos los experimentos se realizaron bajo condiciones controladas mediante fijación de semillas aleatorias, y se documentaron los hiperparámetros relevantes para cada técnica empleada.

2.7 MÉTRICAS DE EVALUACIÓN

Dado que el conjunto de datos no incluye etiquetas diagnósticas de referencia, se utilizaron métricas de validación interna ampliamente reconocidas en la literatura para evaluar la calidad de los agrupamientos generados. Estas métricas permiten analizar tanto la cohesión interna como la separación entre clústeres.

Coefficiente de Silueta (Rousseeuw, 1987): mide la adecuación de cada observación a su propio clúster frente a otros. Para un punto i , se define como:

¹<https://github.com/albertofermer/Trabajo-Fin-de-M-ster>

$$s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \quad (2.1)$$

donde $a(i)$ es la distancia media a los puntos del mismo clúster, y $b(i)$ es la distancia media mínima a los puntos de cualquier otro clúster. Valores próximos a 1 indican buena separación.

Índice de Davies-Bouldin (DB) (Davies and Bouldin, 1979): penaliza la superposición entre clústeres. Se define como:

$$DB = \frac{1}{k} \sum_{i=1}^k \max_{j \neq i} \left(\frac{s_i + s_j}{d(c_i, c_j)} \right) \quad (2.2)$$

donde s_i es la dispersión intra-clúster, y $d(c_i, c_j)$ es la distancia entre centroides. Valores más bajos indican mejor separación.

Índice de Calinski-Harabasz (CH) (Caliński and Harabasz, 1974): evalúa la relación entre la dispersión entre clústeres y la dispersión interna. Se calcula como:

$$CH = \frac{\text{Tr}(B_k)/(k-1)}{\text{Tr}(W_k)/(n-k)} \quad (2.3)$$

donde $\text{Tr}(B_k)$ y $\text{Tr}(W_k)$ son las trazas de las matrices de dispersión inter e intra-clúster, respectivamente.

Cuando fue posible evaluar la estabilidad de los agrupamientos o compararlos entre distintas ejecuciones del mismo algoritmo, se aplicaron también métricas externas:

Índice Rand Ajustado (ARI) (Steinley, 2004):

$$ARI = \frac{RI - \mathbb{E}[RI]}{\max(RI) - \mathbb{E}[RI]} \quad (2.4)$$

donde RI es el índice de Rand, y $\mathbb{E}[RI]$ su valor esperado. Valores cercanos a 1 indican una alta coincidencia entre particiones.

Información Mutua Normalizada (NMI) (Strehl and Ghosh, 2002):

$$NMI(U, V) = \frac{I(U, V)}{\sqrt{H(U) \cdot H(V)}} \quad (2.5)$$

donde $I(U, V)$ es la información mutua entre las particiones U y V , y $H(U)$, $H(V)$ sus respectivas entropías. Valores altos reflejan un mayor solapamiento informativo.

En este trabajo, estas métricas no se utilizaron para comparar los agrupamientos con etiquetas reales (ya que no se dispone de un diagnóstico clínico como *ground truth*), sino como indicadores de estabilidad ante diferentes inicializaciones aleatorias del mismo algoritmo. Para cada configuración, se ejecutó el algoritmo 10 veces con distintas semillas, y se calcularon los valores medios de ARI y NMI entre todos los pares de ejecuciones posibles. De este modo, fue posible estimar la sensibilidad del modelo a las condiciones iniciales y seleccionar configuraciones más robustas.

Esta metodología establece una base sólida para el análisis exploratorio posterior, permitiendo identificar patrones latentes en los datos del desarrollo infantil y evaluar la aplicabilidad de distintas técnicas de agrupamiento no supervisado.

RESULTADOS Y ANÁLISIS

En este capítulo se presentan los resultados obtenidos a partir del análisis de los datos, organizados en función de las distintas fases del estudio. Se explora la distribución de la muestra, la eficacia de las técnicas de agrupamiento y el impacto de la reducción de dimensionalidad. Además, se incluyen interpretaciones preliminares que permiten valorar la utilidad del enfoque propuesto en el análisis de patrones evolutivos del desarrollo infantil.

3.1 EXPLORACIÓN DE LOS DATOS

El conjunto de datos está formado por 639 niños españoles con edades comprendidas entre 1 y 84 meses, evaluados mediante el cuestionario DAUCO. Cada ítem de esta escala está codificado binariamente (0 = no adquirido, 1 = adquirido) y clasificado en una de las áreas del desarrollo infantil: lenguaje, cognitiva, motricidad fina, motricidad gruesa, adaptativa y social.

Como primer paso, se realizó un análisis exploratorio centrado en el grupo de edad de 60 a 71 meses, al ser el más representado en la muestra. Se examinaron los valores ausentes, las distribuciones de frecuencia y los estadísticos básicos por ítem, lo cual permitió obtener una visión preliminar de la consistencia y cobertura de las respuestas.

Adicionalmente, se llevó a cabo un análisis de correlación (Figura 3.1) utilizando el coeficiente de Pearson entre todos los ítems del test. La matriz de correlaciones obtenida reveló múltiples asociaciones significativas, con varios pares de ítems presentando correlaciones superiores a 0,70. Estos resultados sugieren la existencia de redundancias o solapamientos entre dimensiones del desarrollo, especialmente en etapas evolutivas avanzadas.

Una de las correlaciones más destacadas se observó entre el ítem A132 (área adaptativa, «Sigue normas de clase») y el ítem C201 (área cognitiva, «Identifica el primero y último objeto de una fila»), con un coeficiente de 0,70. (Diamond, 2013) mencionan que esta asociación podría reflejar la influencia de las funciones ejecutivas, como la memoria de trabajo o el control inhibitorio, en tareas que exigen tanto autorregulación como procesamiento cognitivo. Además, como señala el artículo de (Guralnick, 2001), los entornos estructurados pueden favorecer simultáneamente el desarrollo de habilidades cognitivas y adaptativas.

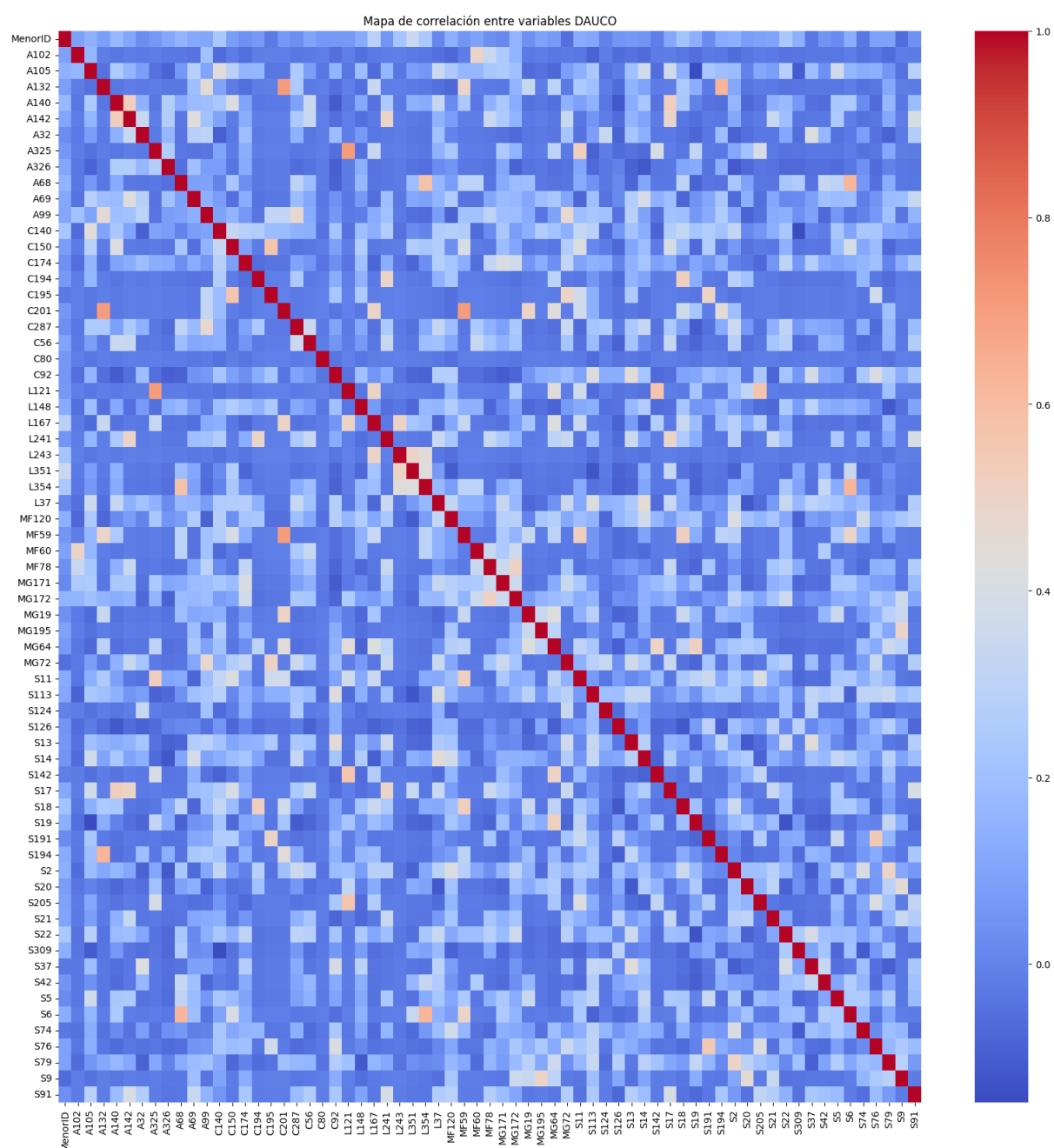


Figura 3.1: Matriz de correlación entre ítems del test DAUCO en el grupo de edad de 60-71 meses.

Otro caso interesante fue la correlación entre el ítem A325 («Controla esfínteres diurnos») del área adaptativa y L121 («Expresa emociones») del área de lenguaje, lo cual sugiere una interdependencia entre la autorregulación fisiológica, las emociones y la comunicación. Según (Wyndaele et al., 2020), la preparación para el control de esfínteres está estrechamente vinculada al desarrollo socioemocional, y particularmente

a la capacidad para expresar y comprender emociones, mediada por el lenguaje.

Asimismo, se identificó una fuerte relación entre C201 («Identifica el primero y último objeto de una fila») y MF59 (motricidad fina, «Copia los números del 1 al 5»), evidenciando la interacción entre dominios cognitivos y motores. Tal como apunta (Diamond, 2013), las funciones ejecutivas no solo son importantes para tareas cognitivas, sino también para habilidades motoras que implican planificación y precisión.

En conjunto, estos resultados refuerzan la idea de que el desarrollo infantil debe entenderse como un proceso integrado, donde los distintos dominios interactúan de forma continua. Esta perspectiva resulta especialmente útil en contextos de reducción de dimensionalidad o selección de características, donde considerar las interrelaciones permite preservar información relevante y evitar la pérdida de componentes funcionales clave.

3.2 RESULTADOS DEL AGRUPAMIENTO

A partir de los datos procesados y reducidos mediante PCA, según lo descrito en el Capítulo 2, se aplicaron distintos algoritmos de agrupamiento con el objetivo de identificar posibles estructuras en el desarrollo infantil. Entre los métodos evaluados se incluyen Spectral Clustering, Clustering Aglomerativo y GMM. La selección del número de clústeres k se guió principalmente por el coeficiente de silueta, descrito en la Sección 2.7.

En la Figura 3.2 se representa la evolución del coeficiente de silueta medio (con desviación estándar sombreada) en función del número de clústeres k , para distintas configuraciones del algoritmo Spectral Clustering. Se comparan afinidades del tipo `nearest_neighbors` y `rbf` con tres valores distintos del parámetro `gamma`.

Los resultados muestran que el uso de una afinidad `rbf` con `gamma = 0.1` mantiene valores de silueta más estables y elevados en un amplio rango de k , superando consistentemente el valor de 0.3 hasta aproximadamente $k = 40$. En cambio, valores más bajos de `gamma` o el uso de afinidad por vecinos (`nearest_neighbors`) generan puntuaciones inferiores y con mayor variabilidad. En todos los casos, se observa una caída pronunciada del coeficiente de silueta a partir de cierto umbral de complejidad (entre 35 y 40 clústeres), lo que indica una pérdida de cohesión interna en los agrupamientos cuando k es demasiado elevado.

La Figura 3.3 muestra la estabilidad del algoritmo Spectral Clustering evaluada

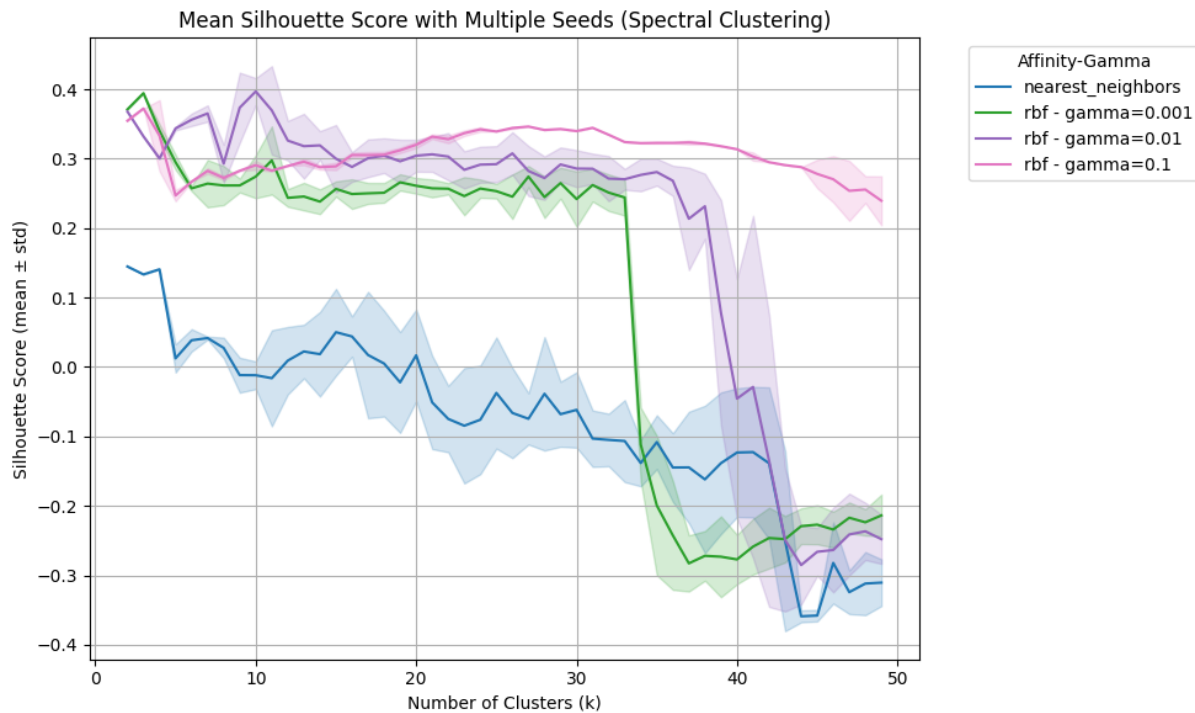


Figura 3.2: Estudio de la evolución del coeficiente de silueta en Spectral Clustering con distintas combinaciones de afinidad y parámetro gamma. La línea representa el promedio de 10 ejecuciones, y la sombra, la desviación estándar.

mediante ARI y NMI entre ejecuciones con diferentes semillas. Cada valor representa la similitud entre dos particiones generadas con la misma configuración pero distinta inicialización aleatoria.

Los resultados evidencian una variabilidad moderada en la asignación de clústeres. Aunque se alcanzan valores de coincidencia elevados en varias comparaciones (por encima de 0.85), también se observan pares con similitudes sustancialmente menores, especialmente en la matriz de ARI. La media de estabilidad obtenida fue de 0,768 para ARI y 0,763 para NMI, lo que indica que el algoritmo presenta cierta sensibilidad a la semilla, aunque no llega a comportarse de forma caótica. Esta fluctuación debe tenerse en cuenta al interpretar los resultados, especialmente en contextos donde la reproducibilidad es un criterio relevante.

El algoritmo de Clustering Aglomerativo fue el que obtuvo los resultados más consistentes. En la Figura 3.4 se presentan los valores del coeficiente de silueta para diversas combinaciones de métricas de distancia y métodos de enlace. En general, las configuraciones con distancia euclídea y enlace simple alcanzaron los valores más altos, superando 0,6 para $k = 2$. No obstante, se observa una disminución progresiva

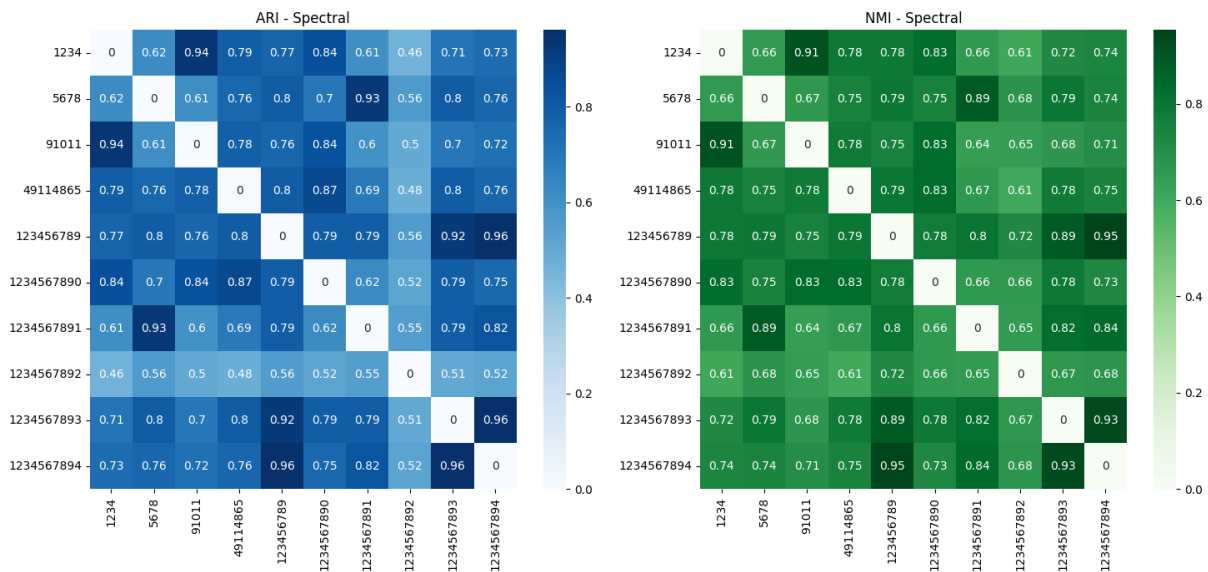


Figura 3.3: Estabilidad de Spectral Clustering con ARI (izquierda) y NMI (derecha) entre diferentes inicializaciones. Cada celda representa la similitud entre dos ejecuciones con la misma configuración de parámetros.

del coeficiente de silueta a medida que aumenta el número de clústeres, una tendencia común en este tipo de métodos jerárquicos.

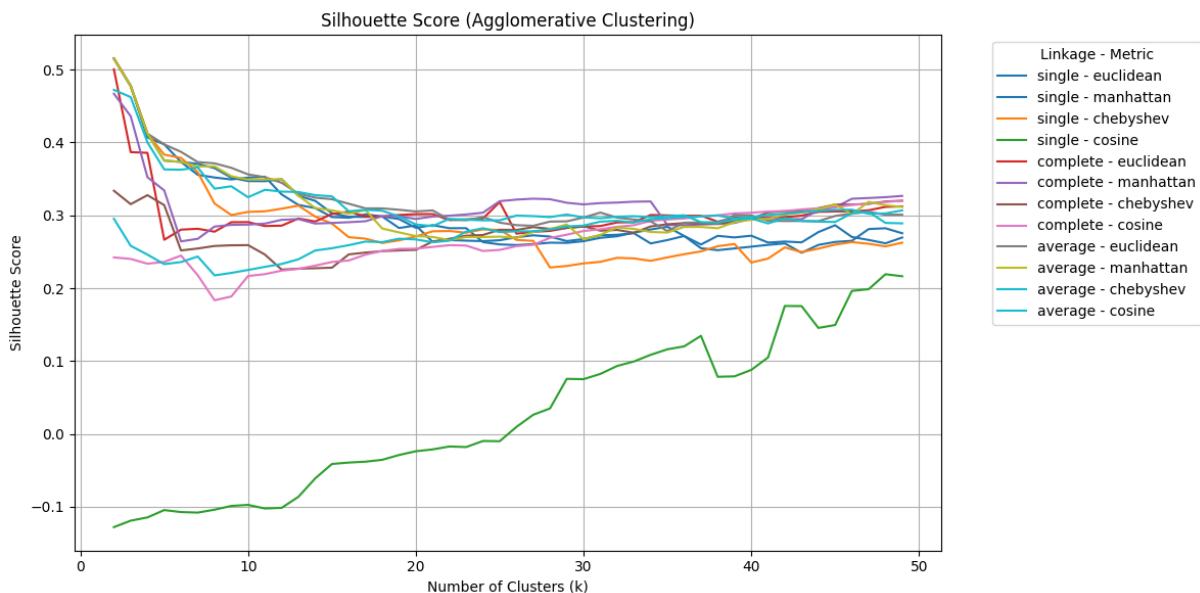


Figura 3.4: Coeficiente de silueta en Clustering Aglomerativo con diferentes combinaciones de distancia y método de enlace.

En cuanto a la estabilidad, los resultados fueron sobresalientes. La Figura 3.5 muestra las matrices de similitud entre ejecuciones con distintas semillas, evaluadas me-

diante ARI y NMI. En ambos casos se obtuvieron valores constantes e iguales a 1, lo que implica que el algoritmo produce exactamente los mismos agrupamientos en cada ejecución. Esto indica que el método es completamente determinista y robusto frente a variaciones en la inicialización, lo cual resulta especialmente valioso en contextos clínicos o educativos donde se requiere reproducibilidad y fiabilidad en la interpretación de los resultados. Al garantizar que un mismo perfil de respuestas genere siempre la misma clasificación, se refuerza la confianza en su uso como herramienta de apoyo a la toma de decisiones.

Esta total reproducibilidad no se debe únicamente a la calidad del modelo, sino a su naturaleza determinista: como se mencionó anteriormente, el algoritmo aglomerativo no utiliza inicializaciones aleatorias ni parámetros estocásticos, por lo que su salida es siempre idéntica dado un mismo conjunto de datos y configuración. Esta característica lo convierte en una opción especialmente atractiva cuando la estabilidad de los resultados es prioritaria.

Esta robustez también se refleja en las métricas promedio: 0,604 para el coeficiente de silueta, 0,277 para Davies-Bouldin y 8,325 para Calinski-Harabasz, con desviaciones estándar nulas. Estos valores posicionan al modelo aglomerativo como la alternativa más fiable dentro del conjunto de algoritmos evaluados.

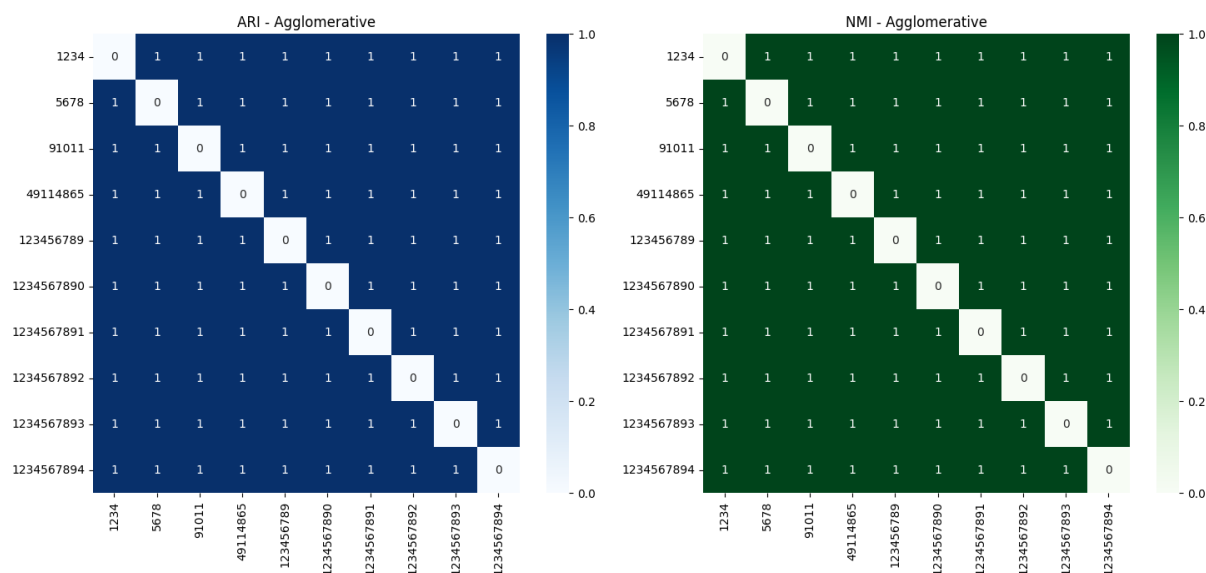


Figura 3.5: Estabilidad del Clustering Aglomerativo con ARI (izquierda) y NMI (derecha) ante variaciones en la semilla.

En comparación con los demás métodos, GMM ofreció los resultados más débiles tanto en calidad de agrupamiento como en estabilidad. En la Figura 3.6 se representa

el coeficiente de silueta promedio para distintas combinaciones de tipo de covarianza e inicialización. Aunque la variante `tied-kmeans` obtuvo los mejores resultados dentro de este enfoque, ninguna configuración alcanzó valores comparables a los observados en Spectral o Clustering Aglomerativo. La puntuación más alta apenas superó los 0,47, con una desviación estándar de $(\pm 0,059)$.

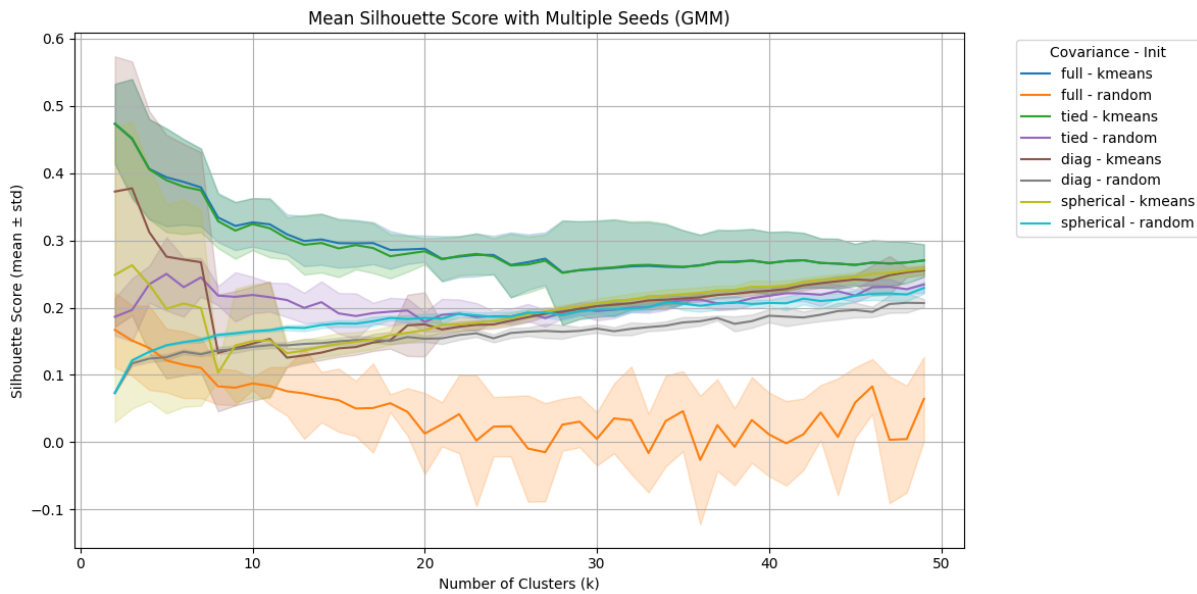


Figura 3.6: Estudio de la evolución del coeficiente de silueta en Gaussian Mixture Models con distintas combinaciones en sus parámetros. La línea representa el promedio de 10 ejecuciones, y la sombra, la desviación estándar.

La Figura 3.7 refleja una baja estabilidad del modelo entre ejecuciones, con valores de ARI y NMI considerablemente inferiores al resto de algoritmos. El promedio de ARI fue de apenas 0,090 y el de NMI, de 0,077, con celdas en las matrices de similitud que incluso toman valores negativos, lo que indica desacuerdo entre particiones por debajo del nivel esperado al azar. Estas cifras evidencian una gran sensibilidad a la semilla y una falta de consistencia en los resultados generados.

Esta inestabilidad puede atribuirse a la propia naturaleza del algoritmo: GMM asume que los datos provienen de una combinación de distribuciones normales multivariantes, una premisa que no se cumple en este caso, ya que el conjunto está compuesto por variables binarias y presenta estructuras que probablemente no siguen una distribución gaussiana. En consecuencia, GMM no se adapta bien a las características de los datos y su desempeño resulta poco fiable para el objetivo exploratorio de este trabajo.

La Tabla 3.1 resume los valores promedio obtenidos en cada métrica. En **negrita** se

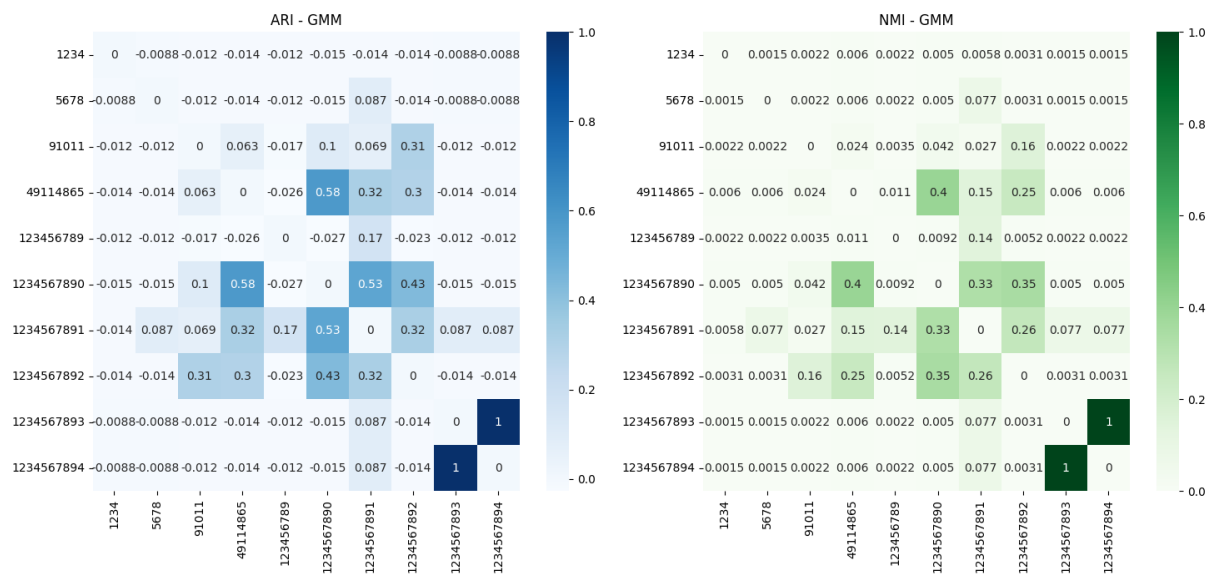


Figura 3.7: Estabilidad del modelo GMM con ARI (izquierda) y NMI (derecha) entre distintas ejecuciones. Los valores bajos y la dispersión reflejan una alta sensibilidad a la inicialización.

señalan los mejores resultados por columna: un mayor valor de Silhouette, Calinski-Harabasz o estabilidad, y un menor valor de Davies-Bouldin, indican mayor calidad del agrupamiento.

Tabla 3.1: Promedio de métricas de rendimiento y estabilidad para las configuraciones seleccionadas de agrupamiento. En negrita se indican los mejores resultados por métrica.

Método	Sil (\uparrow)	DB (\downarrow)	CH (\uparrow)	ARI (\uparrow)	NMI (\uparrow)
Spectral	0,397 \pm 0,019	1,328 \pm 0,095	7,489 \pm 0,164	0,768	0,763
Aglomerativo	0,604 \pm 0,000	0,277 \pm 0,000	8,325 \pm 0,000	1,000	1,000
GMM	0,474 \pm 0,059	1,382 \pm 0,984	6,650 \pm 1,958	0,090	0,077

Estos resultados respaldan la elección del algoritmo aglomerativo como la opción más robusta y fiable para el análisis no supervisado en este estudio, tanto por la calidad de los clústeres generados como por su estabilidad entre ejecuciones.

No obstante, dado el carácter exploratorio del análisis y la homogeneidad de la muestra (niños con desarrollo normotípico), no es posible atribuir un significado clínico directo a los agrupamientos obtenidos. Algunas configuraciones podrían reflejar perfiles más avanzados o rezagados en ciertas áreas funcionales, pero estas interpre-

taciones deben considerarse preliminares y sujetas a validación futura con muestras clínicas etiquetadas.

En conjunto, los hallazgos sugieren la presencia de cierta estructura interna en los datos, aunque no claramente separable. La calidad y estabilidad de los agrupamientos varía entre algoritmos, siendo Clustering Aglomerativo el más consistente. La elevada dimensionalidad inicial y el posible solapamiento entre áreas del desarrollo podrían estar dificultando una segmentación más definida, lo que justifica la exploración de técnicas adicionales en trabajos futuros.

3.3 IMPACTO DEL PREPROCESAMIENTO CON PCA

Para evaluar el impacto del preprocesamiento mediante reducción de dimensionalidad, se comparó el rendimiento de los algoritmos de agrupamiento con y sin la aplicación de PCA. Se utilizaron tres métricas internas: el coeficiente de Silueta, el índice de Davies-Bouldin y el índice de Calinski-Harabasz, cuyos valores medios y desviaciones estándar se resumen en la Tabla 3.2.

Los resultados muestran que el uso de PCA puede mejorar la calidad del agrupamiento en algunos casos, especialmente en el algoritmo aglomerativo, donde el coeficiente de silueta aumentó de 0,577 a 0,604 y se registró una mejora tanto en Davies-Bouldin como en Calinski-Harabasz. En Spectral Clustering también se observa una mejora leve en todas las métricas tras aplicar PCA.

Sin embargo, en el caso de GMM los resultados fueron mixtos: aunque el coeficiente de silueta y el índice de Calinski-Harabasz mejoraron tras la reducción de dimensionalidad, el índice de Davies-Bouldin fue ligeramente superior sin PCA. Este comportamiento sugiere que el impacto del preprocesamiento no es uniforme y depende de la sensibilidad del modelo a la estructura del espacio de representación.

En conjunto, estos resultados sugieren que la aplicación de PCA puede contribuir a mejorar la calidad del agrupamiento, especialmente cuando la estructura de los datos originales presenta redundancias o ruido. No obstante, su impacto no es uniforme y conviene evaluar su efecto empíricamente en función del algoritmo y del dominio de aplicación.

Tabla 3.2: Comparación de resultados de agrupamiento con y sin PCA. Se muestran la media y desviación estándar. En negrita, los mejores valores por algoritmo.

Método	Condición	Silhouette (\uparrow)	Davies-Bouldin (\downarrow)	Calinski-Harabasz (\uparrow)
GMM	Con PCA	0,474 \pm 0,059	1,382 \pm 0,984	6,650 \pm 1,958
GMM	Sin PCA	0,461 \pm 0,048	1,271 \pm 0,772	5,788 \pm 1,377
Aglomerativo	Con PCA	0,604 \pm 0,000	0,277 \pm 0,000	8,325 \pm 0,000
Aglomerativo	Sin PCA	0,577 \pm 0,000	0,298 \pm 0,000	7,558 \pm 0,000
Spectral	Con PCA	0,397 \pm 0,019	1,328 \pm 0,095	7,489 \pm 0,164
Spectral	Sin PCA	0,385 \pm 0,011	1,385 \pm 0,105	6,739 \pm 0,184

3.4 INTERPRETACIÓN EXPLORATORIA DE LOS RESULTADOS

Los resultados obtenidos mediante técnicas de agrupamiento no permitieron identificar clústeres claramente diferenciados en la muestra analizada. En la mayoría de configuraciones, el coeficiente de silueta se mantuvo en valores moderadamente bajos (alrededor de 0,4), lo que refleja una estructura interna débil. Este comportamiento es coherente con la composición de la muestra, formada exclusivamente por niños con desarrollo normotípico, sin presencia de casos clínicos contrastados que pudieran generar patrones marcadamente divergentes.

Pese a ello, el análisis exploratorio de las correlaciones entre ítems permitió detectar asociaciones funcionales relevantes, especialmente en el grupo de edad de 60 a 71 meses. Entre las más destacadas se encontraron:

- Correlaciones entre ítems adaptativos y cognitivos, como seguir normas de clase y realizar tareas de secuenciación visual.
- Conexiones entre aspectos adaptativos y lingüísticos, como el control de esfínteres y la expresión emocional.
- Relaciones entre habilidades cognitivas y de motricidad fina, como la identificación secuencial y la copia de números.

Estos hallazgos sugieren la existencia de trayectorias interdependientes entre distintas áreas del desarrollo, en línea con modelos que defienden una maduración inte-

grada durante la infancia. Aunque los agrupamientos generados no tienen valor diagnóstico ni permiten segmentaciones clínicas en esta muestra homogénea, los patrones de correlación observados ofrecen hipótesis valiosas para futuras investigaciones.

En definitiva, este análisis exploratorio no evidencia la presencia de agrupamientos consistentes, pero sí apunta a posibles vínculos funcionales entre dimensiones del desarrollo. Para validar estas observaciones y explorar su utilidad clínica, será necesario aplicar estos métodos sobre muestras más heterogéneas y clínicamente caracterizadas.

CONCLUSIONES Y TRABAJO FUTURO

Este capítulo resume las principales conclusiones del trabajo, poniendo especial interés en las aportaciones del estudio al campo de la detección temprana de TNDs. Además, se reflexiona sobre las limitaciones encontradas y se proponen líneas de investigación futuras que podrían ampliar y mejorar los hallazgos obtenidos.

4.1 COMPARACIÓN ENTRE ALGORITMOS

Los resultados obtenidos en el análisis de agrupamiento permiten extraer algunas observaciones relevantes sobre el comportamiento de los distintos algoritmos aplicados. En primer lugar, el algoritmo de GMM mostró una menor estabilidad entre ejecuciones, tal y como se refleja en las métricas ARI y NMI. Esta inestabilidad puede deberse a dos factores principales: por un lado, su sensibilidad a la inicialización aleatoria, y por otro, la posible falta de ajuste del modelo a la estructura real de los datos. GMM parte de la suposición de que los datos provienen de distribuciones gaussianas bien separadas, lo cual podría no cumplirse en el conjunto de datos, caracterizado por solapamientos entre dimensiones del desarrollo.

En cambio, el algoritmo aglomerativo mostró una mayor robustez y estabilidad, con puntuaciones altas y consistentes en las métricas de evaluación. Esta fiabilidad se debe a que no depende de ninguna inicialización aleatoria, ya que su estructura jerárquica construye los clústeres de forma determinista a partir de una matriz de distancias. Por su parte, Spectral Clustering presentó resultados intermedios: aunque logró puntuaciones aceptables en algunas configuraciones, su sensibilidad a los parámetros de afinidad lo hizo menos predecible. El comportamiento de K-Means fue correcto en configuraciones simples, pero mostró menor capacidad para modelar estructuras complejas.

Cabe destacar que la existencia de fuertes correlaciones entre áreas funcionales podría estar influenciando la dificultad de los modelos para generar agrupamientos bien definidos, al difuminar los límites entre dominios teóricamente diferenciados.

A diferencia de los enfoques supervisados, el uso de técnicas de agrupamiento sin etiquetas diagnósticas permite explorar la estructura latente del desarrollo infantil sin imponer categorías clínicas predefinidas. Esta estrategia resulta especialmente útil en etapas tempranas, donde los síntomas pueden no estar claramente definidos o no en-

cajar en ningún diagnóstico formal. Al identificar agrupaciones naturales basadas únicamente en patrones conductuales y evolutivos, se abren posibilidades para detectar trayectorias atípicas que podrían ser indicativas de riesgo, sin depender de etiquetas rígidas o retrospectivas.

4.2 LIMITACIONES DEL ENFOQUE

A pesar de los avances logrados, este estudio presenta varias limitaciones. En primer lugar, el conjunto de datos utilizado, aunque extenso, proviene de una población normotípica. Esto implica que los patrones encontrados reflejan principalmente trayectorias de desarrollo típico, lo que limita la capacidad de extrapolación hacia perfiles atípicos o con sospecha de TND.

En segundo lugar, el número de variables sigue siendo elevado incluso tras la aplicación de PCA. Aunque esta técnica ha permitido reducir dimensionalidad y eliminar redundancia, no garantiza que todas las dimensiones relevantes para la detección de TND estén representadas en los componentes seleccionados. Además, el carácter binario de los ítems y la existencia de ítems aplicables solo a ciertos grupos de edad dificultan una modelización homogénea del conjunto.

Otra limitación destacable es la ausencia de etiquetas diagnósticas de referencia, lo que impide validar directamente la utilidad clínica de los agrupamientos generados. Esto condiciona la interpretación de los resultados, que deben considerarse de naturaleza exploratoria.

4.3 IMPLICACIONES PRÁCTICAS Y APLICACIONES FUTURAS

Los resultados obtenidos en este estudio abren la puerta a aplicaciones prácticas en contextos clínicos y educativos. La metodología de análisis propuesta podría utilizarse como una herramienta de cribado preliminar para detectar patrones de desarrollo atípico sin necesidad de contar con un diagnóstico formal en etapas tempranas. Esto permitiría priorizar la evaluación de casos con mayor riesgo y orientar de forma más eficiente los recursos profesionales hacia perfiles que requieran una atención específica.

Para que este enfoque pueda consolidarse como una herramienta de apoyo clínico, serían necesarios varios pasos adicionales. En primer lugar, sería imprescindible validar los agrupamientos obtenidos utilizando muestras clínicas que incluyan niños

con diagnósticos confirmados de TND. Además, sería conveniente evaluar la replicabilidad de los perfiles detectados en otras poblaciones con diferentes características socioculturales y utilizando escalas de evaluación distintas. Por último, el desarrollo de una interfaz automatizada, accesible para profesionales no especializados en análisis de datos, permitiría aplicar estas técnicas de forma sistemática y a gran escala en servicios de atención temprana o en contextos escolares.

El análisis de correlaciones realizado también ha revelado asociaciones significativas entre áreas funcionales del desarrollo, particularmente en el grupo de edad de 60-71 meses. Por ejemplo, la correlación entre el control de esfínteres (A325) y la expresión de emociones (L121) sugiere una interacción entre la autorregulación fisiológica, la competencia emocional y el lenguaje. De igual modo, la relación entre ítems como «Sigue normas de clase» (A132), «Identifica el primero y último objeto de una fila» (C201) y «Copia los números del 1 al 5» (MF59) respalda la existencia de sinergias entre funciones ejecutivas, habilidades cognitivas y motricidad fina. Estas interrelaciones podrían aprovecharse para diseñar versiones abreviadas del cuestionario, eliminando ítems redundantes sin comprometer la validez del instrumento.

Desde el punto de vista técnico, la aplicabilidad del enfoque a entornos reales también dependerá de su escalabilidad. Aunque el tiempo de cómputo fue razonable para esta muestra, en bases de datos más extensas sería necesario optimizar el preprocesamiento y aplicar estrategias que reduzcan la complejidad computacional, particularmente en tareas como la imputación de datos o la evaluación de múltiples configuraciones de agrupamiento.

Como líneas futuras, sería valioso ampliar el tamaño del conjunto de datos, especialmente en aquellos grupos de edad menos representados, para mejorar la robustez de los modelos. Asimismo, se podrían explorar técnicas de reducción de dimensionalidad no lineales, como t-SNE o autoencoders, que podrían capturar estructuras más complejas que las que permite PCA. Finalmente, el seguimiento longitudinal de los sujetos evaluados podría permitir relacionar los patrones de agrupamiento con trayectorias evolutivas y diagnósticos clínicos posteriores, lo que proporcionaría una validación externa de gran relevancia para el enfoque propuesto.

En resumen, este trabajo ha mostrado el potencial de las técnicas no supervisadas de inteligencia artificial para analizar patrones del desarrollo infantil a partir de datos del test DAUCO. Aunque con limitaciones, los resultados apuntan a la viabilidad de este enfoque como herramienta de cribado preliminar, abriendo nuevas vías para la identificación temprana de trayectorias atípicas. La combinación de clustering, re-

ducción de dimensionalidad e interpretabilidad representa una línea de investigación prometedora en el ámbito de la atención temprana.

CONTRIBUCIONES

Este capítulo recoge la contribución principal del trabajo en forma de un artículo enviado al congreso científico de MetroXRINE, lo cual representa un primer paso en la transferencia de los resultados a la comunidad investigadora. Se destaca el potencial de este enfoque basado en aprendizaje no supervisado como apoyo a la evaluación clínica en Atención Temprana.

Las contribución que ha derivado de este Trabajo de Fin de Máster es la siguiente:

- A. Fernández-Merchán, J. M. Alcalde-Llargo, A. Zingoni, A. Sánchez Raya, C. Pérez Dueñas, y E. Yeguas Bolívar, “Early Detection of Neurodevelopmental Disorders in Early Childhood through Machine Learning,” 2025.

Enviado al Congreso MetroXRINE 2025: *IEEE International Conference on Metrology for Extended Reality, Artificial Intelligence and Neural Engineering*: <https://www.metroxraine.org>. El artículo se encuentra actualmente en el último proceso de revisión tras ser aceptado con *minor revisions*. Véase el apéndice Artículo MetroXRINE.

BIBLIOGRAFÍA

- Al-Saoud, S., Nichols, E. S., Brossard-Racine, M., Wild, C. J., Norton, L., and Duerden, E. G. (2024). A transdiagnostic examination of cognitive heterogeneity in children and adolescents with neurodevelopmental disorders. *Child Neuropsychology: A Journal on Normal and Abnormal Development in Childhood and Adolescence*, 31(2):293–311. (página 3).
- American Psychiatric Association (2013). *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5)*. American Psychiatric Publishing, Washington, DC. (página 1).
- Bastiaansen, J. A. J., Veldhuizen, E. E., De Schepper, K., and Scheepers, F. E. (2022). Experiences of siblings of children with neurodevelopmental disorders: Comparing qualitative analysis and machine learning to study narratives. *Frontiers in Psychiatry*, 13:719598. (página 3).
- Caliński, T. and Harabasz, J. (1974). A dendrite method for cluster analysis. *Communications in Statistics*, 3(1):1–27. (página 21).
- Chavas, J., Guillon, L., Pascucci, M., Dufumier, B., Rivière, D., and Mangin, J.-F. (2022). Unsupervised representation learning of cingulate cortical folding patterns. In Wang, L., Dou, Q., Fletcher, P. T., Speidel, S., and Li, S., editors, *Medical Image Computing and Computer Assisted Intervention – MICCAI 2022*, pages 77–87, Cham. Springer Nature Switzerland. (página 3).
- Chien, Y.-L., Lin, H.-Y., Tung, Y.-H., Hwang, T.-J., Chen, C.-L., Wu, C.-S., Shang, C.-Y., Hwu, H.-G., Tseng, W.-Y. I., Liu, C.-F., and Gau, S. S.-F. (2022). Neurodevelopmental model of schizophrenia revisited: similarity in individual deviation and idiosyncrasy from the normative model of whole-brain white matter tracts and shared brain-cognition covariation with adhd and asd. *Molecular Psychiatry*, 27(8):3262–3271. (página 3).
- Cuppens, T., Kaur, M., Kumar, A. A., Shatto, J., Ng, A. C.-H., Leclercq, M., Reformat, M. Z., Droit, A., Dunham, I., and Bolduc, F. V. (2023). Developing a cluster-based approach for deciphering complexity in individuals with neurodevelopmental differences. *Frontiers in Pediatrics*, 11:1171920. (páginas 2, 3 y 4).
- 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. (página 21).

- Diamond, A. (2013). Executive functions. *Annual Review of Psychology*, 64:135–168. (páginas 23 y 25).
- Government of Spain (2018). Organic law on the protection of personal data and guarantee of digital rights. <https://www.boe.es/eli/es/lo/2018/12/05/3/con>. Organic Law 3/2018, of December 5. (página 8).
- Guralnick, M. J. (2001). A developmental systems model for early intervention. *Infants & Young Children*, 14(2):1–18. (página 23).
- Luque de Dios, S. M., Moriana, J. A., Pérez-Dueñas, C., Martínez-Gual, E., Camacho Alcaide, A., González Matilla, R., Antolí, A., and Sánchez-Raya, A. (2025). Propiedades psicométricas y evaluación del desarrollo infantil con la escala dauco test. *Revista de Psicología Clínica con Niños y Adolescentes*. Advance online publication. (páginas 2, 3, 4, 7 y 11).
- Luque de Dios, S. M., Sánchez-Raya, A., and Moriana, J. A. (2023). Quality of child development scales. a systematic review. *International Journal of Educational Psychology*, 12(2):119–148. (página 2).
- Mayer, S., Fanton, S., Riquin, E., Ballon, N., Cachia, A., Henry, C., Plaze, M., Krebs, M.-O., Houenou, J., Dubertret, C., et al. (2022). Unsupervised learning for neurodevelopmental profiles in a transdiagnostic pediatric sample: Combining umap and density-based clustering. *Frontiers in Psychiatry*, 13:960672. (página 2).
- Mengi, M. and Malhotra, D. (2024). Usmda: Unsupervised multisource domain adaptive adhd prediction model using neuroimaging. *Knowledge-Based Systems*, 305:112615. (página 2).
- Morris-Rosendahl, D. J. and Crocq, M. A. (2020). Neurodevelopmental disorders-the history and future of a diagnostic concept. *Dialogues in Clinical Neuroscience*, 22(1):65–72. (páginas 1 y 2).
- Muñiz, J. (2010). Test of theories of the classical theory and theory of responses to items. *Papeles del Psicólogo*, 31(1):57–66. (página 2).
- Rahman, M. M., Usman, O. L., Muniyandi, R. C., Sahran, S., Mohamed, S., and Razak, R. A. (2020). A review of machine learning methods of feature selection and classification for autism spectrum disorder. *Brain Sciences*, 10(12):949. (página 3).

- Reilly, S., Bishop, D. V., and Tomblin, B. (2014). Terminological debate over language impairment in children: forward movement and sticking points. *International Journal of Language and Communication Disorders*, 49(4):452–462. (páginas 1, 2 y 4).
- Rivard, M., Bertrand, M., Lippé, S., and Forget-Dubois, N. (2023). Unsupervised learning reveals distinct developmental profiles in children with autism spectrum disorder. *Autism Research*, 16(5):879–891. (páginas 2 y 3).
- 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. (página 20).
- Silva, M., Mendonça, F., Euclides, J., Mônego, B., and Bandeira, D. (2018). Instruments for multidimensional assessment of child development: a systematic review. *Early Child Development and Care*. (página 2).
- Song, J., Liu, Y., Wang, M., and Jiang, H. (2022). Machine learning for early identification of autism spectrum disorder: a review of data types and modeling approaches. *Frontiers in Psychiatry*, 13:865279. (página 2).
- Steinley, D. (2004). Properties of the hubert-arable adjusted rand index. *Psychological Methods*, 9(3):386–396. (página 21).
- Strehl, A. and Ghosh, J. (2002). Cluster ensembles - a knowledge reuse framework for combining multiple partitions. *Journal of Machine Learning Research*, 3:583–617. (página 21).
- Temprana, G. A. (2000). *Libro Blanco de la Atención Temprana*. Real Patronato sobre Discapacidad, Madrid. (página 1).
- Toki, E. I., Tsoulos, I. G., Santamato, V., and Pange, J. (2024). Machine learning for predicting neurodevelopmental disorders in children. *Applied Sciences*, 14(2):837. (página 2).
- Vincent, J., Fovet, T., Cleret de Langavant, L., et al. (2024). The promise of explainable artificial intelligence for neurodevelopmental disorders: improving interpretability and trust in diagnosis. *npj Science of Learning*, 9(1):17. (página 2).
- Visser, L., Ruiter, S., van der Meulen, B., Ruijsenaars, W., and Timmerman, M. (2012). A review of standardized developmental assessment instruments for young children and their applicability for children with special needs. *Journal of Cognitive Education and Psychology*, 11(2):102–127. (páginas 2 y 3).

- Wang, B., Liu, Z., Li, Y., Xiao, X., Zhang, R., Cao, Y., Cui, L., and Zhang, P. (2020). Un-supervised graph domain adaptation for neurodevelopmental disorders diagnosis. In Martel, A. L., Abolmaesumi, P., Stoyanov, D., Mateus, D., Zuluaga, M. A., Zhou, S. K., Racocanu, D., and Joskowicz, L., editors, *Medical Image Computing and Computer Assisted Intervention – MICCAI 2020*, pages 496–505, Cham. Springer International Publishing. (página 3).
- World Medical Association (2013). World medical association declaration of helsinki: Ethical principles for medical research involving human subjects. *JAMA*, 310:2191–2194. (página 8).
- Wyndaele, J.-J., Kaerts, N., Wyndaele, M., and Vermandel, A. (2020). Development signs in healthy toddlers in different stages of toilet training: Can they help define readiness and probability of success? *Global Pediatric Health*, 7. (página 24).

APÉNDICES

ARTÍCULO METROXRINE

Este apéndice incluye el artículo enviado al congreso Metrology for eXtended Reality, Artificial Intelligence and Neural Engineering (*MetroXRINE*). Para más información, véase: <https://metroxraine.org>. En el artículo se recogen los objetivos, la metodología y los principales resultados del presente Trabajo de Fin de Máster. El texto fue preparado con la intención de resumir de forma clara el trabajo realizado, centrado en la aplicación de técnicas de aprendizaje no supervisado para detectar posibles patrones de desarrollo atípico en la infancia.

El artículo describe el uso de técnicas de reducción de dimensionalidad y distintos algoritmos de agrupamiento aplicados a datos recogidos mediante un cuestionario clínico dirigido a niños de entre 1 y 84 meses. También se comentan algunas correlaciones relevantes entre ítems, así como la comparación de resultados obtenidos con varios métodos de agrupamiento. El artículo ha sido enviado al congreso para su revisión y posible presentación dentro del área de inteligencia artificial aplicada a la salud infantil.

A continuación, se incluye el texto completo tal como fue remitido tras su aceptación con *minor revisions*.

Early Detection of Neurodevelopmental Disorders in Early Childhood through Machine Learning

Alberto Fernández-Merchán

University of Cordoba
Artificial Vision Applications Group
Córdoba, Spain
z42ferma@uco.es

José M. Alcalde-Llargo

Computer Science Department
University of Córdoba
Córdoba, Spain
jmalcalde@uco.es

Andrea Zingoni

DEIM
University of Tuscia
Tuscia, Italy
andrea.zingoni@unitus.it

Araceli Sánchez Raya

Department of Psychology (IMIBIC. CAIT.)
University of Córdoba
Córdoba, Spain
edlsaram@uco.es

Carolina Perez Dueñas

Department of Psychology (IMIBIC. CAIT.)
University of Córdoba
Córdoba, Spain
cpduenas@uco.es

Enrique Yeguas Bolívar

Computer Science Department
University of Córdoba
Córdoba, Spain
eyeguas@uco.es

Abstract—Early detection of Neurodevelopmental Disorders (NDD) allows for interventions that can influence child development. This study applies machine learning techniques to analyze growth patterns and identify deviations. Using data from 638 neurotypically developing Spanish children (aged 1–84 months), we segment the population by age and apply clustering to explore developmental trends. Additionally, we evaluate machine learning models to refine feature selection and improve classification. The results contribute to NDD screening and detection, offering a data-driven approach to support decision-making in early childhood care.

Index Terms—Machine Learning, Clustering, Early Childhood Care, Neurodevelopmental Disorders

I. INTRODUCTION

Early childhood is a critical period for human development, and detecting neurodevelopmental disorders (NDDs) at an early stage is essential for timely interventions. These disorders, such as autism spectrum disorder (ASD) and developmental delays, often co-occur, leading to a shift from traditional diagnostic models to dimensional approaches [1]. However, accurate identification remains challenging due to variability in symptom presentation [2].

Developmental scales have long been used to assess children's progress, but they often face issues such as reliability concerns and arbitrary age-group definitions [3]. In this context, data-driven methods and machine learning offer new opportunities for more precise and individualized analysis of child development [4].

This study explores the application of unsupervised machine learning techniques, such as clustering and dimensionality reduction, to uncover developmental patterns in early childhood and assist in the early detection of NDD.

II. RELATED WORK

Neurodevelopmental disorders present considerable complexity due to factors such as comorbidities, varied clinical characteristics, and overlapping diagnoses, which complicate

early identification and intervention processes [5]. Several studies have explored the application of machine learning techniques, particularly unsupervised methods like clustering, to identify clinical profiles beyond standard diagnostic categories. In [6], cluster analysis is applied to behavioral and cognitive data from children referred for assessment, identifying three distinct profiles that revealed patterns not reflected in conventional diagnoses. These studies suggest that clustering can help detect subgroups of children with different developmental trajectories, potentially informing tailored diagnostic and intervention strategies.

In a related line of work, artificial intelligence has also been applied to data collected through serious games designed to evaluate children's speech and language abilities. One study used a dataset of 473 participants and combined cluster analysis with logistic regression to predict neurodevelopmental disorders, achieving high accuracy and consistency in its measurements [7]. This approach shows how digital tools and machine learning can be integrated to improve early detection and support decision-making in clinical contexts. Building on these approaches, the present work applies clustering and dimensionality reduction techniques to early developmental data, with the goal of examining connections among developmental domains and identifying patterns that may contribute to more efficient screening and diagnosis.

III. EXPERIMENTAL DESIGN

This section outlines the experimental design adopted for this study. First, we describe the dataset preparation process, including data collection and preprocessing. Next, we detail the machine learning models and techniques employed for feature selection and clustering analysis.

A. Dataset

The dataset used in this study comes from the DAUCO TEST survey [8], an assessment designed to measure various

areas of child development. The data includes information from 638 neurotypically developing Spanish children aged between 1 and 84 months. The assessments cover key developmental domains such as language (L), cognition (C), adaptability (A), fine motor (FM), gross motor (GM), and social skills (S).

The data collection process through the DAUCO TEST provides standardized and structured evaluations, offering specific measurements in each of the mentioned areas. The survey was designed to provide a detailed profile of each child's development, aiming to identify potential delays or discrepancies in developmental milestones compared to age-expected ranges.

Data preprocessing included handling missing values and organizing the population into specific age groups to allow a more detailed analysis. The grouping was based on developmental milestones [8]: during the first year of life, children were assessed month by month; from 12 to 24 months, they were grouped into three-month intervals; between 24 and 36 months, into two six-month bands; and from four years onwards, they were grouped annually. The group of children aged 60 to 71 months is the most represented in the dataset (see Fig. 1).

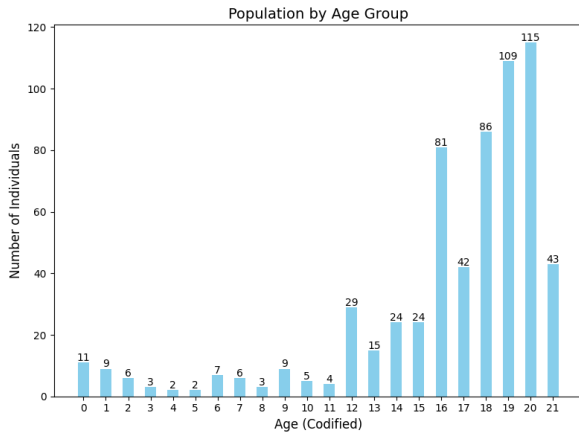


Fig. 1: Number of individuals in each age group.

B. Machine Learning Models

To analyze the data, machine learning techniques were employed. The primary focus was on clustering methods to identify natural patterns in child development and assess whether children deviate from expected growth trajectories. Clustering algorithms, such as *K-Means* [9], *Agglomerative Clustering* [10], *Spectral Clustering* [11], and *Gaussian Mixture Model (GMM)* [12], were applied to this task. Additionally, feature selection techniques, including *Principal Component Analysis (PCA)* [13], for dimensionality reduction, were utilized to identify the most relevant characteristics for accurate clustering.

C. Evaluation Metrics

Given the absence of ground truth labels, internal validation metrics well-known in literature were used to assess clustering quality:

- **Silhouette Score** [14]: Measures cohesion and separation:

$$s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))}$$

where $a(i)$ is the average distance to points in the same cluster, and $b(i)$ is the lowest average distance to points in other clusters. Higher values indicate better-defined clusters.

- **Davies-Bouldin Index** [15]:

$$DB = \frac{1}{N} \sum_{i=1}^N \max_{j \neq i} \left(\frac{s_i + s_j}{d(c_i, c_j)} \right)$$

where s_i is the average distance within cluster i , and $d(c_i, c_j)$ is the distance between centroids. Lower values indicate better clustering.

- **Calinski-Harabasz Index** [16]:

$$CH = \frac{\text{Tr}(B_k)/(k-1)}{\text{Tr}(W_k)/(n-k)}$$

where $\text{Tr}(B_k)$ and $\text{Tr}(W_k)$ are the traces of the between- and within-cluster dispersion matrices, respectively. Higher values suggest better separation.

- **Adjusted Rand Index (ARI)** [17]:

$$ARI = \frac{RI - \mathbb{E}[RI]}{\max(RI) - \mathbb{E}[RI]}$$

where RI is the Rand Index, and $\mathbb{E}[RI]$ its expected value under random clustering. ARI ranges from -1 to 1, with higher values indicating better clustering stability.

- **Normalized Mutual Information (NMI)** [18]:

$$NMI(U, V) = \frac{I(U, V)}{\sqrt{H(U)H(V)}}$$

where $I(U, V)$ is the mutual information, and $H(U)$, $H(V)$ are the entropies of the clusterings. Values close to 1 indicate higher agreement with a reference clustering.

ARI and NMI were also used to assess clustering stability across different random initializations. All metrics were applied to compare algorithms and cluster numbers, as detailed in the Results section.

IV. RESULTS

This section presents the results of the exploratory data analysis, including correlation studies and dimensionality reduction through PCA, as well as the application of clustering techniques. These analyses aim to identify developmental patterns and evaluate their potential for supporting early detection of neurodevelopmental traits.

A. Feature Correlation Analysis

To explore the relationships between the different test items, a correlation analysis was conducted using Pearson's correlation coefficient. The resulting correlation matrix is shown in Fig. 2. Several features exhibited strong correlations, which are detailed below.

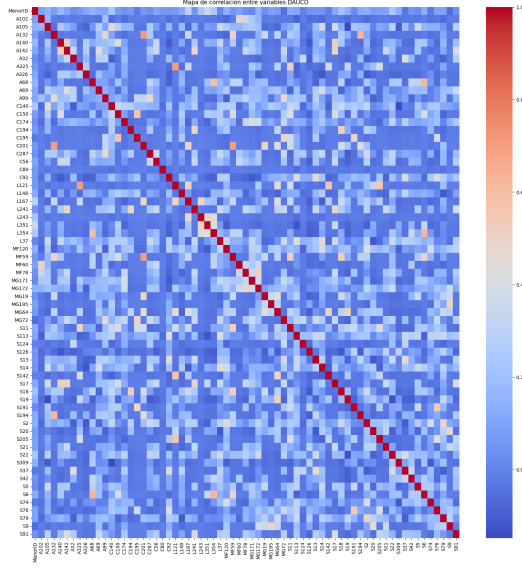


Fig. 2: Correlation matrix of all the items from the DAUCO test, computed using Pearson's correlation coefficient.

Specifically, we examined the pairs of test items that presented a correlation coefficient greater than or equal to 0.70 in the studied population. The following pairs showed high correlations:

- A325 – L121:
 - A325 corresponds to the Adaptive area and assesses the ability to “Controls daytime sphincters,” applicable for ages 30–36, 36–48, 48–59, and 60–71 months.
 - L121 belongs to the Language area and evaluates whether the child “Expresses feelings,” applicable only to the 60–71 months age range.
- A132 – C201:
 - A132 is from the Adaptive area and assesses the ability to “Follows classroom rules,” applicable for ages 60–71 months.
 - C201 is from the Cognitive area and measures the ability to “Identifies the first and last objects in a row,” also applicable only to the 60–71 months age range.
- C201 – MF59:
 - C201 is from the Cognitive area and evaluates the ability to “Identifies the first and last objects in a row,” applicable for the 60–71 months age range.

- MF59 is from the Fine Motor Skills area and assesses the ability to “Copies the numbers 1 through 5,” applicable for the same age range.

These results suggest potential redundancies among certain items, particularly those within the same or overlapping age ranges, which could be considered in the dimensionality reduction and feature selection stages.

B. Dimensionality Reduction with PCA

PCA was applied to identify the optimal number of features for subsequent classification models. The explained variance plot [13] suggests that 32 components capture the 90% of the variance in the dataset as seen in Fig. 3. This result will guide the feature selection process to improve model efficiency while maintaining relevant information.

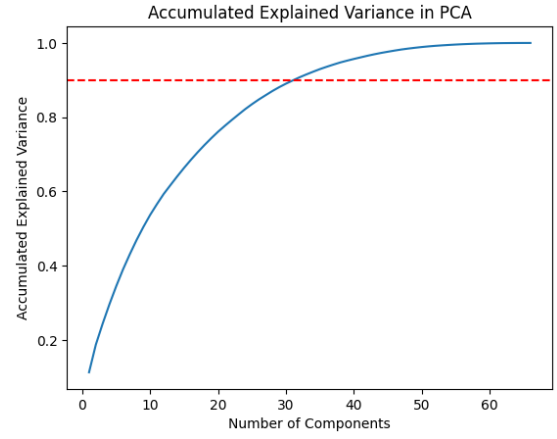


Fig. 3: Accumulated Explained Variance for choosing the number of components of PCA.

C. Clustering Evaluation with Multiple Seeds

To identify the most suitable configurations for the clustering algorithms under study, we conducted a hyperparameter search for Spectral Clustering, Agglomerative Clustering, and GMM. The search involved varying the number of clusters (k), as well as model-specific parameters such as the affinity type, linkage method, or covariance type. Each configuration was evaluated using the silhouette coefficient with a fixed random seed.

For Spectral Clustering, the best silhouette score (0.444) was achieved with $k = 9$, the affinity matrix was obtained using a radial basis function and a kernel coefficient of 0.01. For Agglomerative Clustering, the highest silhouette score (0.604) was observed for $k = 2$, particularly with single, complete, and average linkage methods in combination with Euclidean or Manhattan distances. In contrast, cosine distance led to lower silhouette values, with optimal k values often above 40, suggesting instability or potential overfitting. For GMM, the best result (0.460) was obtained with $k = 3$ components.

TABLE I: Average performance metrics over multiple seeds for selected clustering configurations

Clustering Method	Silhouette Score	Davies-Bouldin Index	Calinski-Harabasz Index
K-Means (k=2)	0.383 \pm 0.014	2.905 \pm 0.084	10.206 \pm 0.234
K-Means (k=5)	0.342 \pm 0.024	1.582 \pm 0.198	8.649 \pm 0.272
K-Means (k=9)	0.322 \pm 0.036	1.140 \pm 0.143	8.418 \pm 0.110
Gaussian Mixture Model	0.451 \pm 0.089	1.227 \pm 0.967	6.685 \pm 1.670
Agglomerative Clustering	0.261 \pm 0.000	1.621 \pm 0.000	7.555 \pm 0.000
Spectral Clustering	0.373 \pm 0.051	1.461 \pm 0.127	7.441 \pm 0.460

After identifying the best-performing hyperparameter sets, we evaluated the stability of each model by repeating the clustering process with multiple random seeds and computing the mean and standard deviation of internal validation metrics. Table I presents the average results for the silhouette score, Davies-Bouldin index, and Calinski-Harabasz index.

Figure 4 illustrates the silhouette scores across various hyperparameter configurations for Spectral Clustering and Figure 5 shows the silhouette scores for Agglomerative Clustering.

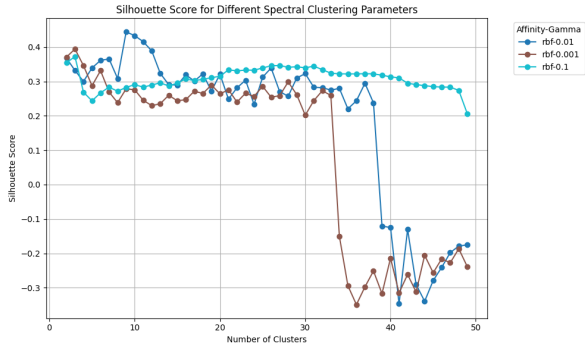


Fig. 4: Silhouette Score for different Spectral Clustering hyperparameters.

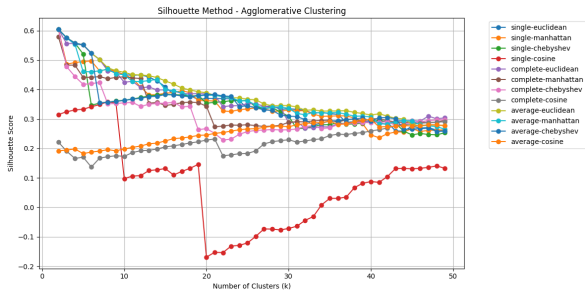


Fig. 5: Silhouette Score for different Agglomerative Clustering hyperparameters.

The variability observed in the ARI and NMI scores across different random initializations underscores the limited stability of the clustering solutions. This analysis was based on the hyperparameters selected using the Silhouette method, as shown in Figures 4, 5, and 6. Despite this optimization, the evaluation metrics indicate a moderate to low clustering performance. These results suggest that the current dataset

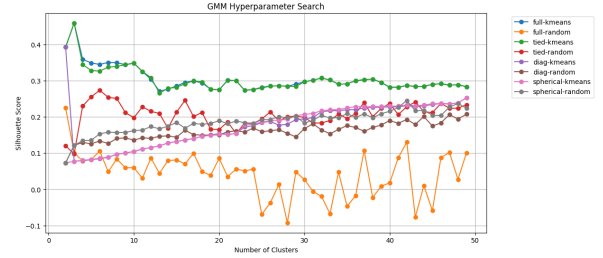


Fig. 6: Silhouette Score for different Gaussian Mixture Model hyperparameters.

may lack a clearly defined cluster structure, potentially due to high feature overlap or the absence of intrinsically separable subgroups (see Table I).

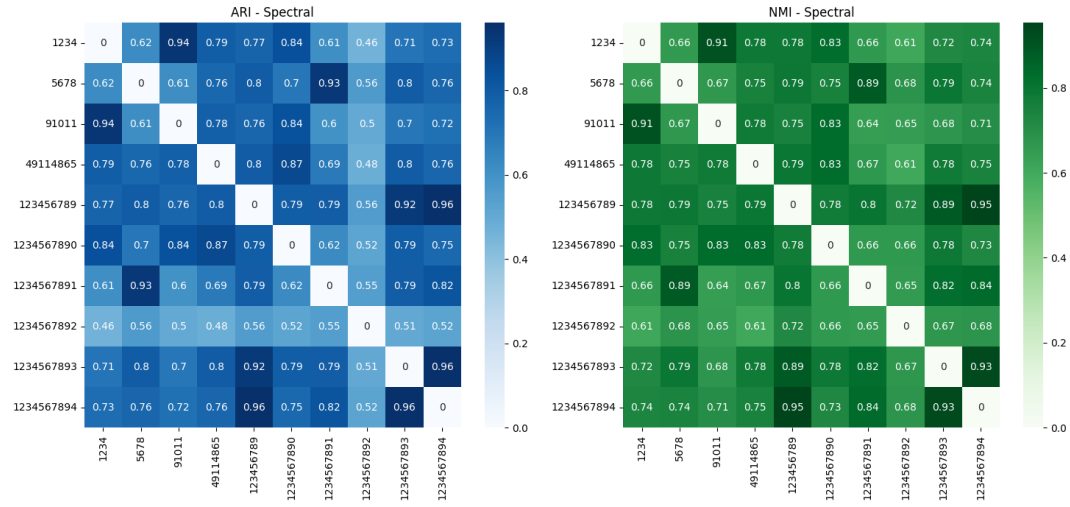
The clustering stability analysis (Figure 7) highlights clear differences between methods. Agglomerative Clustering showed perfect consistency across runs, with ARI and NMI values of 1.0, as expected from its deterministic process. Spectral Clustering also demonstrated acceptable stability, with ARI and NMI typically above 0.5, benefiting from its use of eigenvectors that capture global data structures before the stochastic clustering step.

In contrast, GMM produced unstable results, with ARI and NMI frequently below 0.2 and even negative ARI values in some cases, indicating worse-than-random agreement. This instability is likely due to its sensitivity to initialization and the assumption of Gaussian distributions, which may not hold when the data are overlapping or deviate from this structure.

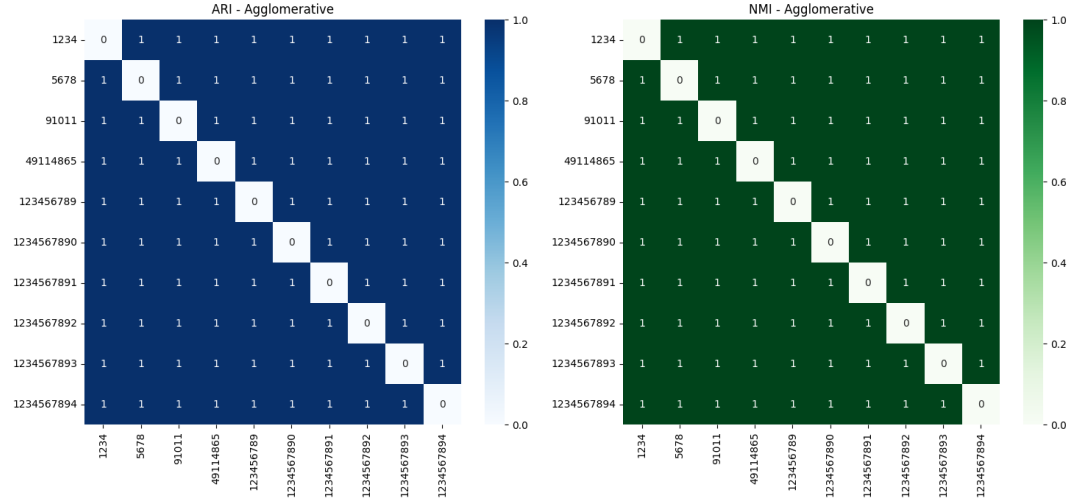
V. DISCUSSION

The exploratory analysis conducted in this study revealed several relevant patterns regarding neurodevelopmental traits assessed through the DAUCO test. One of the most notable findings was the presence of strong correlations between items belonging to different developmental domains, suggesting possible overlaps or synergies in the acquisition of certain skills.

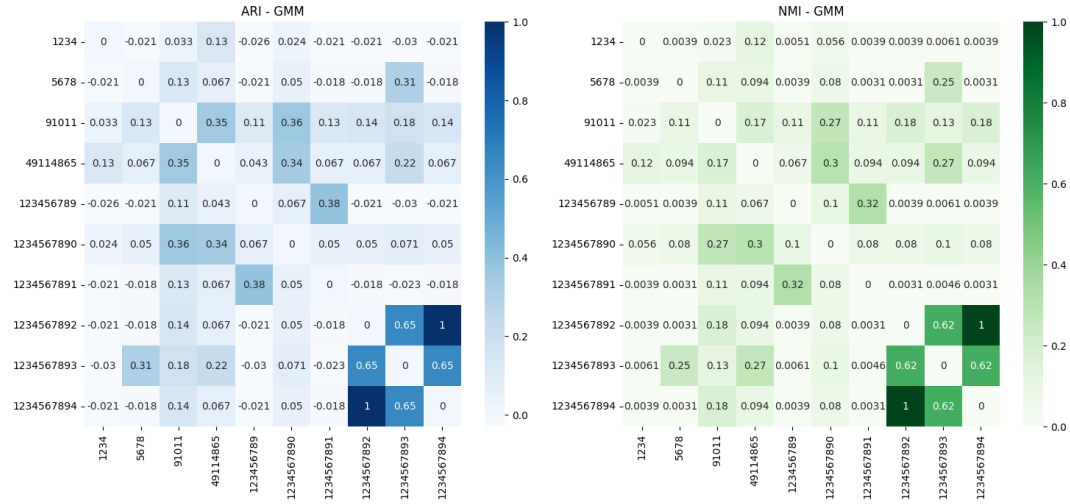
For example, the strong correlation (0.70) between item A132 (“Follows classroom rules”) from the Adaptive area and item C201 (“Identifies the first and last objects in a row”) from the Cognitive area may reflect a relationship between executive functions, which underpin both rule-following behaviors and cognitive tasks involving sequencing and categorization. In [19], author highlights the role of executive functions, such as working memory and inhibitory control, in supporting both adaptive and cognitive domains, which could explain



(a) Comparison of Adjusted Rand Index (ARI) and Normalized Mutual Information (NMI) for Spectral Clustering.



(b) Comparison of ARI and NMI for Agglomerative Clustering.



(c) Comparison of ARI and NMI for Gaussian Mixture Model (GMM).

Fig. 7: Matrix comparison of ARI and NMI for different clustering algorithms.

this observed association. Furthermore, [20] emphasizes that structured environments can simultaneously foster cognitive and adaptive skills development, reinforcing the idea that these areas are interrelated in early childhood.

Similarly, the correlation between item A325 (“Controls daytime sphincters”) from the Adaptive area and item L121 (“Expresses emotions”) from the Language area (0.70) suggests a link between emotional, linguistic, and self-regulation capacities. According to [21], toileting readiness is closely associated with socio-emotional development, particularly the capacity to understand and express emotions, as well as the development of language skills that support these processes. This finding supports the idea that self-regulation behaviors, such as sphincter control, are not isolated milestones but are embedded within broader developmental trajectories involving emotional and communicative competencies.

Moreover, the observed correlation between C201 (“Identifies the first and last objects in a row”) and MF59 (“Copies numbers from 1 to 5”) from the Fine Motor Skills area (0.70) further illustrates the interplay between cognitive and motor domains. As [19] notes, executive functions are not only relevant for cognitive tasks but also play a key role in fine motor tasks that require planning, precision, and sequencing. This convergence aligns with the notion of synergistic development across domains, where progress in one area may facilitate or reflect advancements in another.

Overall, these correlations emphasize the relevance of considering both overlapping and complementary features when applying dimensionality reduction or feature selection strategies. They also highlight the importance of adopting a developmental perspective that acknowledges the interconnected nature of cognitive, adaptive, motor, and socio-emotional skills during early childhood.

VI. CONCLUSIONS AND FUTURE WORK

This study explored neurodevelopmental data using clustering, dimensionality reduction, and correlation analyses. The application of PCA facilitated the reduction of the dataset’s dimensionality while retaining the most relevant variance, improving the visualization and interpretability of the data structure. However, clustering methods did not yield well-defined groups, which may be partly explained by the limited size of the current dataset, particularly in certain age ranges where data scarcity limits the ability to identify stable and generalizable patterns.

Nevertheless, correlation analyses revealed meaningful associations between developmental domains, particularly within the 60–71 months age group. These results highlight the interconnectedness of adaptive, cognitive, fine motor, and language skills, supporting the notion of cross-domain interactions during early development.

Future work will focus on expanding the dataset, particularly by including underrepresented age groups, to enhance the robustness and stability of clustering results. In parallel, refining clustering approaches and investigating how these relationships evolve over time will provide a stronger basis

for developing improved screening tools and early intervention strategies.

REFERENCES

- [1] D. J. Morris-Rosendahl and M. A. Crocq, “Neurodevelopmental disorders—the history and future of a diagnostic concept,” *Dialogues in Clinical Neuroscience*, vol. 22, no. 1, pp. 65–72, 2020.
- [2] S. Reilly, D. V. Bishop, and B. Tomblin, “Terminological debate over language impairment in children: forward movement and sticking points,” *International Journal of Language and Communication Disorders*, vol. 49, no. 4, pp. 452–462, 2014.
- [3] S. M. Luque de Dios, A. Sánchez-Raya, and J. A. Moriana, “Quality of child development scales: a systematic review,” *International Journal of Educational Psychology*, vol. 12, no. 2, pp. 119–148, 2023.
- [4] E. I. Toki, I. G. Tsoulos, V. Santamato, and J. Pange, “Machine learning for predicting neurodevelopmental disorders in children,” *Applied Sciences*, vol. 14, no. 2, p. 837, 2024.
- [5] C. Moreau, C. Deruelle, and G. Auzias, “Machine learning for neurodevelopmental disorders,” in *Machine Learning for Brain Disorders*, O. Colliot, Ed. New York, NY: Springer US, 2023, pp. 977–1007. [Online]. Available: https://doi.org/10.1007/978-1-0716-3195-9_31
- [6] M. Rivard, Z. Mestari, D. Morin *et al.*, “Cluster analysis of clinical features of children suspected to have neurodevelopmental disorders,” *Journal of Autism and Developmental Disorders*, vol. 53, pp. 2409–2420, 2023. [Online]. Available: <https://doi.org/10.1007/s10803-022-05533-y>
- [7] E. I. Toki, I. G. Tsoulos, V. Santamato, and J. Pange, “Machine learning for predicting neurodevelopmental disorders in children,” *Applied Sciences*, vol. 14, no. 2, 2024. [Online]. Available: <https://www.mdpi.com/2076-3417/14/2/837>
- [8] S. M. Luque de Dios, J. A. Moriana, C. Pérez-Dueñas, E. Martínez-Gual, A. Camacho Alcaide, R. González Matilla, A. Antolí, and A. Sánchez-Raya, “Propiedades psicométricas y evaluación del desarrollo infantil con la escala dauco test,” *Revista de Psicología Clínica con Niños y Adolescentes*, 2025, advance online publication.
- [9] J. B. MacQueen, “Some methods for classification and analysis of multivariate observations,” in *Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability*. Berkeley: University of California Press, 1967, pp. 281–297.
- [10] S. C. Johnson, “Hierarchical clustering schemes,” *Psychometrika*, vol. 32, pp. 241–254, 1967.
- [11] U. von Luxburg, “A tutorial on spectral clustering,” 2007. [Online]. Available: <https://arxiv.org/abs/0711.0189>
- [12] C. M. Bishop, *Pattern Recognition and Machine Learning*. Springer, 2006.
- [13] I. T. Jolliffe, *Principal Component Analysis*, 2nd ed., ser. Springer Series in Statistics. New York, NY: Springer, Oct. 2002.
- [14] P. J. Rousseeuw, “Silhouettes: A graphical aid to the interpretation and validation of cluster analysis,” *Journal of Computational and Applied Mathematics*, vol. 20, pp. 53–65, 1987. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/0377042787901257>
- [15] D. L. Davies and D. W. Bouldin, “A cluster separation measure,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-1, no. 2, pp. 224–227, 1979.
- [16] T. Caliński and J. H. and, “A dendrite method for cluster analysis,” *Communications in Statistics*, vol. 3, no. 1, pp. 1–27, 1974. [Online]. Available: <https://www.tandfonline.com/doi/abs/10.1080/03610927408827101>
- [17] D. Steinley, “Properties of the hubert-arable adjusted rand index,” *Psychological Methods*, vol. 9, no. 3, pp. 386–396, 2004. [Online]. Available: <https://doi.org/10.1037/1082-989X.9.3.386>
- [18] A. Strehl and J. Ghosh, “Cluster ensembles - a knowledge reuse framework for combining multiple partitions,” *Journal of Machine Learning Research*, vol. 3, pp. 583–617, 2002.
- [19] A. Diamond, “Executive functions,” *Annual Review of Psychology*, vol. 64, pp. 135–168, 2013. [Online]. Available: <https://doi.org/10.1146/annurev-psych-113011-143750>
- [20] M. J. Guralnick, “A developmental systems model for early intervention,” *Infants & Young Children*, vol. 14, no. 2, pp. 1–18, 2001.
- [21] J.-J. Wyndaele, N. Kaerts, M. Wyndaele, and A. Vermandel, “Development signs in healthy toddlers in different stages of toilet training: Can they help define readiness and probability of success?” *Global Pediatric Health*, vol. 7, 2020.