

Análisis de voces patológicas mediante patrones ordinales y el plano Complexity–Entropy

Materia: Datos, Ecuaciones Diferenciales e Inteligencia Artificial

Cyntia Bonomi

Junio 2026

1 Motivación

Las señales de voz pueden representarse de múltiples formas, pero no está claro qué descriptores capturan mejor las diferencias entre sujetos sanos y patológicos. En este trabajo se explora el espacio entropía–complejidad como representación compacta de la dinámica de la señal, con el objetivo de caracterizar posibles patrones diferenciales entre ambos grupos.

2 Datos

Se empleó la base de datos Saarbrücken Voice Database [1], un conjunto de grabaciones de voz de personas sanas y con distintas patologías vocales en alemán. El corpus incluye la pronunciación sostenida de las vocales /a/, /i/ y /u/ y además la frase “Guten Morgen, wie geht es Ihnen?”.

Los audios están muestreados a 16 kHz y, para este trabajo, se utiliza el subconjunto correspondiente a la vocal sostenida /a/, con una duración promedio aproximada de 1,3 s. Se seleccionó la vocal sostenida /a/ porque permite analizar la voz en condiciones más controladas. A diferencia del habla en frases, que incluye distintos sonidos y variaciones de entonación, una vocal sostenida produce una señal más estable, donde las variaciones observadas están principalmente asociadas al proceso de fonación. Además, en el habla continua, medidas como la entropía pueden reflejar no solo las características intrínsecas de la voz, sino también la variabilidad introducida por cambios fonéticos y prosódicos.

Por este motivo, el uso de vocales sostenidas reduce estos factores de variabilidad, permitiendo obtener una representación más precisa de las características de la señal de voz.

3 Método

3.1 Patrones ordinales

A partir de la serie temporal $\{x_t\}_{t=1}^N$, donde x_t representa el valor de la señal en el instante t , se construyen vectores de embedding de dimensión d con un retardo temporal τ :

$$(x_t, x_{t+\tau}, x_{t+2\tau}, \dots, x_{t+(d-1)\tau})$$

Cada vector se transforma en un patrón ordinal definido por el orden relativo de sus valores. Es decir, en lugar de utilizar las magnitudes originales, se conserva únicamente la información de orden entre los elementos.

Por ejemplo, si tenemos la señal $[4, 7, 2]$, su orden ascendente es $2 < 4 < 7$. El patrón ordinal correspondiente es $(2, 3, 1)$, donde cada entrada indica la posición original de los valores una vez ordenados.

Este procedimiento permite construir una distribución de probabilidad de patrones P , independiente de la amplitud de la señal, ya que los valores son reemplazados por su orden relativo. De este modo, la medida no depende de la escala ni del valor absoluto de la señal, sino que se centra únicamente en cómo cambian sus valores entre sí.

3.2 Entropía (H)

A partir de P se calcula la entropía de permutaciones:

$$H(P) = \frac{-\sum_{i=1}^{d!} p_i \log(p_i)}{\log(d!)}$$

donde:

- $P = \{p_i\}$ es la distribución de patrones ordinales,
- d es la dimensión del embedding.

La entropía mide el grado de desorden de los patrones ordinales. Valores cercanos a 1 indican alta aleatoriedad, mientras que valores cercanos a 0 indican una dinámica más regular o predecible.

3.3 Complejidad (C)

La complejidad mide la estructura intermedia entre orden y aleatoriedad, basada en la desviación respecto a la distribución uniforme:

$$C(P) = Q_J(P, P_e) H(P)$$

donde:

- Q_J es la distancia entre la distribución observada y la distribución uniforme,
- P_e es la distribución uniforme,
- $H(P)$ es la entropía.

Cada señal se representa como un punto en el plano Complejidad–Entropía (H, C), lo que permite comparar su grado de orden y estructura dinámica. Los detalles formales del cálculo pueden encontrarse en [3].

4 Selección de parámetros

La construcción de patrones ordinales requiere definir la dimensión de embedding d y el retardo temporal τ . El parámetro d determina la longitud de los patrones considerados, mientras que τ controla la separación temporal entre las muestras utilizadas para construir cada patrón.

La elección de τ se realizó considerando las escalas temporales características de la voz humana. Estudios previos [2] reportan frecuencias fundamentales promedio cercanas a 120 Hz para hombres y 210 Hz para mujeres, lo que corresponde a períodos de aproximadamente 8,3 ms y 4,8 ms, respectivamente. En este sentido, los valores de τ se seleccionaron para cubrir desde escalas locales entre muestras consecutivas hasta escalas del orden del período fundamental de la señal.

En particular, valores pequeños de τ capturan variaciones rápidas asociadas a la forma local de la onda y a componentes de alta frecuencia, mientras que valores intermedios incorporan información relacionada con la frecuencia fundamental de la voz (percepción de voz más grave o más aguda). Para valores mayores, el análisis se centra en estructuras temporales más lentas asociadas a la periodicidad glotal.

En este trabajo se consideraron valores de $\tau = 4, 8, 16, 32, 64, 128$ y dimensiones de embedding $d = 3, 5, 6, 7$.

5 Resultados

5.1 Analisis Entropía - Complejidad

Con el objetivo de analizar la influencia de los parámetros de embedding sobre la distribución de las señales en el plano entropía-complejidad (H-C), se evaluaron múltiples combinaciones de dimensión de embedding d y retardo temporal τ . A continuación se presentan algunos de los resultados representativos para la discusión. La Figura 1 corresponde a $d = 4$ y $\tau = 4, 16, 32, 64$, mientras que la Figura 2 muestra los resultados para $d = 7$ utilizando los mismos valores de τ .

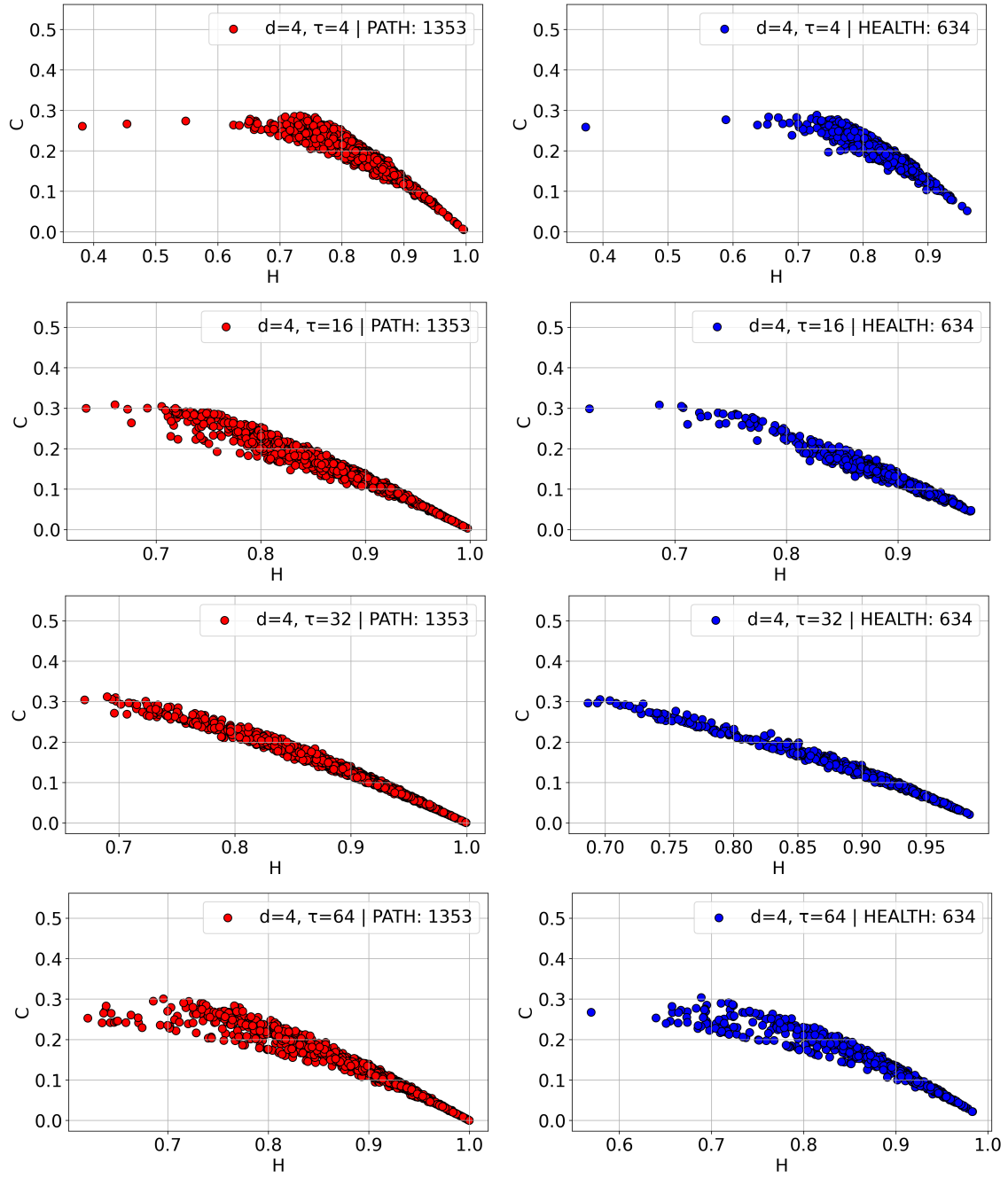


Figure 1: Plano entropía-complejidad para las dimensiones de embedding $d = 4$ y utilizando distintos retardos temporales $\tau = 4, 16, 32, 64$. Se muestran por separado las distribuciones correspondientes a voces sanas y patológicas.

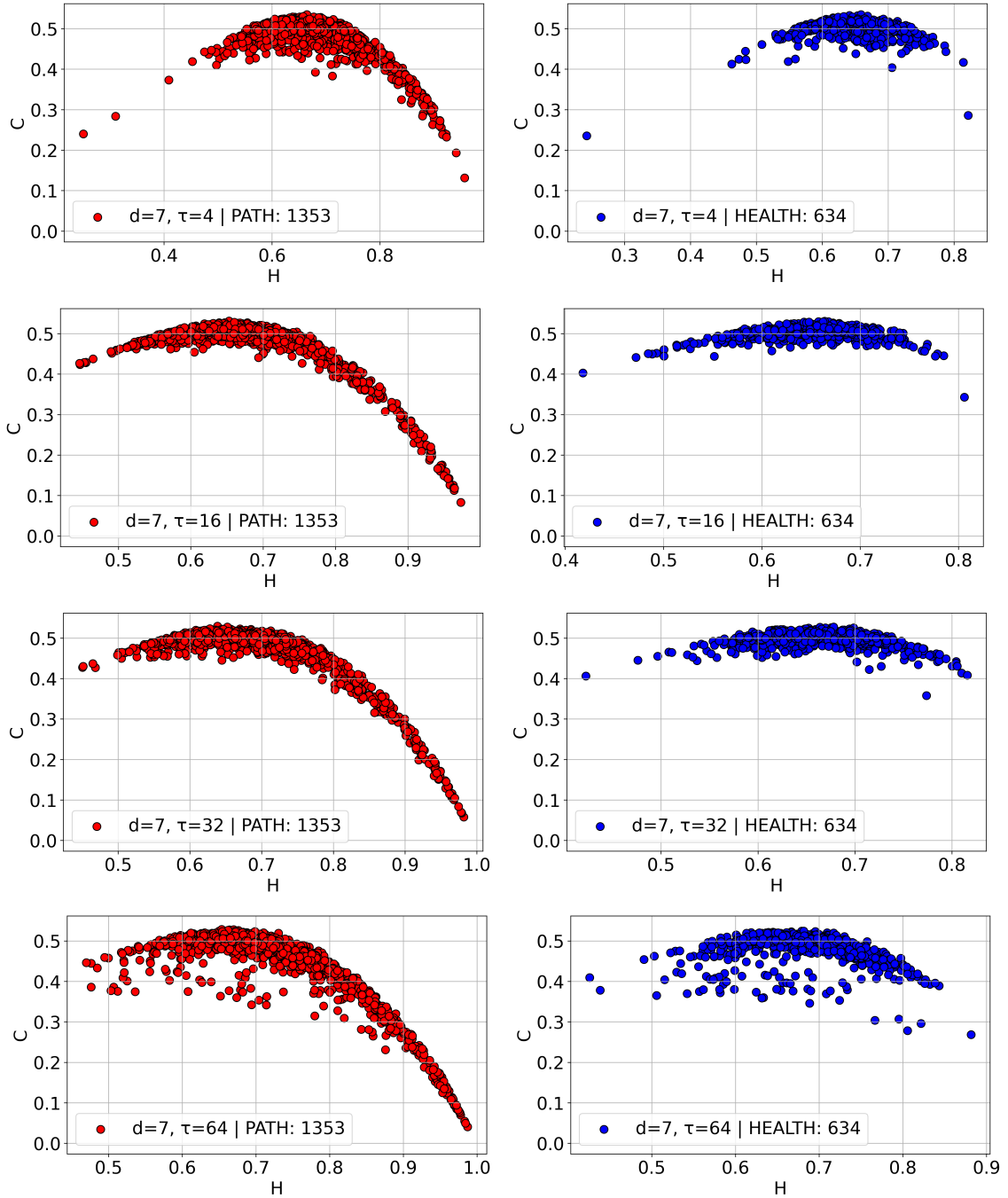


Figure 2: Plano entropía-complejidad para las dimensiones de embedding $d = 7$ y utilizando distintos retardos temporales $\tau = 4, 16, 32, 64$. Se muestran por separado las distribuciones correspondientes a voces sanas y patológicas.

Las Figuras 1, 2 presentan los resultados obtenidos para $d = 4$ y $d = 7$ respectivamente, seleccionados como casos representativos de baja y alta dimensión, respectivamente. Para $d = 4$, las distribuciones se concentran en una región relativamente acotada del plano H - C y alcanzan valores moderados de complejidad. Al incrementar la dimensión hasta $d = 7$, la estructura arqueada característica del plano se vuelve más evidente y se observa una expansión significativa del rango de complejidad alcanzado por las señales. Este comportamiento sugiere que dimensiones mayores permiten capturar información temporal adicional de la dinámica vocal. Por otra parte, el retardo temporal también modifica la distribución de los puntos dentro del plano H - C . A medida que

aumenta τ , las distribuciones presentan una mayor dispersión y se observan cambios en la ocupación de determinadas regiones del espacio H-C. Sin embargo, en todas las configuraciones analizadas persiste una importante superposición entre las muestras sanas y patológicas.

Particularmente para $d = 7$, ciertos valores de τ revelan una mayor dispersión de las muestras patológicas hacia regiones de menor complejidad y alta entropía. No obstante, esta tendencia no define una separación clara entre grupos, ya que las distribuciones continúan compartiendo gran parte del espacio H-C.

5.2 Análisis Entropía Complejidad ($d = 7$ $\tau = 32$)

Para analizar con mayor detalle la posible separación entre voces sanas y patológicas en el plano entropía-complejidad, se considera la configuración $d = 7$ y $\tau = 32$, la cual había mostrado previamente una ligera concentración de muestras patológicas en regiones de baja complejidad. En la Figura 3 se presenta la superposición de ambas clases en el mismo espacio H-C. Esta visualización permite evaluar de manera directa si dicha tendencia corresponde a una separación real entre grupos o a un efecto de visualización observado en representaciones separadas. Podemos ver que en términos generales, ambos grupos presentan una importante superposición y ocupan la misma región principal del espacio H-C. Sin embargo, se observa una diferencia sutil en la región de alta entropía y baja complejidad. Mientras que las muestras sanas se concentran principalmente en la zona superior de la curva, un subconjunto de las voces patológicas se extiende hacia valores de complejidad progresivamente menores, generando una cola que alcanza regiones próximas a $H \approx 1$ y $C \approx 0$. Esta tendencia sugiere la presencia de señales patológicas que, para esta escala temporal de análisis, exhiben menos estructura y más cercanía a un comportamiento aleatorio.

No obstante, dicha cola representa únicamente una fracción del conjunto patológico y no produce una separación clara entre clases, ya que la mayor parte de las muestras de ambos grupos permanece fuertemente superpuesta.

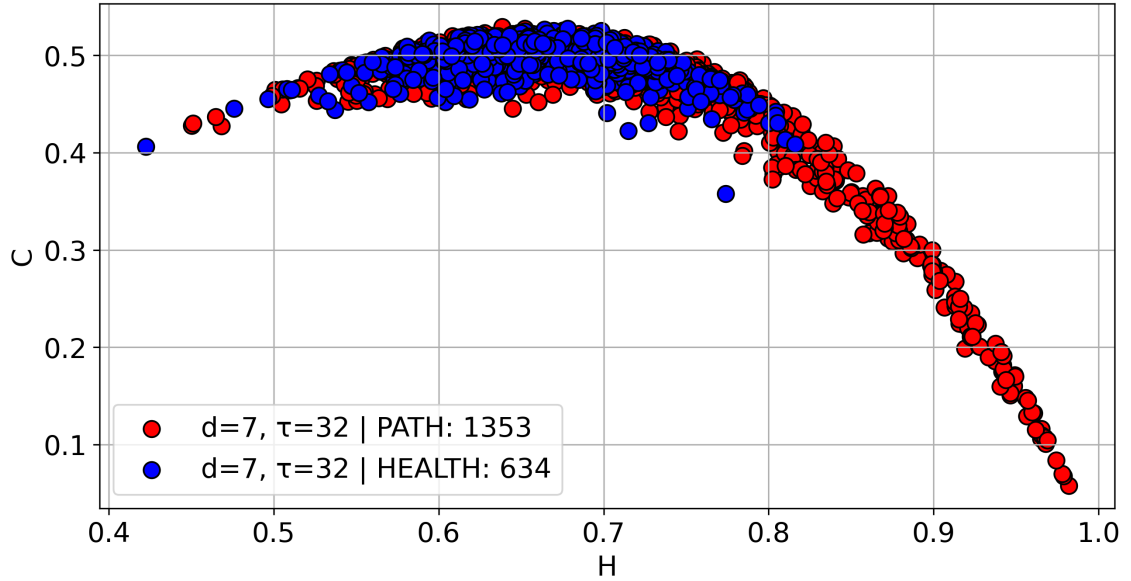


Figure 3: Plano entropía-complejidad para la distribución conjunta de voces sanas y patológicas en el plano entropía-complejidad para $d = 7$ y $\tau = 32$

5.3 Análisis Entropía Complejidad ($d = 7$ best τ)

Con el fin de evaluar si la dispersión observada para valores fijos de τ estaba asociada a la elección de una escala temporal común para todas las señales, se realizó un análisis adaptativo en el que a cada señal se le asignó el valor de τ que maximiza su complejidad. La Figura 4 muestra la

distribución resultante en el plano H-C y permite compararla con el caso de $\tau=32$ presentado en la Figura 3.

En la Figura 4, se observa una reducción significativa de la cola en la región de alta entropía y baja complejidad. En paralelo, las muestras patológicas tienden a desplazarse hacia valores más altos de complejidad, concentrándose en la zona superior del plano H-C.

A pesar de este cambio en la distribución, las muestras sanas y patológicas continúan mostrando una fuerte superposición en casi toda la región del espacio H-C, sin que se observen áreas claramente exclusivas para alguno de los grupos.

La reducción de la cola sugiere que parte de la dispersión observada para τ fijo estaba asociada a diferencias en la escala temporal característica de cada señal. Al permitir que cada muestra sea analizada en el τ donde alcanza su máxima complejidad, la distribución se vuelve más homogénea. Esto indica que algunas señales, particularmente dentro del grupo patológico, alcanzan niveles elevados de complejidad en escalas temporales diferentes, lo que podría reflejar una mayor heterogeneidad dinámica.

Sin embargo, esta reorganización de las muestras en el plano H-C no se traduce en una mejor separabilidad entre clases. Incluso bajo este enfoque adaptativo, las métricas globales de entropía y complejidad no permiten una discriminación clara entre voces sanas y patológicas.

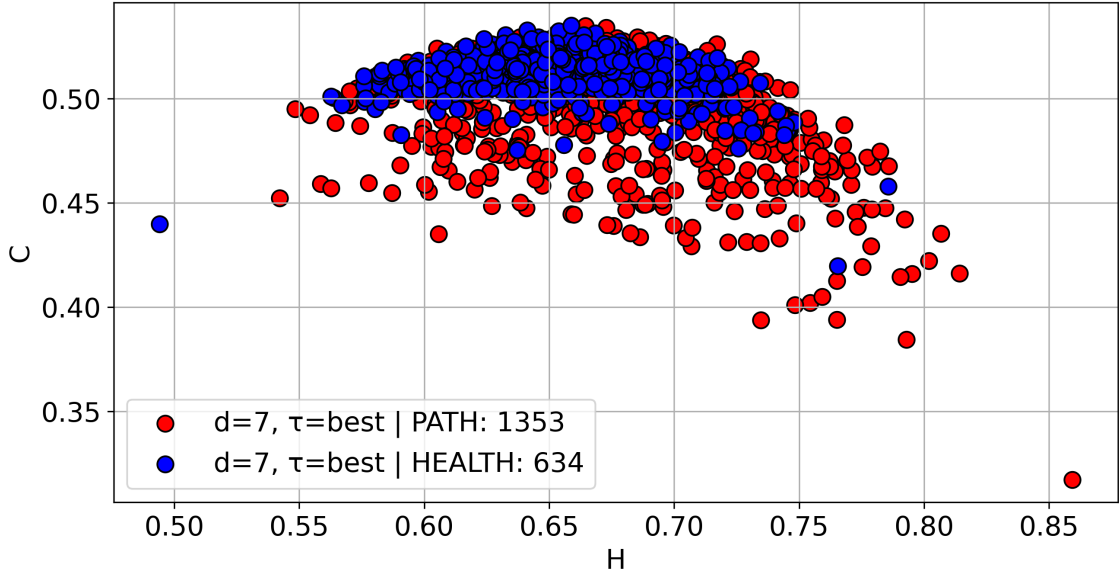


Figure 4: Plano entropía-complejidad para la distribución conjunta de voces sanas y patológicas en el plano entropía-complejidad para $d = 7$ y τ que maximiza la complejidad, manteniendo fija la dimensión de embedding en $d = 7$

5.4 Análisis entropía-complejidad en subgrupos

Dado que las características acústicas de la voz pueden verse influenciadas por factores demográficos, se exploró la posible influencia de la edad y el género sobre la distribución de las muestras en el plano H-C. Para ello, se analizó la distribución correspondiente al conjunto completo de datos y a distintos subgrupos definidos por género, rango etario y la combinación de ambas variables.

En todos los casos, la distribución mantuvo la misma estructura general, sin cambios sustanciales en el patrón global. Sin embargo, se observó una mayor presencia de la cola de alta entropía y baja complejidad en sujetos mayores de 50 años.

Es importante destacar que los subgrupos analizados presentan un desbalance significativo en el número de muestras de cada clase (sanos y patológicos). En particular, algunos rangos etarios muestran una mayor proporción de sujetos patológicos (adultos mayores), mientras que en otros predominan los sujetos sanos (jóvenes).

Este aspecto debe considerarse al interpretar los resultados de discriminación obtenidos en cada subgrupo. Si bien la métrica AUC es menos sensible al desbalance de clases que otras medidas de desempeño, diferencias importantes en la composición de los grupos pueden afectar la robustez y comparabilidad de las estimaciones.

5.5 Análisis por subgrupos del AUC para entropía (H) y complejidad (C)

Si bien las distribuciones en el plano entropía-complejidad (H-C) permiten identificar tendencias generales, su interpretación es principalmente visual y no cuantifica de forma directa la capacidad discriminativa de las métricas consideradas. Para complementar este análisis, se evaluó cuantitativamente el desempeño de la entropía y la complejidad mediante el área bajo la curva ROC (AUC).

Para cada combinación de parámetros (d, τ) , se calculó el AUC utilizando los valores de H o C como variable de entrada y la condición clínica (sano/patológico) como etiqueta. El AUC permite cuantificar la capacidad de separación entre ambas clases de manera independiente de un umbral de decisión específico.

Este análisis se realizó sobre el conjunto completo de datos y sobre distintos subgrupos definidos por género, rango etario y su combinación. Sin embargo, se presentan principalmente los resultados correspondientes al análisis por rango etario, ya que en este caso se observan los patrones más consistentes. En cambio, las particiones más finas (combinación de género y edad) presentan tamaños de muestra reducidos, lo que puede afectar la estabilidad de las estimaciones debido al desbalance entre clases.

Resultados AUC para el subgrupo: rango etario

Table 1: AUC para complejidad (C)

Grupo	AUC(C)	d	τ óptimo	sanos	patológicos
(18–30]	0.58	7	4	456	172
(30–50]	0.70	7	16	136	446
> 50	0.71	6	8	42	735

Table 2: AUC para entropía (H)

Grupo	AUC(H)	d	τ óptimo	sanos	patológicos
(18–30]	0.57	3	4	456	172
(30–50]	0.65	4	3	136	446
> 50	0.70	7	4	42	735

Las Tablas 1 y 2 presentan los valores de AUC obtenidos para la complejidad y la entropía, respectivamente, junto con los valores óptimos de (d, τ) y el número de sujetos por clase en cada subgrupo.

En ambos casos se observa una tendencia de un leve incremento del AUC con la edad. Para la complejidad, los valores aumentan desde 0.58 en el grupo de 18–30 años hasta 0.71 en el grupo de mayor edad. Un comportamiento similar se observa en la entropía, con un incremento desde 0.57 hasta 0.70.

Este patrón sugiere una mayor capacidad de separación entre voces sanas y patológicas en los grupos de mayor edad, mientras que en los grupos más jóvenes la discriminación es menor. Asimismo, la complejidad muestra valores de AUC ligeramente superiores a los de la entropía en todos los subgrupos, aunque ambos indicadores preservan la misma tendencia global. En cuanto a los parámetros óptimos, se observa que los valores de (d, τ) varían entre subgrupos, con una tendencia general a bajos en casi todos los grupos.

6 Conclusiones

En este trabajo se evaluó la capacidad discriminativa de las métricas entropía y complejidad, considerando su dependencia con los parámetros de embedding d y retardo temporal τ . El análisis se realizó tanto sobre el conjunto completo de datos como sobre subgrupos definidos por género y rango etario, lo que permitió estudiar la estabilidad de las métricas frente a variaciones demográficas.

Los resultados obtenidos mediante el área bajo la curva ROC (AUC) muestran que ambas métricas presentan capacidad de discriminación entre sujetos sanos y patológicos, aunque con diferencias sistemáticas en su desempeño. En general, la complejidad tiende a presentar valores de AUC ligeramente superiores a los de la entropía, especialmente en subgrupos de mayor edad, donde se alcanzan los mejores resultados del estudio. Aunque cabe señalar que los adultos mayores cuentan con el mayor desbalance entre sanos (42 muestras) y patológicos (735 muestras).

Asimismo, se observó que los subgrupos más jóvenes presentan una menor separabilidad entre clases, lo que se refleja en valores de AUC más bajos para ambas métricas. Este comportamiento podría estar relacionado con una menor manifestación de patrones patológicos en etapas tempranas o con una mayor similitud dinámica entre grupos en estas edades.

Desde el punto de vista metodológico, se identificó un desbalance importante en la cantidad de sujetos sanos y patológicos dentro de los subgrupos analizados. Este factor puede influir en la estabilidad de la estimación del AUC, especialmente en aquellos grupos con menor número de muestras. Por lo tanto, aunque el AUC constituye una métrica adecuada para evaluar separabilidad, las diferencias observadas entre subgrupos deben interpretarse considerando esta limitación.

En conjunto, los resultados sugieren que las métricas basadas en entropía y complejidad capturan propiedades relevantes de la señal de voz, aunque su capacidad discriminativa es limitada, siendo la complejidad ligeramente más informativa en la mayoría de los escenarios evaluados. Asimismo, su desempeño está condicionado por las características demográficas de la población y por la distribución de los datos, lo que destaca la importancia de considerar análisis estratificados y balanceados.

References

- [1] Manfred Pützer and Jacques Koreman. *A German Database of Pathological Vocal Fold Vibration*. Online database. Available at: <https://stimddb.coli.uni-saarland.de/>. 1997.
- [2] Hartmut Traunmüller and Anders Eriksson. “The frequency range of the voice fundamental in the speech of male and female adults”. In: *Unpublished manuscript* 11 (1995).
- [3] Massimiliano Zanin and Felipe Olivares. “Ordinal patterns-based methodologies for distinguishing chaos from noise in discrete time series”. In: *Communications Physics* 4.1 (2021), p. 190.