GRADO en INGENIERÍA BIOMÉDICA

UVa

Taller 3 de Escritura del TFG/TFM

Capítulos de Resultados y Discusión

Gonzalo C. Gutiérrez Tobal, Adrián Martín Montero, Verónica Barroso García, Ana Martín Fernández, Jorge Jiménez García, María Herrero Tudela, Víctor Gutiérrez de Pablo

Lugar: **Aula B04 de la Facultad de Medicina**
Fecha: **16/05/2025**
Hora: **16:00-18:00**

Taller enmarcado en el Proyecto de Innovación Docente “Creación de un espacio virtual colaborativo para la formación, tutorización integral y comunicación técnica en TFGs y TFMs de Ingeniería Biomédica”

Virt UVa

Contenidos

Parte 1: Capítulo de Resultados

1.1 ¿Para qué sirve?

1.2 Contenidos y estructura

1.3 Tablas y figuras típicas

Parte 2: Capítulo de Discusión

2.1 ¿Para qué sirve?

2.2 Contenidos y estructura

2.3 Checklist del Capitulo de Discusión

Ejercicios prácticos

Desarrollo de una figura de Resultados con Flourish

Ejemplo de texto asociado a la figura

Ejemplo de estructura de Discusión

Diferencia entre Resultados y Discusión: Kahoot

Parte 1. Capítulo de Resultados

GRADO en INGENIERÍA BIOMÉDICA

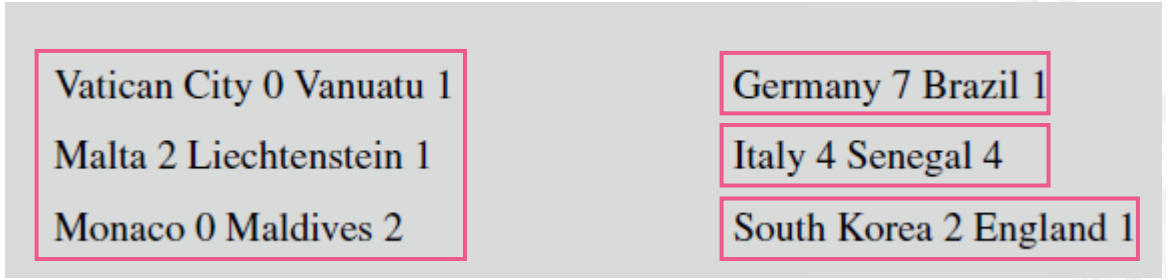
Parte 1: Capítulo de Resultados

1.1 ¿Para qué sirve?

Lo obvio: para presentar los resultados más **relevantes** de vuestro trabajo
Lo menos obvio: para presentar los resultados **inesperados o negativos** de vuestro trabajo

Muy importante en este capítulo es el "cómo":

1. De manera **completa**: todos los resultados de interés, en texto, figuras, tablas…
2. De manera **sencilla**: que el lector lo entienda fácilmente
3. De manera **complementaria**. **NO** se pone la misma información en el texto y en las figuras y en las tablas. En el texto se resalta lo más importante de figuras y tablas, incluidas posibles conexiones entre diferentes resultados.
4. Se **muestran / presentan** los resultados, pero **NO** se analizan los resultados (no se da opiniones sobre los resultados, no se interpretan, no se compara con otros trabajos, etc.)



Idea clave: primero decidir qué resultados son representativos, y luego organizarlos en secuencia para resaltar las respuestas a los objetivos, hipótesis o preguntas del TFG / TFM

GRADO en INGENIERÍA BIOMÉDICA

Parte 1: Capítulo de Resultados

1.2 Contenidos y estructura

El contenido debe responder a:
¿Qué he encontrado?
¿Qué NO he encontrado?
¿Qué he encontrado y no esperaba encontrar? (lo que contradiga vuestra hipótesis / pregunta de investigación / objetivo

Estructura típica:

Seguir el orden que pusisteis en la sección de Métodos y poner tablas y figuras secuencialmente para responder a las preguntas de arriba

Acordaos de que estáis contando una historia y que precisamente la contáis de acuerdo con los resultados

De manera práctica: vais comentando figura a figura en el orden que habéis elegido, y para cada una:

1. Remarcáis los resultados que responden a vuestro problema
2. Esbozais los resultados secundarios
3. Dais información de apoyo
4. Mencionáis resultados contradictorios y explicáis porqué son anómalos

GRADO en INGENIERÍA BIOMÉDICA

Parte 1: Capítulo de Resultados

1.2 Contenidos y estructura

2 posibles formas de comenzar:

- Panorama general de tus resultados: “En general, nuestros resultados mostrados más abajo muestran que…”; “Los tres principales resultados del trabajo son”:
- Más común: simplemente empezar “En la Tabla / Figura 1 se muestra…”, “ En la Tabla / Figura 2 se presenta…”

Tiempos verbales:

Pasado o presente alternando voz pasiva y activa

Parte 1: Capítulo de Resultados

1.3 Tablas y figuras típicas

Tabla de comparación de grupos

TABLE III
FEATURE VALUES FOR THE SEVERITY GROUPS (MEAN ± SD)

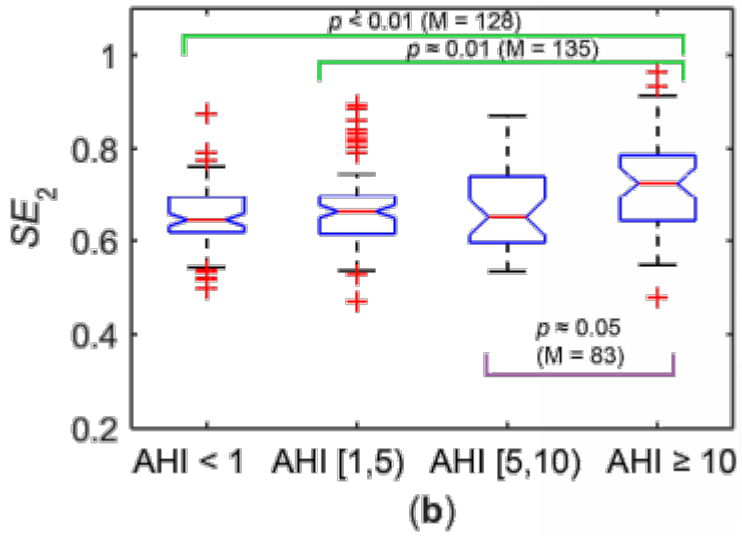
Feat.	no-SAHS	Mild	Moderate	Severe	*p*-value
MA (10^{-4})	2.012 ± 1.091	2.854 ± 1.460	5.148 ± 3.134	13.736 ± 11.360	$\ll 0.01$
mA (10^{-4})	1.359 ± 0.729	1.849 ± 0.930	2.903 ± 1.294	6.225 ± 4.498	$\ll 0.01$
M_{f1} (10^{-4})	1.670 ± 0.912	2.296 ± 1.131	3.900 ± 1.886	9.400 ± 7.295	$\ll 0.01$
M_{f2} (10^{-5})	2.140 ± 1.424	3.193 ± 2.428	7.418 ± 8.268	24.864 ± 27.774	$\ll 0.01$
M_{f3}	0.190 ± 0.540	0.259 ± 0.512	0.149 ± 0.619	0.429 ± 0.689	0.19^{+}
M_{f4}	2.154 ± 0.590	2.269 ± 0.569	2.298 ± 0.637	2.608 ± 1.115	0.41^{+}
WD	0.046 ± 0.019	0.052 ± 0.029	0.063 ± 0.041	0.086 ± 0.056	0.003^{+}
MF	0.038 ± 0.001	0.038 ± 0.002	0.037 ± 0.002	0.036 ± 0.002	0.004^{+}
SpecEn (10^{-1})	9.963 ± 0.032	9.958 ± 0.046	9.924 ± 0.168	9.882 ± 0.134	0.024^{+}
CTM (10^{-1})	9.993 ± 0.007	9.988 ± 0.015	9.987 ± 0.009	9.963 ± 0.023	$\ll 0.01$
LZC	0.057 ± 0.009	0.057 ± 0.007	0.057 ± 0.006	0.058 ± 0.007	0.71^{+}
SampEn	0.059 ± 0.012	0.063 ± 0.014	0.062 ± 0.016	0.058 ± 0.014	0.18^{+}

$^{+}$ *Not lower than Bonferroni correction (p-value = 0.01/6).*

Parte 1: Capítulo de Resultados

1.3 Tablas y figuras típicas

Boxplots y Violin plots



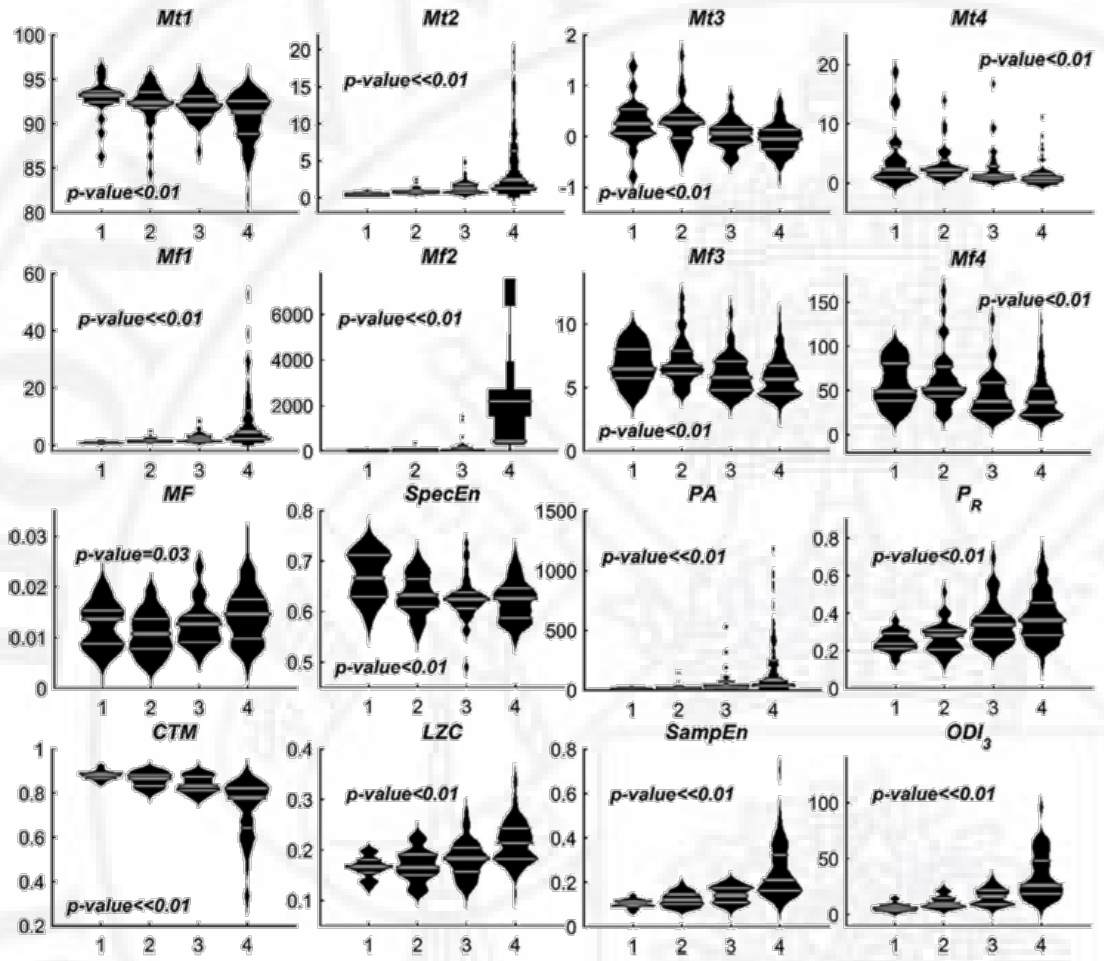


Fig. 2. Violin plots of each extracted feature divided by SAHS-severity degree (only training set). Numbers in x-axis represent the severity of SAHS: 1 stands for no-SAHS, 2 for mild, 3 for moderate, and 4 for severe. All *p*-values from Kruskal-Wallis test were corrected using the Bonferroni criterion.

GRADO en INGENIERÍA BIOMÉDICA

Parte 1: Capítulo de Resultados

1.3 Tablas y figuras típicas

Radar o Spider plots

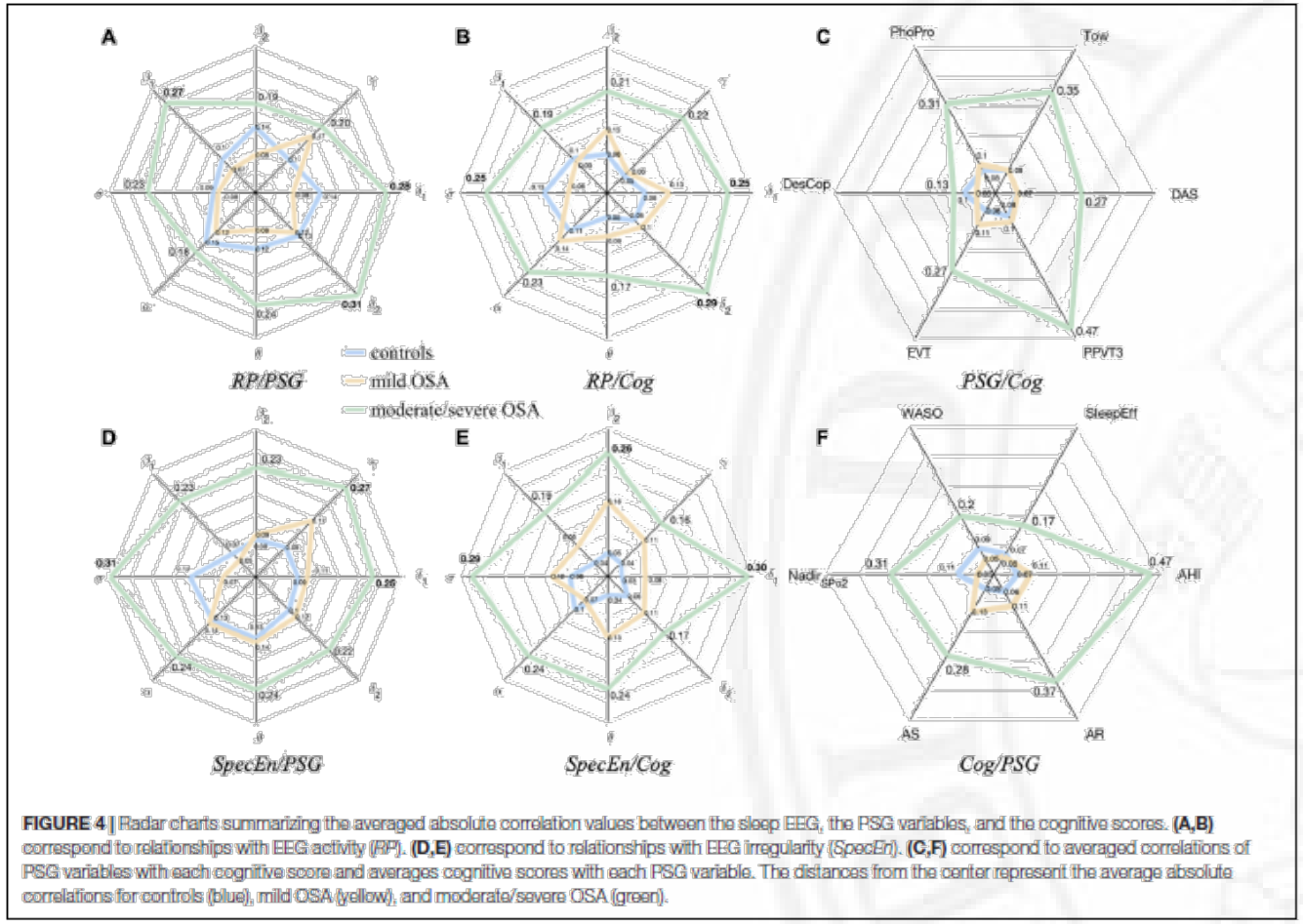


FIGURE 4 | Radar charts summarizing the averaged absolute correlation values between the sleep EEG, the PSG variables, and the cognitive scores. **(A,B)** correspond to relationships with EEG activity (*RP*). **(D,E)** correspond to relationships with EEG irregularity (*SpecEn*). **(C,F)** correspond to averaged correlations of PSG variables with each cognitive score and averages cognitive scores with each PSG variable. The distances from the center represent the average absolute correlations for controls (blue), mild OSA (yellow), and moderate/severe OSA (green).

Parte 1: Capítulo de Resultados

1.3 Tablas y figuras típicas

Redes de asociación o correlación

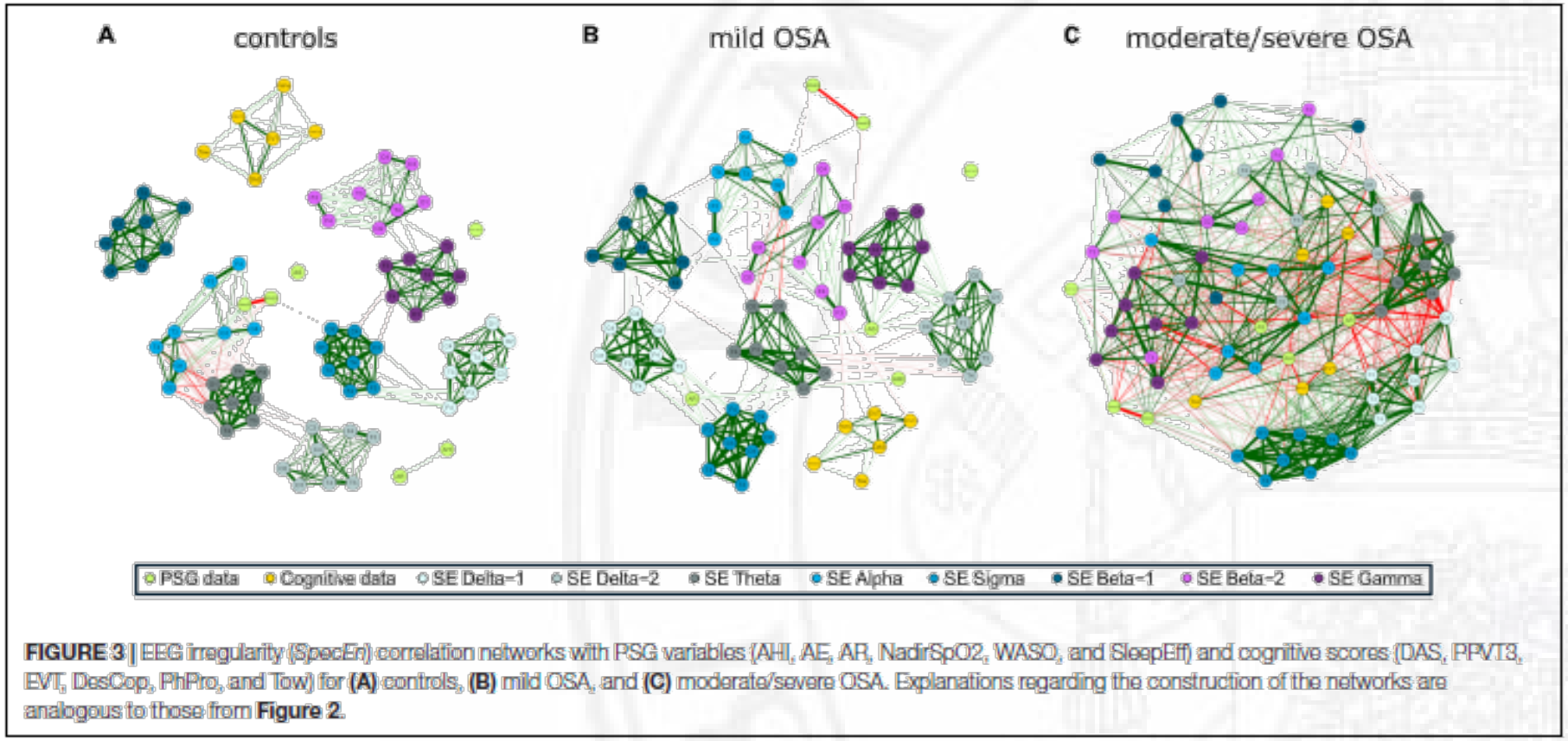


FIGURE 3 | EEG irregularity (SpecEn) correlation networks with PSG variables (AHI, AE, AR, NadirSpO2, WASO, and SleepEff) and cognitive scores (DAS, PPVT3, EVT, DesCop, PhPro, and Tow) for **(A)** controls, **(B)** mild OSA, and **(C)** moderate/severe OSA. Explanations regarding the construction of the networks are analogous to those from **Figure 2**.

GRADO en INGENIERÍA BIOMÉDICA

Parte 1: Capítulo de Resultados

1.3 Tablas y figuras típicas

Matriz de confusión (tabla) y Tabla de rendimiento

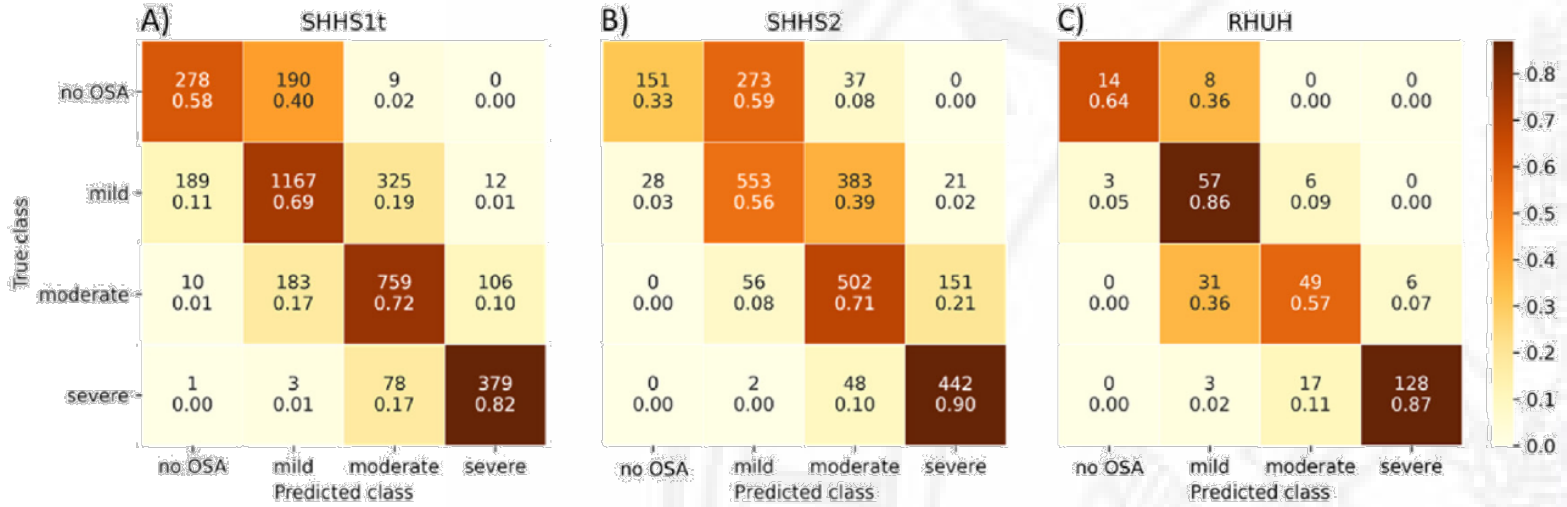


Table 3
Diagnostic performance on the clinical AHI thresholds used to demarcate OSA severity categories (5 e/h, 15 e/h, and 30 e/h).

	$SHHS1_t$			SHHS2			RHUH		
	5 e/h	15 e/h	30 e/h	5 e/h	15 e/h	30 e/h	5 e/h	15 e/h	30 e/h
Se (%)	93,77	87,03	82,21	98,70	95,17	89,83	99,00	85,47	86,49
Sp (%)	58,28	84,06	96,34	32,80	69,50	92,01	63,64	93,18	96,55
PPV (%)	93,89	79,26	76,26	87,44	72,16	71,99	97,38	97,09	95,52
NPV (%)	58,16	90,25	97,43	84,36	95,54	97,54	82,35	70,69	89,36
LR+	2,25	5,46	22,46	1,47	3,12	11,24	2,72	12,53	25,07
LR-	0,11	0,15	0,18	0,04	0,07	0,11	0,02	0,16	0,14
Acc (%)	89,18	85,28	94,58	87,23	81,14	91,61	96,58	87,58	91,93

Acc: accuracy, LR+/LR-: positive and negative likelihood ratio, PPV/NPV: positive and negative predictive value, Se/Sp: sensitivity and specificity.

Parte 2. Capítulo de Discusión

GRADO en INGENIERÍA BIOMÉDICA

Parte 2: Capítulo de Discusión

2.1 ¿Para qué sirve?

El capítulo más importante del documento

Es donde se demuestra todo lo que habéis aprendido y que lo dominái s

Es el culmen de la trama de vuestra historia: vamos a descubrir al asesino

¿Qué se discute (debate)? Los métodos seguidos, los experimentos realizados y los RESULTADOS principales

Parte 2: Capítulo de Discusión

2.2 Contenidos y estructura

El contenido debe responder a:

1. ¿Cuáles son mis hallazgos más importantes?
2. ¿Apoyan estos hallazgos lo que me propuse demostrar al inicio del trabajo?
3. ¿Cómo se comparan mis hallazgos con los de otros investigadores? ¿Qué tan consistentes son?
4. ¿Cuál es mi interpretación personal de mis hallazgos?
5. ¿Qué otras interpretaciones posibles existen?
6. ¿Cuáles son las limitaciones de mi estudio? ¿Qué otros factores podrían haber influido en mis hallazgos?
7. ¿Contribuyen mis interpretaciones a una nueva comprensión del problema que he abordado? En ese caso, ¿sugieren un avance respecto al trabajo de otros?

Parte 2: Capítulo de Discusión

2.2 Contenidos y estructura

Propuesta de estructura (muy generalizable): con o sin subapartados (en TFG/TFM mejor con)

1. Pequeño resumen de contribución o contribuciones **más importantes** con información accesoria mínima pero suficiente. NO hacer un resumen de todo lo que habéis hecho

In this study, we have accomplished substantial advances in SpO_2 characterization by extracting up to 32 overnight features from different analytical approaches. They serve to develop a novel OSA-specific LSBoost-based model that has the ability to accurately estimate AHI from single-channel oximetry data obtained in the patients' home. Furthermore, this model displays a high level of agreement with the actual PSG-derived AHI, and high diagnostic performance in both referral and non-referral cohorts. Our analytical approaches not only allowed us to develop a clinically useful tool, but also explain the model through the SpO_2 data used by each of the 199 base classifiers of the ensemble, thus decreasing the common 'black box' perception of automatic models.

GRADO en INGENIERÍA BIOMÉDICA

Parte 2: Capítulo de Discusión

2.2 Contenidos y estructura

Propuesta de estructura (muy generalizable): con o sin subapartados (en TFG/TFM mejor con)

2. Discutir párrafo a párrafo, o sección a sección, todo lo que sea relevante para justificar ese párrafo anterior

4.1. Explaining the AHI estimation

ODI3 accounted for 85.7% of the relative importance, i.e., the contribution to the final overall AHI estimation. This is consistent with previous studies in which *ODI3* reflected the principal OSA-related informative component regarding SpO_2 [17,43]. However, *ODI3* alone underestimates AHI as not all apneic events lead to a desaturation [17,18,35]. Underestimation is not generally observed in our model, suggesting that the information contained in the remaining features is counteracting this effect. This would be supported by the similarity between the relative importance accounted by these features (14.3%) and the average amount of apneic events not accompanied by a 3% desaturation in SHHS dataset (11.5%) [18]. Most of this remaining relative importance is composed of the eight features mentioned in Results section, which are related to signal complexity (*LZC*, $msEnt_{area}$), irregularity ($mSEnt_{scale}$, *SampEn*), SpO_2 values distribution over time (*M4t*) and frequencies (*WD*), and the number and amplitude of recursive desaturations lasting 30 to 70 s irrespective of specific percentages of decrease ($M1f_{BOI}$, MA_{BOI}). Therefore, we posit that the overnight patterns characterized by these features reflect additional information about OSA, regardless of whether it is related to other events involved in AHI definition (e.g. arousals) or not (e.g. hypoxic burden). However, further research will be required to define the specific relationships between these parameters and other OSA-related effects.

4.3. Diagnostic ability and comparison with the state-of-the-art studies

Our OSA-specific LSBoost model reached high diagnostic ability in both non-referral and referral datasets. It additionally outperformed other regression and classification machine-learning methods evaluated on the same datasets. Previous studies also focused on the analysis of at-home SpO_2 to automate OSA detection (Table 5). Some of these models also reached high diagnostic ability, yet only assessed either one of the two cohort types.

Five studies have focused on non-referral cohorts. All but one used the SHHS dataset, at least partially. In contrast, Chung et al. directly evaluated different thresholds of 4% ODI to determine OSA severity in a sample of 475 surgical patients [64]. Accuracies were high at the cost of unbalanced Se/Sp pairs, which differed for more than 20 percentage points in all cases. Schlotthauer et al. and Rolón et al. (2018 and 2020) used a subset of the SHHS2 database to respectively evaluate estimations of *ODI3* and AHI in the 15 e/h severity threshold [19–21]. Schlotthauer et al. used empirical mode decomposition as the main analytical tool, whereas Rolón et al. applied discrepancy measures (2018) and structured dictionary learning (2020). The three studies reported high Acc with balanced Se/Sp. Finally, Deviaene et al. used the whole SHHS dataset to develop a random forest model focused on detecting 3% desaturations caused by apneic events [14]. They subsequently estimated AHI by counting these events and applied a robust regression methodology to correct biases. Their results reached lower κ than our LSBoost model in SHHS1, mainly due

Table 5
State-of-the-art studies focused on analyzing SpO_2 recordings acquired at home or using SHHS database with other signals.

Study	# Subjects	Purpose and main predictor	AHI (e/h)	Se (%)	Sp (%)	Acc (%)	Four class κ
Non-referral							
Chung et al. (2012)	475	ODI4 direct assessment (univariate)	5	96.3	67.3	87.0	nd*
			15	70.0	92.5	84.0	
			30	76.0	97.2	93.7	
[64] Schlotthauer et al. (2014)	996	ODI3 estimation and regression using	15	83.8	85.5	nd	nd
[65] Airflow and SpO_2	988(SHHS1)	ad-hoc automatic algorithm to detect apneic events	15	94.7	88.5	91.0	nd
			30	87.8	98.2	96.7	

*nd: not enough data to estimate; *estimated from reported data.

In summary, the single LSBoost model proposed performed similarly to all other methods that exhibited the highest diagnostic ability among non-referral cohorts, while clearly outperforming all the proposed approaches focused on referral databases. Additionally, our model also outperformed the two studies that used the SHHS dataset with other signals [66,67], thus suggesting the superiority of SpO_2 when following a single-channel approach to simplify OSA diagnosis. Uddin et al. however, followed a two-channel approach by the jointly use of airflow and SpO_2 [68]. They involved 988 subjects from the SHHS1 dataset to develop and test a new ad-hoc detection algorithm. Their method, when evaluated in the 15 e/h AHI threshold, reached clearly higher performance than the results of our model in our SHHS1 test group. Moreover, very similar figures were reached when comparing the results from 5 e/h and 30 e/h AHI thresholds, with their proposal increasing the complexity of the test due to the extra airflow channel.

GRADO en INGENIERÍA BIOMÉDICA

Parte 2: Capítulo de Discusión

2.2 Contenidos y estructura

Propuesta de estructura (muy generalizable): con o sin subapartados (en TFG/TFM mejor con)

2. Discutir párrafo a párrafo, o sección a sección, todo lo que sea relevante para justificar ese párrafo anterior

4.4. Clinical usefulness of the proposal

Our OSA-specific model offers high diagnostic capability regardless of whether the strategy used focuses on primary care services (low pre-test probability) or specialized sleep facilities (high pre-test probability). In a context in which approximately 80% of moderate and severe patients remain undiagnosed [69], a reasonable purpose in a primary care setting is to conduct a protocol to screen as many hidden OSA positive subjects as possible. The high Se (>93%) and PPV (>87%) reached by our model in 5 e/h for both SHHS1t and SHHS2 groups suggest the suitability of our model for this task. Moreover, the decision curves showed that, when used to establish the non-referral subjects that should undergo PSG, our model produces higher net benefit than not conducting any protocol for almost any probability threshold a clinician would consider. This is an important result as sending no one to PSG is the current strategy for non-referral subjects in most healthcare systems [70]. It also showed that our model reaches higher net benefit than sending all the subjects to PSG

In this study, we have accomplished substantial advances in SpO_2 characterization by extracting up to 32 overnight features from different analytical approaches. They serve to develop a novel OSA-specific LSBoost-based model that has the ability to accurately estimate AHI from single-channel oximetry data obtained in the patients' home. Furthermore, this model displays a high level of agreement with the actual PSG-derived AHI, and high diagnostic performance in both referral and non-referral cohorts. Our analytical approaches not only allowed us to develop a clinically useful tool, but also explain the model through the SpO_2 data used by each of the 199 base classifiers of the ensemble, thus decreasing the common 'black box' perception of automatic models.

GRADO en INGENIERÍA BIOMÉDICA

Parte 2: Capítulo de Discusión

2.2 Contenidos y estructura

Propuesta de estructura (muy generalizable): con o sin subapartados (en TFG/TFM mejor con)

3. Discutir Limitaciones del trabajo y posibles líneas futuras

4.5. Limitations and future work

We have already discussed the need for future evaluation of the performance of our model in the presence of co-morbidities, and the need for further analysis of the extracted SpO_2 information to find associations with other common OSA events. Another limitation is the sample size of the referred database. Although it is one of the largest among the studies focused on SpO_2 recordings, it is remarkably smaller than the SHHS dataset. Therefore, future assessment of larger referral databases would help match the statistical power of our results on referral and non-referral cohorts. In addition, our validation strategy led to a smaller proportion of subjects for training than for testing. It was the result of trying to avoid biases in the model AHI estimation. However, using more training instances could derive into even more accurate models. There is also a need for assessing recordings from younger subjects as only 25% of those included in RHUH dataset are less than 46 years old and none from the SHHS dataset are less than 40 years of age. However, OSA is known to increase its prevalence with age, with only around 1.2% of subjects below 44 years old presenting AHI $\geq$ 5 e/h [72]. Similarly, although the SHHS dataset provides race data, it is only detailed for white and black subjects. Moreover, only white subjects were included in the RHUH dataset as a result of the natural demographics in the area of Valladolid (Spain). Therefore, a comprehensive assessment of our proposal involving subjects from other ethnicities is still pending. In this sense, despite significant differences in the proportion of black subjects included in our SHHS1 training set comparing with our SHHS1 test set, our model rightly classified a similar proportion of black and all other races subjects (65.5% *vs.* 70.3%), showing non-significant *p*-values ($>$ 0.01 in Fisher's exact test). The current use of the AHI thresholds to predict OSA-related adverse outcomes or mortality is also under discussion [73,74]. Although it is accepted as the main diagnostic option to establish OSA and its different severity categories [73], it is also as true that there is an active search for improving the AHI capability to predict the related risks [73,74]. In this regard, one strength of our regression model is that changes in AHI thresholds could be easily adopted in our AHI estimations. Future research, however, would be needed to assess to what extent our estimated AHI is sensitive to OSA-related negative consequences different from respiratory events. In addition, a comprehensive cost-effectiveness study would be also useful to complement our findings. Finally, the investigation of new automated diagnostic techniques is another future goal. Deep learning algorithms could be interesting alternatives at the cost of reduced model interpretability.

Parte 2: Capítulo de Discusión

2.3 Checklist del capítulo de Discusión

1. ¿Está clara la contribución de mi trabajo? ¿He subrayado la importancia de mis hallazgos? ¿He relacionado mis hallazgos y observaciones con otros estudios relevantes?
2. ¿He explicado de manera muy clara lo que considero importante, sin exagerar? ¿Me he asegurado de no atribuir interpretaciones a los resultados que en realidad no pueden ser respaldadas?
3. ¿He interpretado verdaderamente mis resultados, en lugar de simplemente repetirlos? ¿He mostrado la relación (confirmación o rechazo) entre mis resultados y mi hipótesis / objetivo original?
4. ¿He distinguido claramente hechos de especulaciones? ¿Podrá el lector entender fácilmente cuándo solo estoy sugiriendo una posible interpretación y no presentando evidencia concluyente?
5. ¿Me he asegurado de no ocultar ninguno de mis datos ni resultados inesperados?
6. ¿He discutido mis hallazgos en el contexto de lo que mencioné en la Introducción?
7. ¿Han sido justificadas y constructivas mis críticas a la literatura previa?
8. ¿Están todas las afirmaciones que he hecho respaldadas por los datos contenidos en mis figuras y tablas?
9. ¿He eliminado cualquier información trivial? ¿He sido lo más conciso posible?

Taller Práctico

GRADO en INGENIERÍA BIOMÉDICA

Ejercicios prácticos

Poner en práctica lo aprendido

📌 **1. Figuras en Flourish**

- ✓ https://app.flourish.studio: Box Plot / Violin Plot, Radar Chart, Correlation / Association Networks, Matriz de confusión

📌 **2. Redacta un texto para describir una de las figuras generadas con Flourish**

📌 **3. Ejemplo de estructura de discusión**

📌 **4. Kahoot: diferencias entre Resultados y Discusión**