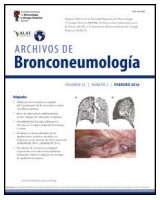
Archivos de Bronconeumología xxx (xxxx) xxx–xxx

Sociedad Española de Neumología y Cirugía Torácica SEPAR

ARCHIVOS DE
Bronconeumología

www.archbronconeumol.org



Discussion Letter

## Authors Reply to: "Reevaluating Feature Selection in Machine Learning Models for Identifying Disease-Modifying Agents in Obstructive Sleep Apnea"

*To the Director,*

We sincerely appreciate the critical analysis of our study,[1] and we welcome the opportunity to further clarify our findings.

The primary objective of our investigation was to apply machine learning (ML) techniques to capture the complex and nonlinear relationships between microRNAs (miRNAs) and obstructive sleep apnea (OSA). The biology of miRNAs is inherently intricate, involving regulatory networks with multiple targets, where even subtle variations trigger significant biological responses.[2] While classical statistical methods provide valuable insights, they may not fully account for these complexities.

This biological context guided our approach. We employed Boruta and VSURF, two algorithms based on the Random Forest method designed for variable selection. These methods aim to exclude variables with negligible importance in distinguishing study groups. Boruta evaluates the relevance of original attributes by comparing them against artificially generated, randomized shadow variables. If an original variable demonstrates significantly greater importance than its shadow counterparts, it is retained; otherwise, it is discarded. This iterative process continues until an optimal set of predictors is identified, mitigating the risk of including irrelevant features. Conversely, VSURF follows a three-phase approach to filter out non-contributory variables. In the first phase, it assesses the importance of each variable in differentiating the study groups, discarding those with low or near-zero relevance. The second phase refines the model by eliminating redundant or less stable variables. Finally, in the third phase, the subset of remaining variables that minimizes prediction error is selected. Additionally, we employed sPLS-DA (Sparse Partial Least Squares Discriminant Analysis), a dimensionality reduction technique designed to address high-dimensional data while preventing the inclusion of irrelevant variables. By introducing a penalty on the weights of variables, sPLS drives some weights to zero, a process known as sparse regularization, effectively selecting the most relevant variables within a relational component of miRNAs that correlate with the study group.

We acknowledge the recommendation to assess multicollinearity using the Variance Inflation Factor (VIF). While VIF is a standard tool in classical statistical modeling, the three ML methods implemented in our study inherently mitigate redundancy during variable selection. These methods evaluate all variables collectively, allowing for the identification of their relative importance within complex interactions. The relationship between miRNAs and OSA may not conform to simple monotonic associations, such as those detected through Spearman correlations or similar approaches. Regarding collinearity, both sPLS and VSURF address this concern through L1 regularization and the elimination of redundant variables, respectively.

In conclusion, we value the feedback received, as it contributes to a constructive dialog on the application of ML in molecular profiling, particularly in the field of noncoding RNAs.[3] While we employed algorithms specifically designed for variable selection and the analysis of complex interactions, our study is not without limitations. Despite efforts to improve generalizability by minimizing redundancy and irrelevant variables, the challenges of high dimensionality, the risk of overfitting and the need for external validation remain critical considerations. Further research is essential to expand our understanding of the role of miRNAs in the pathology of OSA.

### Authors' Contribution

All authors were involved in the drafting and critical revision of the work. All authors approved the final version of the manuscript.

### Artificial Intelligence Involvement

The authors declare no artificial intelligence involvement.

### Funding Statement

This study was funded by the Instituto de Salud Carlos III (ISCIII) through the projects "PI20/00577" and "PI22/00636" and cofunded by the European Union. DdGC has received financial support from Instituto de Salud Carlos III (Miguel Servet 2020: CP20/00041) cofunded by European Union. MSdlT has received financial support from the "Ramón y Cajal" grant (RYC2019-027831-I) from the "Ministerio de Ciencia e Innovación—Agencia Estatal de Investigación" cofunded by the European Social Fund (ESF)/"Investing in your future". CIBERES (CB07/06/2008) is an initiative of the Instituto de Salud Carlos III.

### Conflicts of Interests

The authors declare no competing interests.

### Acknowledgements

The authors would like to express their gratitude to Manel Perez-Pons and María C. García-Hidalgo for their invaluable technical support.

### References

1. Belmonte T, Benitez ID, García-Hidalgo MC, Molinero M, Pinilla L, Mínguez O, et al. Synergic integration of the miRNome, machine learning and bioinformatics

https://doi.org/10.1016/j.arbres.2025.03.010


Please cite this article as: I.D. Benitez, M. Sánchez-de-la-Torre and D. de Gonzalo-Calvo, Authors Reply to: "Reevaluating Feature Selection in Machine Learning Models for Identifying Disease-Modifying Agents in Obstructive Sleep Apnea", Archivos de Bronconeumología, https://doi.org/10.1016/j.arbres.2025.03.010

for the identification of potential disease-modifying agents in obstructive sleep apnea. Arch Bronconeumol. 2024;S0300-2896(24)00449-6.
2. Ebert MS, Sharp PA. Roles for MicroRNAs in conferring robustness to biological processes. Cell. 2012;149:515–24.
3. García-Hidalgo MC, Benítez ID, Perez-Pons M, Molinero M, Belmonte T, Rodríguez-Muñoz C, et al. MicroRNA-guided drug discovery for mitigating persistent pulmonary complications in critical COVID-19 survivors: a longitudinal pilot study. Br J Pharmacol. 2025;182:380–95.

Iván D. Benitez [a,b,c], Manuel Sánchez-de-la-Torre [b,c,d,e], David de Gonzalo-Calvo [a,b,*]

[a] *Translational Research in Respiratory Medicine, University Hospital Arnau de Vilanova and Santa Maria, IRBLleida, Lleida, Spain*
[b] *CIBER of Respiratory Diseases (CIBERES), Institute of Health Carlos III, Madrid, Spain*
[c] *Precision Medicine in Chronic Diseases, University Hospital Arnau de Vilanova and Santa Maria, IRB Lleida, Department of Nursing and Physiotherapy, Faculty of Nursing and Physiotherapy, University of Lleida, Lleida, Spain*
[d] *Group of Precision Medicine in Chronic Diseases, Hospital Nacional de Parapléjicos, IDISCAM, Spain*
[e] *Department of Nursing, Physiotherapy and Occupational Therapy, Faculty of Physiotherapy and Nursing, University of Castilla-La Mancha, Toledo, Spain*

*Corresponding author.
*E-mail address:* dgonzalo@irblleida.cat (D. de Gonzalo-Calvo).