

## Método de *random forest* para el reconocimiento de patrones de sensibilidad y resistencia en antibiogramas

### *Random forest method for the recognition of susceptibility and resistance patterns in antibiograms*

Estimado Editor:

La resistencia a los antimicrobianos es un tema de preocupación mundial y ha tenido un inesperado aumento en Chile en los últimos 10 años. En este país, los estándares de calidad y control de infecciones asociadas a la atención en salud han implementado un monitoreo continuo de patrones de sensibilidad y resistencia en bacterias alto interés clínico (alta virulencia y mortalidad). En los hospitales se vigila la resistencia antibiótica a través de métodos de estadística descriptivos. Se utilizan métodos de tendencia central, razones y proporciones que tienen limitaciones en la sensibilidad metodológica de detección de patrones de resistencia en el procesamiento de miles y millones de datos en un recinto asistencial o incluso a nivel regional y nacional. Tampoco tienen la capacidad de detectar desviaciones a lo largo del tiempo en análisis de series temporales. Los métodos analíticos ofrecen mayor desempeño para identificar variables dependientes y niveles de correspondencia (Pearson, Spearman Rank) al igual que los modelos de regresión lineal simple y múltiple. Sin embargo, son sensibles a la varianza de valores extremos y no tienen la capacidad de clasificar patrones en variables categóricas en miles de millones de datos<sup>1</sup>.

Las técnicas de *machine learning* han tenido un amplio uso desde la pandemia del COVID-19<sup>2</sup> y tienen un desempeño superior, en aplicaciones de resistencia antimicrobiana, que los estadísticos centrales y métodos analíticos. En microbiología, las técnicas de *clustering* son útiles para identificar patrones que a simple vista o con el uso de la estadística descriptiva o analítica sería difícil de lograr. El algoritmo de *random forest* (RF) es una metodología que genera múltiples árboles de decisión, y a partir de la combinación de ellos, entrega un resultado de alta precisión. Es un método de aprendizaje supervisado que en su fase de entrenamiento permite lidiar con problemas asociados a sobreajuste o subajuste de los modelos. En la etapa de entrenamiento, RF genera el mayor número posible de árboles de decisión para luego realizar una clasificación para el caso de estudio a evaluar: bacterias sensibles o resistentes a un antimicrobiano. El uso de “*bagging*” y “*feature randomness*” corresponde al que modelo disminuye el sesgo del modelo. En primera instancia, el clasificador RF genera

multitud de árboles de decisión, cada árbol de decisión es entrenado con una muestra de los datos de entrenamiento generada a partir del método de “*bagging*” y en cada nodo, el algoritmo busca en un subconjunto aleatorio de atributos (“*feature randomness*”) para determinar la división<sup>3</sup>. Con el RF creado, se puede aplicar en su versión de clasificación de variables categóricas para definir variables de importancia (VIM) y los resultados de clasificación. En múltiples centros asistenciales europeos se han aplicado para identificar bacterias emergentes, patrones de resistencia antibiótica, tendencias y predicciones de resistencia en uso de series temporales. Esto ha permitido alertar de forma temprana al equipo clínico y las autoridades. Por ejemplo, se ha utilizado la metodología de clasificación multi-etiqueta, basada en RF, como modelo predictivo de alta eficacia para *Escherichia coli* resistente a ciprofloxacina, ceftriaxona y CTZ (cloranfenicol-tetraciclina-zinc-eugenol)<sup>4</sup>. Un estudio multicéntrico, aplicando modelos de *machine learning*, con un registro de 8.342 infecciones de una unidad de urgencias y 15.806 infecciones del tracto urinario, entregaron una tasa de cobertura de antimicrobianos personalizados del 85,9% en el espectro de acción antibiótica. Esto fue similar al desempeño realizado por los médicos (84,3% ; valor-p = 0,11). La aplicación de modelos de *machine learning* (RF, XGBoost, regresión logística L1 y L2) basado en valores categóricos permitió simular una disminución del 69% de las prescripciones empíricas de vancomicina más piperacilina-tazobactam a solo el uso piperacilina-tazobactam<sup>5</sup>. Así, se podría controlar la prescripción innecesaria de antimicrobianos, limitando el aumento de mecanismos de aprendizaje y/o resistencia a estos fármacos.

Sugerimos que sería recomendable aplicar estos métodos de “*clustering*” y aprendizaje supervisado en los centros asistenciales chilenos para avanzar cualitativa y cuantitativamente en la tecnología de monitoreo del uso clínico de antimicrobianos en pacientes ambulatorios y hospitalizados. *Random forest* es un método complementario para otorgar confianza a la toma de decisiones del personal clínico.

## Referencias bibliográficas

1. Weaver C G W, Basmadjian R B, Williamson T, McBrien K, Sajobi T, Boyne D, et al. Reporting of model performance and statistical methods in studies that use machine learning to develop clinical prediction models: Protocol for a systematic review. *JMIR Res Protoc* 2022; 11(3): e30956. <https://doi.org/10.2196/30956>.
2. Alballa N, Al-Turaiki I. Machine learning approaches in COVID-19 diagnosis, mortality, and severity risk prediction: A review. *Inform Med Unlocked* 2021; 24: 100564. <https://doi.org/10.1016/j.imu.2021.100564>.
3. Chen C C M, Schwender H, Keith J, Nunkesser R,

Mengersen K, MacRossan P. Methods for identifying SNP interactions: A review on variations of logic regression, Random Forest and bayesian logistic regression. *IEEE/ACM Trans Comput Biol Bioinform* 2011; 8(6):1580-91. <https://doi.org/10.1109/TCBB.2011.46>.

4. Ren Y, Chakraborty T, Doijad S, Falgenhauer L, Falgenhauer J, Goesmann A, et al. Multi-label classification for multi-drug resistance prediction of *Escherichia coli*. *Comput Struct Biotechnol J* 2022; 20: 1264-70. <https://doi.org/10.1016/j.csbj.2022.03.007>.
5. Corbin C K, Sung L, Chattopadhyay A, Noshad M, Chang A, Deresinski S, et al. Personalized antibiograms for machine learning driven antibiotic selection. *Communications Medicine* 2022; 2: 38. <https://doi.org/10.1038/s43856-022-00094-8>.

**Nicolás Ayala-Aldana<sup>1</sup> y Leticia Gonzalez-Valdés<sup>2</sup>**  
<sup>1</sup>BSc. MBA. Predoctoral Researcher. Departamento de Genética, Microbiología y Bioestadística-Universidad de Barcelona. Barcelona, España.  
<sup>2</sup>Tecnóloga Médica, Chile.

Conflicto de interés: la autora Leticia González-Valdés tiene afiliación en Chile en la Compañía Becton-Dickinson donde se desempeña como especialista de aplicaciones.

**Correspondencia a:**  
 Nicolas Ayala-Aldana  
 nayalaaldana@gmail.com