

Resumen

Klebsiella pneumoniae es un patógeno oportunista asociado al ambiente intrahospitalario y que provoca infecciones graves en pacientes en riesgo. El tratamiento antibiótico puede incluir carbapenémicos en episodios infecciosos causados por *K. pneumoniae* resistente a otros antibióticos. No obstante, *K. pneumoniae* ha adquirido resistencia a múltiples antibióticos, convirtiéndose en un problema de salud pública. El diagnóstico tardío y la prescripción no específica han aumentado las tasas de mortalidad. Las herramientas de Inteligencia Artificial (AI) se han convertido en un apoyo a las técnicas tradicionales para el diagnóstico y prescripción de un tratamiento. En este proyecto se implementó un modelo basado en *Machine Learning* (ML) para predecir la resistencia de *K. pneumoniae* a imipenem y meropenem. La extensa cantidad de datos de secuenciación de nueva generación (NGS) permitió formar un set de datos para el entrenamiento del modelo de ML. El preprocesamiento basado en estudios de asociación de genoma completo (GWAS) corrigió los problemas asociados a la estructura poblacional bacteriana y la dimensionalidad. Los modelos se entrenaron mediante un algoritmo *Extreme Gradient Boosting* (XGBoost) optimizando tiempo y recursos computacionales. Los resultados de aprendizaje demostraron la capacidad de los modelos basados en ML para predecir fenotipos de susceptibilidad antimicrobiana. No obstante, las métricas de evaluación se podrían mejorar aumentando la cantidad de aislados.

Palabras clave:

- **KLEBSIELLA PNEUMONIAE**
- **CARBAPENÉMICOS**
- **MACHINE LEARNING (ML)**
- **EXTREME GRADIENT BOOSTING (XGBOOST)**

Abstract

Klebsiella pneumoniae is an opportunistic pathogen associated with the hospital environment and able to cause serious infections in vulnerable patient groups. The treatment, in infections caused by resistant isolates includes the administration of carbapenems. However, *K. pneumoniae* has acquired resistance to multiple antibiotics, becoming a public health problem. Untimely diagnosis and non-targeted prescription have increased mortality rates. Artificial intelligence (AI) tools have become a support for traditional techniques for diagnosing and prescribing a treatment. In this project, a model based on Machine Learning (ML) was implemented to predict the resistance of *K. pneumoniae* to imipenem and meropenem. The large amount of New Generation Sequencing (NGS) data made it possible to create a data set for training the ML model. The preprocessing based in genome-wide association studies (GWAS) corrected problems associated with bacterial population structure and dimensionality. The models were trained using an Extreme Gradient Boosting (XGBoost) algorithm, optimizing time and computational resources. The learning outcomes demonstrated the ability of ML-based models to predict antimicrobial susceptibility phenotypes. However, the evaluation metrics could be improved by increasing the number of isolates.

Key words:

- **KLEBSIELLA PNEUMONIAE**
- **CARBAPENÉMICOS**
- **MACHINE LEARNING (ML)**
- **EXTREME GRADIENT BOOSTING (XGBOOST)**