

## **Generación de modelos predictivos de péptidos anti-angiogénicos para el proyecto CICLOGEN**

Carlos Fernández-Lozano<sup>1</sup> y Jose Liñares Blanco<sup>1</sup>

<sup>1</sup>Department of Computer Science and Information Technologies, Faculty of Computer Science, University of A Coruña, CITIC. Campus Elviña s/n, 15071, A Coruña, Spain

El cribado y el modelado in silico son actividades críticas para la reducción de los costes experimentales. También aceleran notablemente la investigación y refuerzan el marco teórico, permitiendo así a los investigadores cuantificar numéricamente la importancia de un determinado subconjunto de información. Por ejemplo, en campos como el cáncer y otras enfermedades de alta prevalencia, es crucial disponer de un método de predicción fiable. El objetivo de este trabajo es clasificar las secuencias peptídicas en base a su actividad anti-angiogénica para entender los principios subyacentes a través del aprendizaje automático. En primer lugar, las secuencias peptídicas fueron convertidas en tres tipos de descriptores moleculares numéricos basados en la composición de aminoácidos. Se realizaron diferentes experimentos con los descriptores y se fusionaron para obtener resultados de línea de base para el desempeño de los modelos, particularmente de cada subconjunto de descriptores moleculares. Se aplicó un proceso de selección de características para reducir la dimensionalidad del problema y eliminar las características ruidosas, que están muy presentes en los problemas biológicos. Después de un robusto diseño experimental de aprendizaje de máquina en igualdad de condiciones, se superó estadística y significativamente el mejor modelo antiangiogénico previamente publicado con un modelo lineal generalizado (glmnet), alcanzando un valor medio de AUC superior a 0,96 y con una precisión de 0,86 con 200 descriptores moleculares, mezclados de los tres grupos. De acuerdo con nuestros hallazgos, los péptidos de actividad antiangiogénica están fuertemente asociados con las secuencias de aminoácidos SP, LSL, PF, DIT, PC, GH, RQ, QD, TC, SC, AS, CLD, ST, MF, GRE, IQ, CQ y HG.