ML model used

We included a four ML methods in our study and are briefly described below. These ML methods were used in the regression mode.

Random forest: Random forest method is a non-parametric, supervised method, that can be used as both classification and regression. The heart of tree-based learners is the decision tree, wherein a series of decision rules are chained and learned. In a decision tree, every decision rule occurs at a decision node [49]. This model proposed by Tin Kam Ho and further adapted by Leo Breiman and Adele Cutler [50].

Support Vector Machine: Support Vector machines classify data by finding the hyperplane that maximizes the margin between the classes in the training data. A support vector machine can be represented like: $f(x) = \beta_0 + \sum_{i \in S} \alpha_i K(x_i, x_{\hat{i}})$, where β_0 is the bias, S is the set of all support vector observations, α is the parameters in the model to be learned, $(x_i, x_{\hat{i}})$ are pairs of two support vector operations and K is the kernel function which compares the similarity between $x_i, x_{\hat{i}}$.

k Nearest Neighbors: The k-nearest Neighbors algorithm is a supervised machine learning algorithm that can be used as both classification and regression problems, especially when there is little or no prior knowledge about the distribution of the data. Let X_i be an input sample with p features $(x_{i1}, x_{i2}, \dots x_{ip})$, The Euclidean distance between the sample x_i and $x_l (l = 1,2, \dots, n)$ is defined as $d(X_1, X_l) = \sqrt{(x_{i1} - x_{l1})^2 + \cdots + (x_{ip} - x_{lp})^2}$, and its neighborhood as: $R_i = \{X \in \mathbb{R}^p : d(X, X_i) \leq: d(X, X_m), \forall\, i \neq m\}$, where R_i represents the clusters of elements with class m, and X the set of points belong to it. The predicted class of the new sample x is set equal to the most frequent class among the k nearest training samples, which follow the rule: $d(m_i, X) = \min\{d(m_i, X)\}$, where d is the distance function.

Multi-Layer perceptron (MLP): A MLP is composed of multiple perceptrons or neurons, developed originally by Frank Rosenblatt [51], commonly arranged in three layers known as input layer, hidden layer (can have more than one stack of neurons) and output layer, and this kind of configuration is called Artificial Neural Network (ANN). Each input of the neurons x_i are associated with a weight w_i and

computed as a sum as follows $z = x_1 w_1 + \cdots + x_n w_n = X^T W$, then an activation function is calculated as $f(z)$, where $f(z)$ can be any continuously differentiable function like a linear function, sigmoid or even the modern ReLU commonly used in deep learning [52].