



Figure 2. Performance evaluation of ForSyn

(A–E) According to Nemenyi test, the average rank of all algorithms tested on data 1–8 and five metrics: (A) F1 score, (B) AUPR, (C) recall, (D) MCC, and (E) G-mean. The average rank of each algorithm in eight datasets is marked as a red dot, and a horizontal line crossing the red dot indicates the range of CD value in Nemenyi test. The smaller the overlap between two horizontal bars, the more significant the difference between the two algorithms.

(F) The performance difference between ForSyn and other algorithms on F1 score under data 1–8. The y axis denotes $\Delta F1$ between ForSyn and other comparison algorithms, $\Delta F1 = F1_{\text{comparison algorithms}} - F1_{\text{ForSyn}}$. A positive number indicates that the performance value of the comparison algorithm exceeds ForSyn, while a negative number indicates that ForSyn is superior to the comparison algorithm.

(G) The impact of the number of base classifiers and the under-sampling ratio on performance of ForSyn's RF-CSU unit. The y axis represents the F1 score, and the x axis represents the under-sampling ratio for the majority class with a value range of 0.1–1. The blue, red, and green lines represent the RF-CUS unit containing 100, 300, and 500 decision trees, respectively.

(H) The impact of the number of base classifiers and the retained feature dimension on performance of ForSyn's ETF-DR unit.