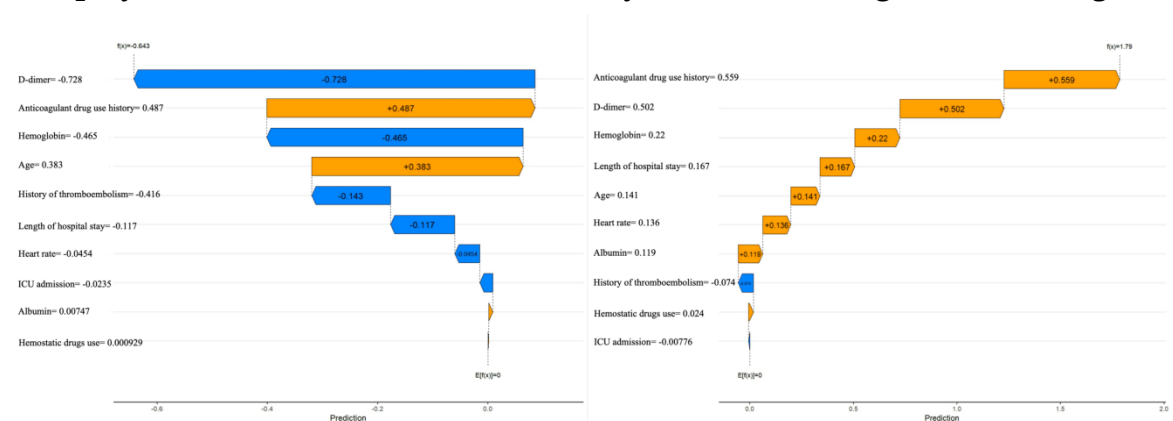**Supplementary Figure 1 The top 20 variables based on variable importance and SHapley Additive exPlanations value analysis from the categorical boosting model.**

Supplementary Figure 2 The arrows represent the impact of each factor on the prediction, with blue and red arrows indicating decreased (blue) or increased (orange) thromboembolic risk, respectively. In the negative patient, the SHapley Additive exPlanations score (-0.643) fell below the baseline (0), whereas in the positive patient, the SHapley Additive exPlanations score (1.79) was above baseline (0)

Supplementary Table 1 The performance of five machine learning models and univariate D-dimer using all variables in the internal validation sets

Model	Accuracy	Precision	Sensitivity	Specificity	F1	Area under the receiver operating characteristic curve	P value
L1 regularized logistic	0.743 (0.705-0.778)	0.444 (0.371-0.511)	0.724 (0.637-0.797)	0.748 (0.705-0.787)	0.547	0.778 (0.73-0.826)	0.0000007

regression

Support vector machines	0.720 (0.666-0.739)	0.318 (0.310-0.451)	0.612 (0.521-0.696)	0.738 (0.677-0.787)	0.469	0.725 (0.678-0.772)	0.002913
Categorical boosting	0.795 (0.769-0.821)	0.517 (0.436-0.598)	0.682 (0.574-0.787)	0.838 (0.803-0.872)	0.586	0.823 (0.784-0.863)[1]	0.0000000
Random Forest	0.771 (0.734-0.804)	0.476 (0.403-0.551)	0.690 (0.590-0.787)	0.794 (0.740-0.848)	0.566	0.796 (0.748-0.844)	0.0000000
Extreme gradient boosting	0.736 (0.697-0.771)	0.428 (0.359-0.499)	0.669 (0.561-0.761)	0.740 (0.686-0.787)	0.524	0.772 (0.723-0.821)	0.0000000
D-Dimer	0.621 (0.579-0.668)	0.309 (0.253-0.371)	0.621 (0.534-0.704)	0.621 (0.574-0.671)	0.413	0.621 (0.571-0.671)	0.0000000

[1]Categorical boosting algorithm achieved the highest area under the receiver operating characteristic curve value among all machine learning models.