Machine learning models were developed using logistic regression (LR), random forest, and support vector machine classifiers. Five models were developed: (1) a model that predicts the probability of good outcome from data with successful recanalization (SR) (i.e., modified treatment in cerebral infarction [mTICI] 2b–3); (2) a model that predicts the probability of a good outcome from data with unsuccessful recanalization (i.e., mTICI 0–2a); (3) a model that predicts the probability of SR from endovascular thrombectomy (EVT) treatment; (4) a model that predicts the probability of a good outcome from data with EVT treatment; and (5) a model that predicts the probability of good outcome from data without EVT treatment. Fivefold cross-validation was performed using a randomized search for hyperparameter optimization of each classifier in Scikit-learn.

For external validation, we used baseline DWI images from 54 patients with EVT and 18 patients without EVT, who were from either University of California Los Angeles (UCLA) or Samsung Medical Center (SMC). The perfusion-weighted imaging (PWI) and DWI lesion volumes were estimated using RAPID software for the UCLA data and a custom software tool for the SMC data. The mismatch ratio was calculated as the ratio of the PWI lesion volume to the DWI lesion volume. All 42 features were used as inputs for ML predictions. For each ML method, we obtained 10 different train/validation splits using 10 different random seeds and performed 10 five-fold cross-validations to 10 different models, which were tested on an external validation cohort to calculate 10 area under the curve (AUC) values.

Statistical analysis
Statistical analysis was conducted using R version 3.5.0 (R Foundation for Statistical Computing, Vienna, Austria). Descriptive demographics and clinical and radiological data are shown as mean±standard deviations or numbers or median and interquartile range. A two-sample unpaired Student’s t-test was performed to determine if the development and validation data were significantly different in terms of age, National Institutes of Health Stroke Scale score, lesion volume, and onset-to-magnetic resonance imaging time. For the external validation data, we calculated the receiver operating characteristic (ROC) AUC values for the comparison of the prediction of good outcome. The Mann–Whitney U test was used to test the statistical significance of AUC differences between unconditional and conditional ML models. Statistical significance was set at P<0.05.

An interaction term of the ML-estimated gain of thrombectomy (GoT) and EVT (binary variable) was used for a LR model with an mRS score ≤2 at 90 days as the outcome binary vari-
able. The GoT×treatment interaction was considered significant if the $P$-value was <0.05. We divided the patients treated with EVT into two groups (high and low GoT groups) based on the threshold in the ML-estimated GoT. The cutoff point was obtained by calculating the Youden index from the ROC curve, where the binary functional outcome was the dependent variable. The same dichotomization scheme was used for the patients treated without EVT.