Accuracy vs runtime (per dataset × method)