**Supplementary Data 2.**

**Performance of the machine learning and traditional models to predict in hospital mortality, as well as good functional outcome (mRS 0-2) and mortality at 3 months in the Geneva Stroke (internal validation) and MIMIC-III datasets (external validation).**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Model** | **ROC AUC** | **Matthew's Coefficient** | **Accuracy** | **Precision (PPV)** | **Recall (Sensitivity)** | **Specificity** |
| GENEVA STROKE DATASET | **Mortality (in hospital)** | | | | | | |
| Transformer | 0.911 (0.843-0.951) | 0.486 (0.318-0.629) | 0.935 (0.914-0.957) | 0.425 (0.278-0.576) | 0.636 (0.440-0.794) | 0.920 (0.469-0.967) |
| XGB | 0.908 (0.857-0.951) | 0.295 (0.096-0.494) | 0.951 (0.933-0.969) | 0.667 (0.250-1.000) | 0.143 (0.037-0.296) | 0.692 (0.043-1.000) |
| LSTM | 0.884 (0.808-0.943) | 0.358 (0.169-0.529) | 0.947 (0.927-0.965) | 0.500 (0.267-0.750) | 0.292 (0.136-0.474) | 0.786 (0.154-0.992) |
| THRIVE-C | 0.840 (0.815-0.866) | 0.254 (0.215-0.291) | 0.825 (0.813-0.837) | 0.163 (0.137-0.186) | 0.625 (0.565-0.689) | 0.768 (0.571-0.845) |
| THRIVE | 0.746 (0.710-0.783) | 0.134 (0.108-0.158) | 0.485 (0.470-0.501) | 0.074 (0.063-0.085) | 0.842 (0.793-0.896) | 0.624 (0.455-0.886) |
| HIAT | 0.731 (0.697-0.765) | 0.118 (0.096-0.138) | 0.402 (0.386-0.416) | 0.065 (0.056-0.074) | 0.892 (0.848-0.935) | 0.618 (0.364-0.928) |
| SPAN-100 | 0.675 (0.639-0.710) | 0.261 (0.206-0.312) | 0.901 (0.892-0.909) | 0.225 (0.182-0.265) | 0.425 (0.356-0.494) | 0.720 (0.367-0.932) |
| Custom LR | 0.750 (0.630-0.846) | 0.000 (0.000-0.000) | 0.949 (0.927-0.965) | 0.000 (0.000-0.000) | 0.000 (0.000-0.000) | 0.500 (0.000-1.000) |
| **Mortality (3 months)** | | | | | | |
| Transformer | 0.893 (0.839-0.933) | 0.565 (0.437-0.676) | 0.907 (0.880-0.934) | 0.773 (0.638-0.902) | 0.485 (0.358-0.606) | 0.823 (0.373-0.989) |
| XGB | 0.862 (0.801-0.907) | 0.521 (0.396-0.635) | 0.900 (0.871-0.927) | 0.771 (0.613-0.900) | 0.421 (0.299-0.545) | 0.793 (0.317-0.990) |
| LSTM | 0.890 (0.839-0.930) | 0.577 (0.466-0.684) | 0.898 (0.866-0.923) | 0.633 (0.520-0.745) | 0.643 (0.522-0.759) | 0.864 (0.544-0.958) |
| THRIVE-C | 0.787 (0.765-0.807) | 0.331 (0.291-0.370) | 0.820 (0.808-0.832) | 0.375 (0.340-0.411) | 0.504 (0.460-0.547) | 0.711 (0.467-0.878) |
| THRIVE | 0.723 (0.700-0.745) | 0.226 (0.199-0.251) | 0.537 (0.522-0.552) | 0.204 (0.188-0.221) | 0.837 (0.807-0.868) | 0.651 (0.476-0.864) |
| HIAT | 0.700 (0.679-0.723) | 0.201 (0.178-0.225) | 0.456 (0.439-0.472) | 0.183 (0.168-0.198) | 0.894 (0.870-0.924) | 0.634 (0.374-0.920) |
| SPAN-100 | 0.637 (0.617-0.656) | 0.324 (0.278-0.368) | 0.861 (0.851-0.872) | 0.483 (0.429-0.539) | 0.330 (0.290-0.367) | 0.665 (0.297-0.952) |
| Custom LR | 0.805 (0.747-0.860) | 0.259 (0.117-0.384) | 0.862 (0.828-0.891) | 0.526 (0.308-0.733) | 0.193 (0.097-0.296) | 0.686 (0.109-0.984) |
| **Good functional outcome (mRS 0-2; 3 months)** | | | | | | |
| Transformer | 0.894 (0.863-0.922) | 0.638 (0.567-0.711) | 0.829 (0.796-0.863) | 0.848 (0.805-0.887) | 0.877 (0.837-0.911) | 0.832 (0.703-0.906) |
| XGB | 0.895 (0.864-0.924) | 0.640 (0.560-0.706) | 0.829 (0.789-0.863) | 0.844 (0.797-0.885) | 0.883 (0.841-0.917) | 0.826 (0.691-0.912) |
| LSTM | 0.889 (0.857-0.918) | 0.607 (0.532-0.683) | 0.812 (0.774-0.847) | 0.856 (0.812-0.899) | 0.828 (0.782-0.871) | 0.811 (0.732-0.866) |
| THRIVE-C | 0.811 (0.797-0.824) | 0.383 (0.354-0.411) | 0.711 (0.697-0.727) | 0.694 (0.678-0.712) | 0.937 (0.927-0.947) | 0.663 (0.346-0.945) |
| THRIVE | 0.717 (0.703-0.731) | 0.368 (0.340-0.393) | 0.667 (0.653-0.681) | 0.805 (0.787-0.824) | 0.594 (0.574-0.613) | 0.687 (0.577-0.796) |
| HIAT | 0.704 (0.688-0.718) | 0.349 (0.321-0.377) | 0.631 (0.616-0.647) | 0.836 (0.816-0.857) | 0.486 (0.468-0.507) | 0.672 (0.470-0.870) |
| SPAN-100 | 0.594 (0.584-0.605) | 0.317 (0.291-0.342) | 0.676 (0.661-0.691) | 0.656 (0.640-0.671) | 0.981 (0.976-0.987) | 0.605 (0.190-0.986) |
| Custom LR | 0.799 (0.748-0.840) | 0.519 (0.436-0.599) | 0.774 (0.734-0.812) | 0.795 (0.744-0.840) | 0.843 (0.796-0.886) | 0.780 (0.611-0.879) |
| MIMIC-III | **Mortality (in hospital)** | | | | | | |
| Transformer | 0.876 (0.791-0.932) | 0.293 (0.168-0.382) | 0.636 (0.579-0.692) | 0.194 (0.121-0.267) | 0.880 (0.714-1.000) | 0.684 (0.559-0.966) |
| XGB | 0.864 (0.807-0.915) | 0.113 (-0.033-0.323) | 0.903 (0.870-0.939) | 0.500 (0.000-1.000) | 0.038 (0.000-0.138) | 0.594 (0.000-1.000) |
| LSTM | 0.836 (0.765-0.902) | 0.340 (0.179-0.501) | 0.834 (0.789-0.879) | 0.308 (0.182-0.459) | 0.591 (0.375-0.767) | 0.806 (0.409-0.898) |
| THRIVE-C | 0.728 (0.658-0.805) | 0.238 (0.143-0.322) | 0.650 (0.608-0.695) | 0.185 (0.127-0.239) | 0.750 (0.615-0.909) | 0.669 (0.599-0.875) |
| THRIVE | 0.676 (0.617-0.741) | 0.192 (0.142-0.246) | 0.405 (0.360-0.450) | 0.136 (0.097-0.179) | 0.941 (0.917-1.000) | 0.653 (0.308-1.000) |
| HIAT | 0.656 (0.586-0.728) | 0.177 (0.105-0.252) | 0.465 (0.416-0.513) | 0.140 (0.096-0.183) | 0.867 (0.769-1.000) | 0.605 (0.379-1.000) |
| SPAN-100 | 0.587 (0.503-0.650) | 0.174 (0.006-0.299) | 0.854 (0.821-0.887) | 0.250 (0.083-0.400) | 0.250 (0.083-0.385) | 0.686 (0.125-0.943) |
| Custom LR | 0.622 (0.511-0.724) | 0.000 (0.000-0.000) | 0.903 (0.866-0.939) | 0.000 (0.000-0.000) | 0.000 (0.000-0.000) | 0.500 (0.000-1.000) |
| **Mortality (3 months)** | | | | | | |
| Transformer | 0.875 (0.820-0.919) | 0.448 (0.338-0.553) | 0.733 (0.676-0.781) | 0.426 (0.333-0.515) | 0.848 (0.733-0.936) | 0.750 (0.651-0.926) |
| XGB | 0.872 (0.819-0.913) | 0.443 (0.305-0.561) | 0.785 (0.733-0.834) | 0.486 (0.371-0.600) | 0.686 (0.558-0.808) | 0.773 (0.574-0.856) |
| LSTM | 0.869 (0.816-0.910) | 0.425 (0.321-0.515) | 0.688 (0.631-0.745) | 0.390 (0.298-0.474) | 0.887 (0.780-0.962) | 0.729 (0.580-0.954) |
| THRIVE-C | 0.785 (0.739-0.835) | 0.323 (0.222-0.412) | 0.688 (0.644-0.732) | 0.369 (0.293-0.443) | 0.706 (0.611-0.800) | 0.692 (0.625-0.784) |
| THRIVE | 0.713 (0.661-0.760) | 0.280 (0.220-0.343) | 0.495 (0.455-0.541) | 0.282 (0.231-0.330) | 0.939 (0.897-1.000) | 0.660 (0.339-1.000) |
| HIAT | 0.698 (0.644-0.751) | 0.265 (0.197-0.343) | 0.542 (0.497-0.589) | 0.293 (0.239-0.350) | 0.857 (0.800-0.938) | 0.662 (0.414-0.926) |
| SPAN-100 | 0.601 (0.551-0.651) | 0.274 (0.144-0.396) | 0.801 (0.762-0.842) | 0.533 (0.384-0.733) | 0.258 (0.167-0.348) | 0.661 (0.179-0.967) |
| Custom LR | 0.777 (0.705-0.839) | 0.344 (0.219-0.466) | 0.737 (0.680-0.789) | 0.410 (0.304-0.520) | 0.630 (0.500-0.755) | 0.721 (0.519-0.814) |

*Legend*: Metrics are reported as median (95% confidence interval). LSTM: long short-term memory; XGB: XGBoost; ROC AUC: area under the receiver operating characteristic curve; PPV: positive predictive value; THRIVE: Totaled Health Risks in Vascular Events Score; HIAT: Houston Intra-Arterial Therapy score; Span100: Stroke Prognostication using Age and NIH Stroke Scale index; Custom LR: logistic regression trained on training dataset and using three clinical variables.