Supplementary Table 4: Evaluation of Model and Clinical Concordance

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Patient ID** | **Days Since Onset** | **FVC Baseline** | **Mean FVC** | **FVC Change(% change /mth)** | **FVC SD** | **ALSFRS-R Baseline** | **Mean ALSFRS-R** | **ALSFRS-R/ Decrease** | **ALSFRS-R/SD** | **Future Slope (change in pts/mth)** | **Slope Category** | **BLSTM Model Prediction** | **BLSTM Confidence (Percentile, 0 - least confident, 1 - most)** | **XGB Model Prediction** | **Clinician 1 Assessment** | **Clinician 2 Assessment** | **Clinician 3 Assessment** |
| **Mean of All Patients** | 813 | 76.1 | 69.3 | -2.0 | 0.066 | 35.2 | 34.6 | -0.79 | 1.57 | -0.819 |  |  |  |  |  |  |  |
| **Mean of Fast Patients** | 652 | 70.4 | 66.2 | -1.9 | 0.071 | 34.9 | 33.9 | -1.13 | 1.97 | -2.33 | Fast |  |  |  |  |  |  |
| **Mean of Non-Fast Patients** | 866 | 81.8 | 72.4 | -2.1 | 0.060 | 35.7 | 34.9 | -0.65 | 1.40 | -0.43 | Non-Fast |  |  |  |  |  |  |
| **Group 1: Correctly Predicted by Models and Clinicians** | | | | | | | | | | | | | | | | |  |
| **23959** | 288 | 83.2 | 68.6 | -8.2 | 0.0992 | 37.2 | 39.6 | 0.00 | 4.16 | -3.47 | Fast | Fast | 0.18 | Fast | Fast | Fast | Fast |
| **65548** | 323 | 60.1 | 45.4 | -5.74 | 0.108 | 39.6 | 38.4 | -0.58 | 1.20 | -2.46 | Fast | Fast | 0.49 | Fast | Fast | Fast | Fast |
| **12856** | 2248 | 68.8 | 64.6 | -2.8 | 0.0096 | 34.8 | 36 | 0.61 | 1.20 | -0.77 | Non-Fast | Non-Fast | 1.00 | Non-Fast | Non-Fast | Non-Fast | Non-Fast |
| **20417** | 712 | 72.7 | 73.4 | 3.28 | 0.120 | 37.2 | 38 | 0.62 | 0.69 | -0.58 | Non-Fast | Non-Fast | 0.17 | Non-Fast | Non-Fast | Non-Fast | Non-Fast |
| **Group 2: Correctly Predicted by Models , Mixed Accuracy in Clinicians** | | | | | | | | | | | | | | | | |  |
| **348** | 501 | NA | 59.2 | -4.4 | 0.091 | 39 | 38.5 | -0.49 | 0.71 | -0.28 | Non-Fast | Non-Fast | 0.54 | Non-Fast | Non-Fast | Non-Fast | ***Fast*** |
| **11081** | 984 | 56.0 | 52.2 | -6.3 | 0.028 | 40.8 | 39 | -1.20 | 1.55 | -1.51 | Fast | Fast | 0.01 | Fast | Fast | Fast | ***Non-Fast*** |
| **Group 3: Incorrectly Predicted By Models and Clinicians** | | | | | | | | | | | | | | | | |  |
| **100626** | 887 | 87.6 | 88.4 | 0.76 | 0.015 | 36 | 34.5 | -0.09 | 1.51 | -2.41 | Fast | ***Non-Fast*** | 0.70 | ***Non-Fast*** | ***Non-Fast*** | ***Non-Fast*** | ***Non-Fast*** |
| **202517** | 1017 | 91.8 | 94.1 | 0.75 | 0.043 | 42 | 42 | -1.17 | 1.96 | -1.94 | Fast | ***Non-Fast*** | 0.83 | ***Non-Fast*** | ***Non-Fast*** | ***Non-Fast*** | ***Non-Fast*** |
| **7399** | 382 | 92.8 | 84.9 | -5.3 | 0.083 | 42 | 38.7 | -2.19 | 2.66 | 0.01 | Non-Fast | ***Fast*** | 0.46 | ***Fast*** | ***Fast*** | ***Fast*** | ***Fast*** |
| **Group 4: Incorrectly Predicted By Models, Mixed Accuracy In Clinicians** | | | | | | | | | | | | | | | | |  |
| **104310** | 1361 | 54.7 | 39.4 | -7.75 | 0.159 | 26.4 | 24.4 | -1.55 | 1.83 | -3.55 | Fast | ***Non-Fast*** | 0.69 | ***Non-Fast*** | Fast | Fast | ***Non-Fast*** |
| **43715** | 611 | NA | 51.1 | -1.68 | 0.059 | 40 | 39 | -1.15 | 1.41 | 0.11 | Non-Fast | ***Fast*** | 0.29 | ***Fast*** | Non-Fast | Non-Fast | ***Fast*** |
| **Group 5: Incorrectly Predicted By Models but Correctly By Clinicians** | | | | | | | | | | | | | | | | |  |
| **216936** | 522 | 89.1 | 87.4 | 0.15 | 0.021 | 31.2 | 32.16 | 0.24 | 1.31 | -0.64 | Non-Fast | ***Fast*** | 0.53 | ***Fast*** | Non-Fast | Non-Fast | Non-Fast |
| **Group 6: Mixed Accuracy in Both Models and Clinicians** | | | | | | | | | | | | | | | | |  |
| **88338** | 515 | 55.9 | 50.8 | 6.5 | 0.141 | 42 | 41.2 | -0.12 | 1.39 | -3.60 | Fast | Fast | 0.23 | ***Non-Fast*** | Fast | ***Non-Fast*** | Fast |
| **31916** | 592 | 74.2 | 83.9 | 1.72 | 0.067 | 30 | 31.6 | 1.03 | 1.39 | -1.61 | Fast | ***Non-Fast*** | 0.52 | Fast | ***Non-Fast*** | Fast | ***Non-Fast*** |
| **190553** | 436 | NA | 65.5 | -0.796 | 0.042 | 43 | 44 | 0.12 | 1.00 | -1.57 | Fast | Fast | 0.29 | ***Non-Fast*** | Fast | ***Non-Fast*** | Fast |
| **119554** | 429 | NA | 73.6 | 0.390 | 0.0047 | 41 | 41.25 | 0.12 | 0.50 | -1.78 | Fast | ***Non-Fast*** | 0.16 | Fast | ***Non-Fast*** | ***Non-Fast*** | Fast |
| **8471** | 500 | 60.0 | 40.9 | -6.3 | 0.111 | 25.2 | 24.6 | -1.29 | 0.85 | -0.25 | Non-Fast | ***Non-Fast*** | 0.52 | Fast | Fast | Fast | ***Non-Fast*** |
| **71211** | 811 | 82.8 | 78.4 | -0.85 | 0.024 | 34.8 | 34.8 | 0.00 | 0.00 | -0.80 | Non-Fast | ***Fast*** | 0.01 | Non-Fast | Non-Fast | ***Fast*** | Non-Fast |
| **Group 7: Mixed Accuracy in Models, Incorrect in Clinicians** | | | | | | | | | | | | | | | | |  |
| **5936** | 470 | 100.9 | 98.8 | -0.7 | 0.037 | 45.6 | 44 | -1.13 | 1.39 | -0.72 | Non-Fast | ***Fast*** | 0.65 | Non-Fast | ***Fast*** | ***Fast*** | ***Fast*** |
| **18216** | 346 | 87.3 | 85.2 | -2.2 | 0.049 | 34.8 | 32 | -2.08 | 2.50 | -1.34 | Non-Fast | Non-Fast | 0.53 | ***Fast*** | ***Fast*** | ***Fast*** | ***Fast*** |
| **Classification Accuracy for Subset of 20 Samples** | | | | | | | | | | | | **0.5** |  | **0.5** | **0.65** | **0.55** | **0.45** |

Table categorised by accuracy and concordance of machine learning and clinician predictions. Incorrect predictions are italicised, bolded and highlighted.