

Supplementary Information

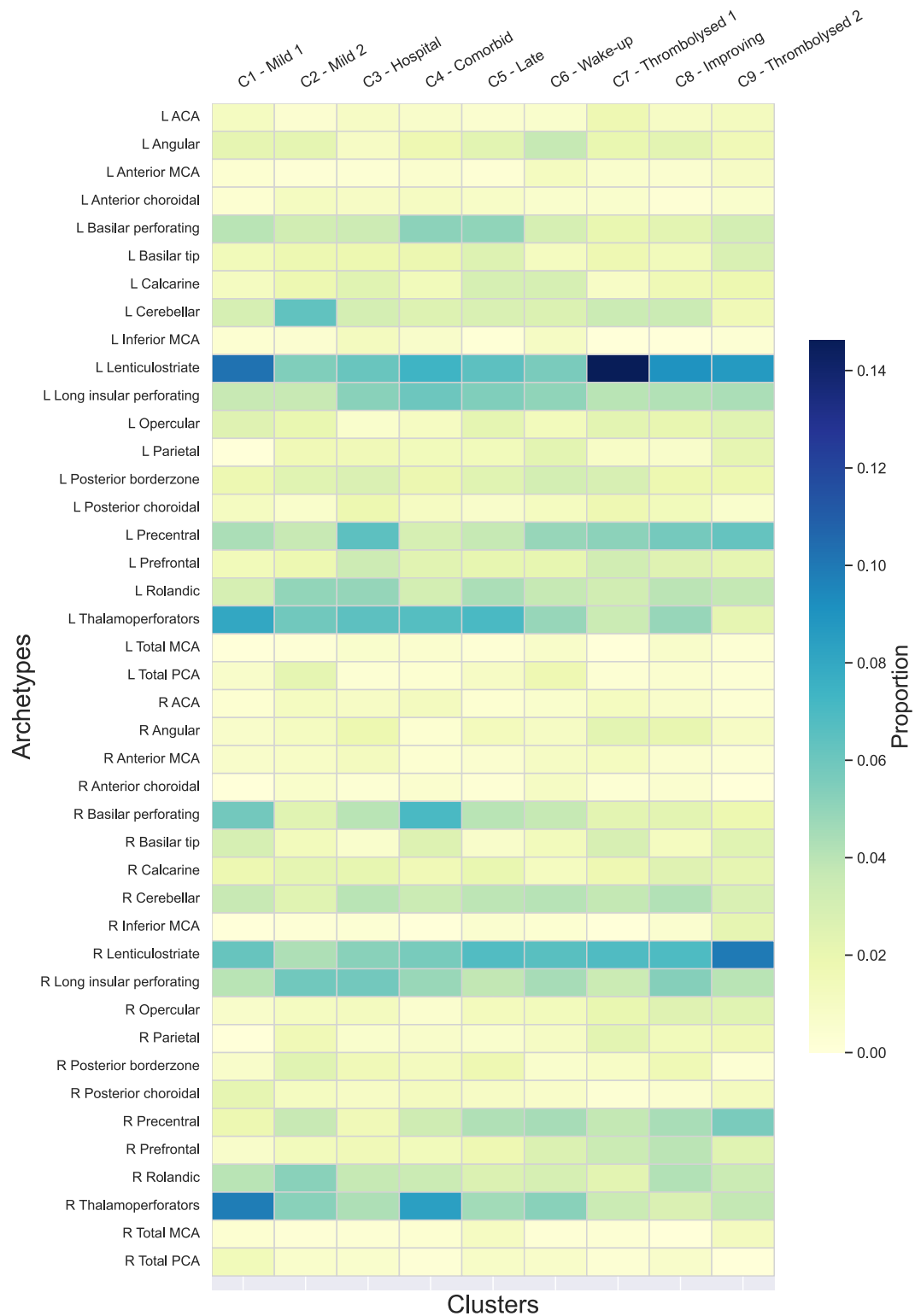


Figure S1. Heatmap showing the distribution of lesion archetypes across the nine clinical–demographic clusters.

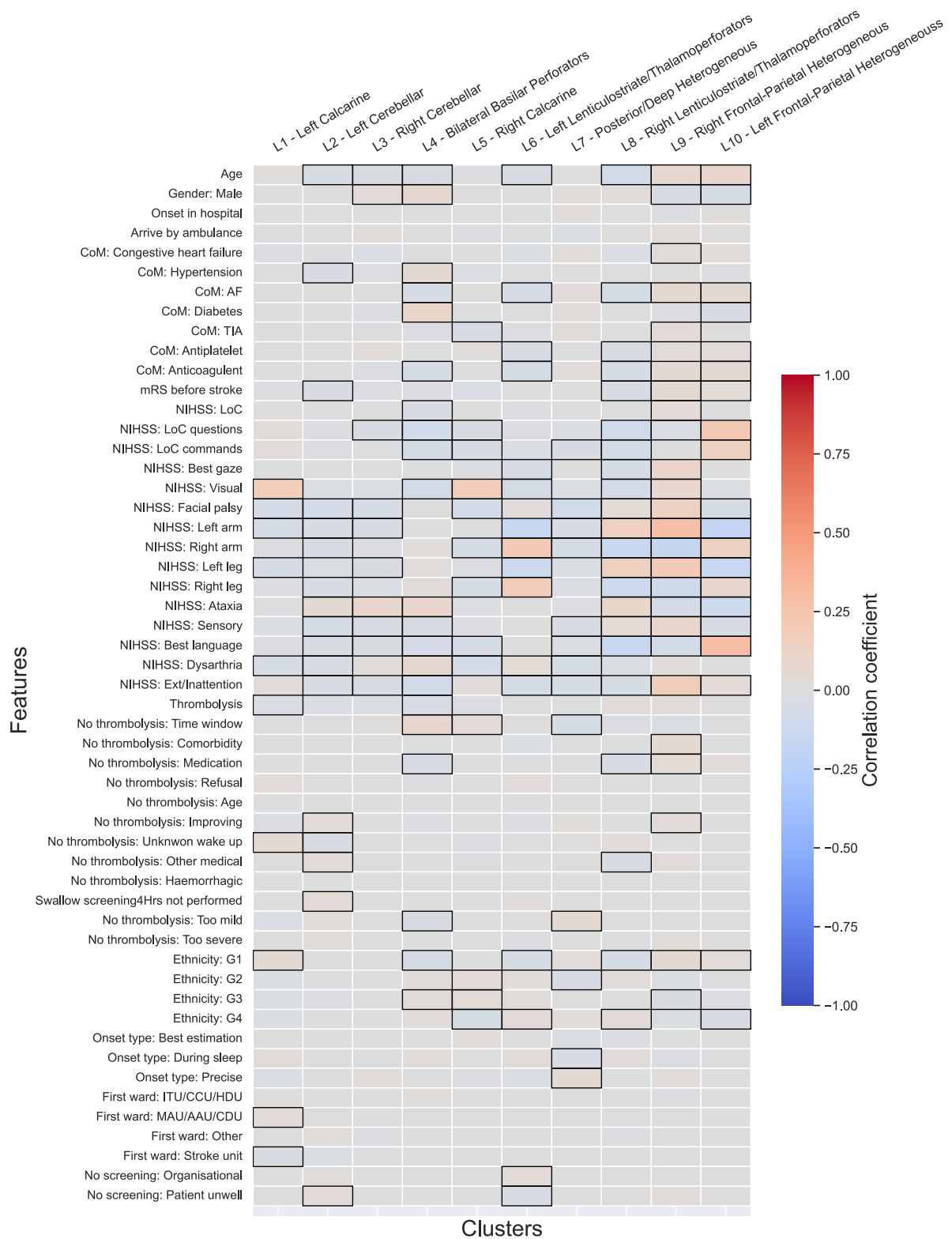


Figure S2. Heatmap showing the distribution of clinical-demographic features across the ten lesion clusters.

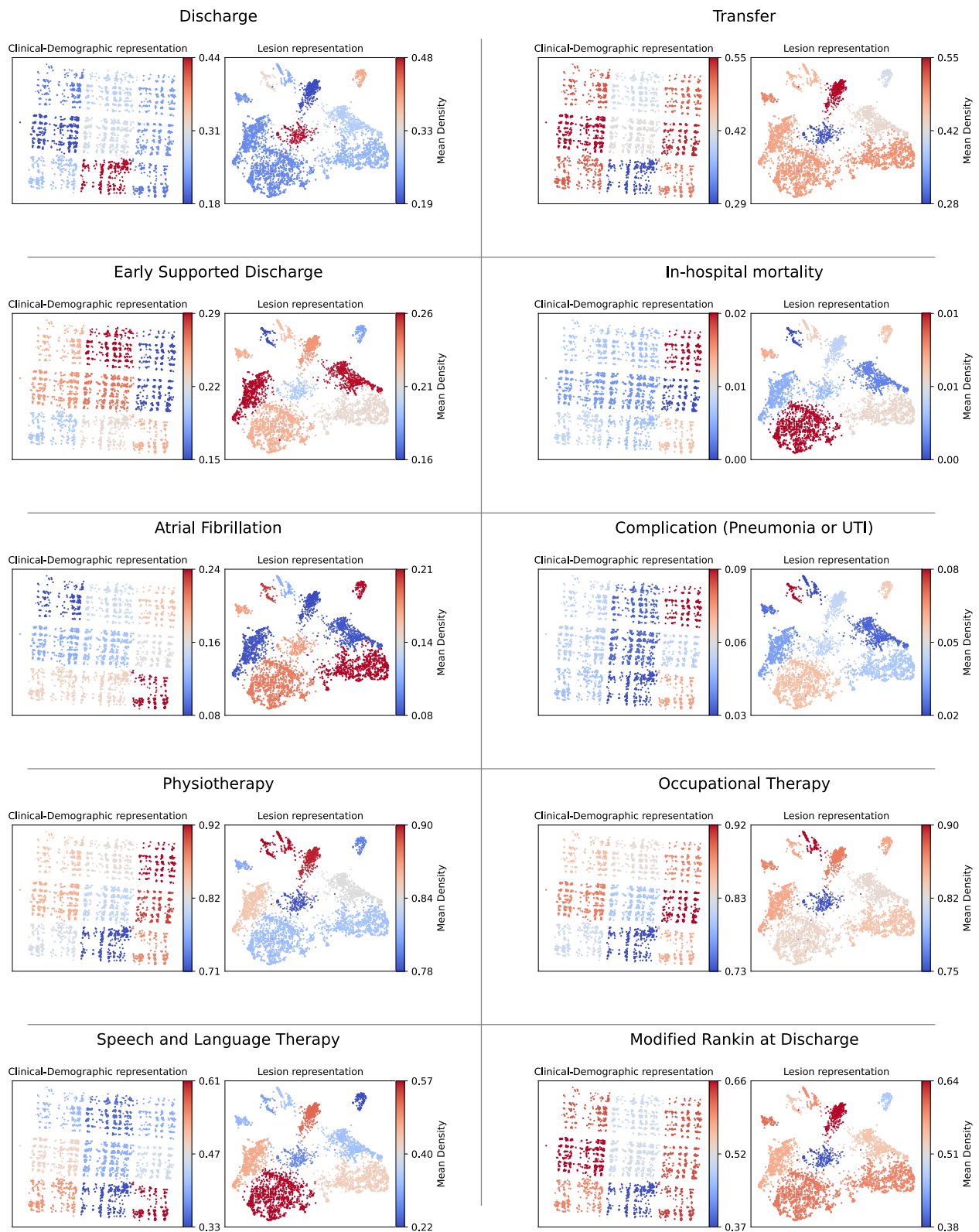


Figure S3. Each panel corresponds to a specific prediction target and contains two scatter plots: the clinical-demographic latent space (left) and the lesion latent space (right). Points represent latent clusters and are coloured according to the mean density of the corresponding target within each cluster, using a continuous colourmap with the colour bar normalised to the minimum and maximum values in each representation. The axes values are not shown; the range is identical across all plots.

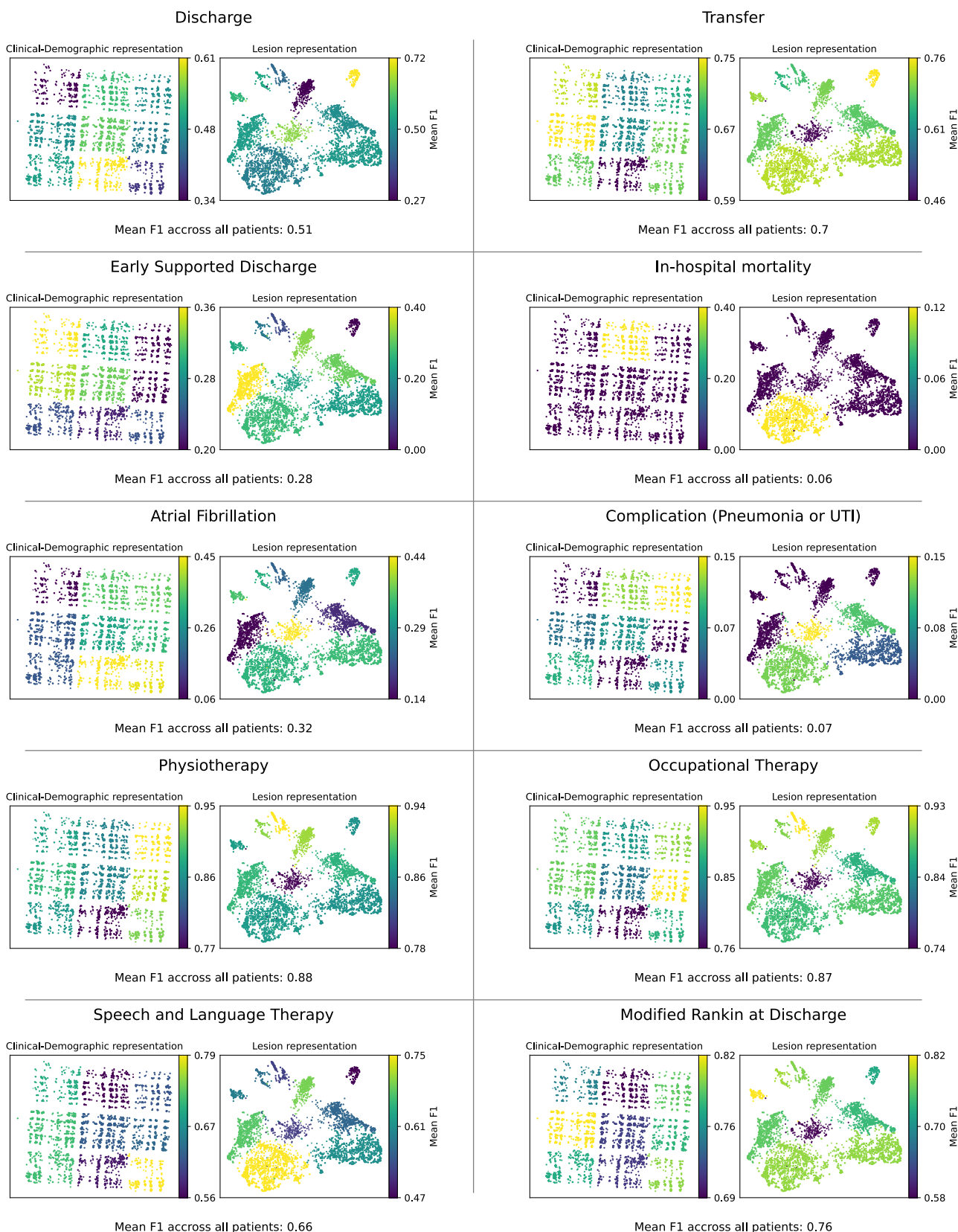


Figure S4. Identical to Fig. 5 in the main text, except that it reports the F1 score per cluster instead of balanced accuracy. The axes values are not shown; the range is identical across all plots.

Sentinel Stroke National Audit Programme (SSNAP) data

SSNAP attributes used as predictors

Attribute Name	Description
S1AgeOnArrival	Age of patient on arrival at hospital
S1Gender	Patient's gender
S1Ethnicity	Patient's ethnicity
S1OnsetInHospital	Whether stroke onset occurred while patient was already in hospital
S1OnsetDateType	Type/precision of stroke onset date (e.g., precise time, best estimate)
S1ArriveByAmbulance	Whether patient arrived at hospital by ambulance
S1FirstWard	First ward the patient was admitted to
S1FirstStrokeUnitArrivalNA	Whether arrival at first stroke unit is not applicable
S2CoMCongestiveHeartFailure	Comorbidity: Congestive heart failure
S2CoMHypertension	Comorbidity: Hypertension
S2CoMAtrialFibrillation	Comorbidity: Atrial fibrillation
S2CoMDiabetes	Comorbidity: Diabetes
S2CoMStrokeTIA	Comorbidity: Previous stroke or TIA (Transient Ischemic Attack)
S2CoMAFAntiplatelet	Atrial fibrillation patient on antiplatelet therapy
S2CoMAFAnticoagulant	Atrial fibrillation patient on anticoagulant therapy
S2RankinBeforeStroke	Modified Rankin Scale score before stroke (pre-stroke disability level)
S2NihssArrivalLoc	NIHSS on arrival: Level of consciousness
S2NihssArrivalLocQuestions	NIHSS on arrival: Level of consciousness questions
S2NihssArrivalLocCommands	NIHSS on arrival: Level of consciousness commands
S2NihssArrivalBestGaze	NIHSS on arrival: Best gaze
S2NihssArrivalVisual	NIHSS on arrival: Visual field testing
S2NihssArrivalFacialPalsy	NIHSS on arrival: Facial palsy
S2NihssArrivalMotorArmLeft	NIHSS on arrival: Motor function left arm
S2NihssArrivalMotorArmRight	NIHSS on arrival: Motor function right arm
S2NihssArrivalMotorLegLeft	NIHSS on arrival: Motor function left leg
S2NihssArrivalMotorLegRight	NIHSS on arrival: Motor function right leg
S2NihssArrivalLimbAtaxia	NIHSS on arrival: Limb ataxia
S2NihssArrivalSensory	NIHSS on arrival: Sensory
S2NihssArrivalBestLanguage	NIHSS on arrival: Best language
S2NihssArrivalDysarthria	NIHSS on arrival: Dysarthria
S2NihssArrivalExtinctionInattention	NIHSS on arrival: Extinction and inattention
S2Thrombolysis	Whether thrombolysis was performed
S2ThrombolysisNoButTimeWindow	Thrombolysis not given: Outside time window

Attribute Name	Description
S2ThrombolysisNoButComorbidity	Thrombolysis not given: Due to comorbidity
S2ThrombolysisNoButMedication	Thrombolysis not given: Due to medication contraindication
S2ThrombolysisNoButRefusal	Thrombolysis not given: Patient/family refusal
S2ThrombolysisNoButAge	Thrombolysis not given: Due to age
S2ThrombolysisNoButImproving	Thrombolysis not given: Symptoms improving
S2ThrombolysisNoButTooMildSevere	Thrombolysis not given: Symptoms too mild or too severe
S2ThrombolysisNoButTimeUnknownWakeUp	Thrombolysis not given: Time of onset unknown/wake-up stroke
S2ThrombolysisNoButOtherMedical	Thrombolysis not given: Other medical reason
S2ThrombolysisNoButHaemorrhagic	Thrombolysis not given: Hemorrhagic stroke
S2SwallowScreening4HrsNotPerformed	Whether swallow screening within 4 hours was not performed
S2SwallowScreening4HrsNotPerformedReason	Reason why swallow screening within 4 hours was not performed

SSNAP attributes used as outcomes

Attribute Name	Description
S7DischargeType – 'H'	Discharged to own home
S7DischargeType – 'T', 'TN'	Discharged to transfer - another inpatient care team ('T') or non-stroke specialist team ('TN')
S7DischargedEsdmr – 'NS', 'SNS'	Discharged with Early Supported Discharge and Multidisciplinary Team - Nursing home/Residential care with stroke support ('NS') or without stroke support ('SNS')
S7DischargeType – 'D'	Patient died during admission
S7DischargeAtrialFibrillation + S2CoMAtrialFibrillation - 'Y'	Patient with atrial fibrillation (either known on admission or diagnosed during admission)
S5UrinaryTractInfection7Days + S5PneumoniaAntibiotics7Days – 'Y'	Patient developed urinary tract infection and/or pneumonia requiring antibiotics within first 7 days
S4Physio – 'Y'	Patient received physiotherapy during admission
S4OccTher – 'Y'	Patient received occupational therapy during admission
S4SpeechLang – 'Y'	Patient received speech and language therapy during admission
S7RankinDischarge - >=3	Modified Rankin Scale score at discharge of 3 or higher (moderate to severe disability or death)

Ethnicity Codes and Groupings

Below is a descriptive mapping of the SSNAP ethnicity codes (A–S, Z, 99) used in our analysis. These codes were aggregated into five groups (G1–G5) as described in the Methods. The definitions and grouping scheme were derived from the SSNAP data dictionary.

Code	Description
A	White British
B	White Irish
C	Any other white background
D	White and Black Caribbean
E	White and Black African
F	White and Asian
G	Any other mixed background
H	Indian
J	Pakistani
K	Bangladeshi
L	Any other Asian background
M	Caribbean
M	African
P	Any other Black background
R	Chinese
S	Any other ethnic background
Z	Not stated
99	Not known