

RESEARCH

Open Access



Stroke recovery phenotyping through network trajectory approaches and graph neural networks

Sanjukta Krishnagopal^{1*} , Keith Lohse² and Robynne Braun³

Abstract

Stroke is a leading cause of neurological injury characterized by impairments in multiple neurological domains including cognition, language, sensory and motor functions. Clinical recovery in these domains is tracked using a wide range of measures that may be continuous, ordinal, interval or categorical in nature, which can present challenges for multivariate regression approaches. This has hindered stroke researchers' ability to achieve an integrated picture of the complex time-evolving interactions among symptoms. Here, we use tools from network science and machine learning that are particularly well-suited to extracting underlying patterns in such data, and may assist in prediction of recovery patterns. To demonstrate the utility of this approach, we analyzed data from the NINDS tPA trial using the Trajectory Profile Clustering (TPC) method to identify distinct stroke recovery patterns for 11 different neurological domains at 5 discrete time points. Our analysis identified 3 distinct stroke trajectory profiles that align with clinically relevant stroke syndromes, characterized both by distinct clusters of symptoms, as well as differing degrees of symptom severity. We then validated our approach using graph neural networks to determine how well our model performed predictively for stratifying patients into these trajectory profiles at early vs. later time points post-stroke. We demonstrate that trajectory profile clustering is an effective method for identifying clinically relevant recovery subtypes in multidimensional longitudinal datasets, and for early prediction of symptom progression subtypes in individual patients. This paper is the first work introducing network trajectory approaches for stroke recovery phenotyping, and is aimed at enhancing the translation of such novel computational approaches for practical clinical application.

Keywords: Stroke recovery, Disease subtyping, Network science, Network medicine, Graph neural networks

1 Introduction

1.1 Dynamic and multi-domain nature of stroke recovery

The process of neurological recovery after brain injuries such as stroke entails complex interactions among multiple variables that change dynamically over time [1, 2]. It is well known that the degree of recovery after stroke varies widely between individuals [3–6], where each patient's recovery pattern uniquely reflects the combined

influence of their lesion size and location [7], baseline health status, time to initial treatment [8], and response to medical treatment or rehabilitation, among many other intrinsic and extrinsic factors. Recovery trajectories furthermore vary depending on the specific neurological domain(s) affected (i.e., for motor, language, or sensory impairments) [9, 10], and each of these symptoms may show varying responsiveness to treatment. For example, language problems (aphasia), right-sided motor symptoms, and spatial perceptual problems (hemineglect) are reportedly less responsive than other symptoms to treatment with tissue plasminogen activator (tPA) [11]. Stroke recovery is therefore notoriously heterogeneous in terms

*Correspondence: s.krishnagopal@ucl.ac.uk

¹ Gatsby Computational Neuroscience Unit, University College London, London W1T 4JG, UK

Full list of author information is available at the end of the article

of the type and severity of residual symptoms, as well as the timecourse of progression and/or resolution of those symptoms [12].

An important goal for stroke research is to reduce the ‘noise’ arising from this inherent heterogeneity by stratifying patients who are likely to have similar symptom trajectories. The heterogeneity of symptoms and time-varying recovery patterns inherent to stroke make it an area especially well-suited area for data-driven approaches. The increasing availability of large scale stroke datasets has led to a recent explosion in the use of data-science methods for stroke research [13]. For example, machine learning analyses of stroke clinical data [12, 14] have been used to characterize symptom clusters [15], predict outcomes [3], and define composite measures of recovery [16].

1.2 Limitations of conventional regression and machine learning approaches versus network science

While machine learning (ML) approaches have been successful in a variety of analytical tasks, they often present challenges for interpretation and subsequent application. By contrast, network science tools are explicit in their modeling, making them more useful for studying medical data where clinical interpretation is paramount. Additionally, typical ML approaches focus on prediction, while taking the outcome itself at face value; in contrast, network approaches attempt to improve how the outcome itself is captured. While ML tools may not be as readily interpretable as network approaches for various types of analyses (such as understanding the interactions that underlie disease recovery patterns) ML still presents several desirable properties, particularly in terms of data-driven predictive ability, and may therefore be useful for prognostication of patient recovery patterns.

Conventionally, statistical tools such as mixed-effects regression are used for modeling longitudinal data in disciplines where repeated measures designs are particularly relevant, such as education, motor learning, and psychology [17–19]. Mixed-effects models have tremendous flexibility in their ability to accommodate different types of study designs and data types [20–22]. Such models are thus increasingly used in fields like neurorehabilitation where serial measures of recovery constitute a central focus [2, 18, 23]. Most importantly for our present purposes, mixed-effect models provide a means of estimating a unique trajectory for each person by combining the models fixed-effects with random-slopes and intercepts to obtain a unique (non)linear trajectory for each person. These trajectories can then be compared across different domains of recovery to see which domains covary or vary independently from each other.

1.3 Network science and trajectory profile clustering for stroke research

Insights into the complex patterns of symptom evolution can be gained through the computational power of network analysis. The field of network medicine [24] studies disease manifestation and progression as a function of multiple interacting disease variables, which may be of similar or different types. Network approaches also produce intuitive data visualizations that can facilitate interaction between clinicians and data scientists to yield novel insights on disease. However, with few exceptions, most network medicine studies have focused on biomolecular data [25, 26] rather than characterizing patients’ patterns of symptom progression over time.

Recently, Krishnagopal et al. [27] introduced a network-based approach called Trajectory Profile Clustering (TPC) that groups patients based on similar patterns of symptom evolution. The intuitiveness and ability of TPC to integrate variables on multiple different scales make it especially useful for studying disease severity, progression, and recovery. Multi-layer [28] types of trajectory clustering have also shown success in clinically validated disease trajectory prediction in Parkinson’s disease. We argue that TPC offers unique advantages for stroke recovery research based on its ability to simultaneously: (i) identify the dominant variables that differentiate stroke recovery subtypes; (ii) account for temporal disease progression patterns; and (iii) delineate distinct symptom groupings. This paper is the first work introducing network trajectory approaches for stroke recovery phenotyping, and is aimed at enhancing the translation of such novel computational approaches for practical clinical application.

When analyzing recovery trajectories with TPC, an obvious question that arises is at what stage do patients begin to stratify into distinct trajectory clusters (i.e., when do they begin to show symptom patterns unique to their recovery subtype)? The timing of medical treatments might be one important influence on the timecourse of recovery subtype stratification. For example in stroke, stratification might be expected to occur based on when patients receive treatments such as tPA or clot retrieval. Naively, this may appear to be a problem of simply measuring the differences between trajectory subtypes at each timepoint. However, since treatment efficacy for the same individual at different timepoints is not unrelated, more sophisticated tools are required to extract the timescale of separation. We can investigate these questions through graphical tools such as graph neural networks in machine learning.

1.4 Graph neural networks for the study of neurological disorders

The field of machine learning has been revolutionized by recent advances in deep neural networks, especially convolutional neural networks (CNNs) [29]. Conventionally, CNNs use local connections, shared weights and multiple layers to extract representations of data. However, CNNs work in a Euclidean domain and are best suited for use with images. By contrast, other deep learning methods can operate on a graph domain (i.e., graph neural networks or GNNs) [30]. Convolutional variants of graph neural networks provide a framework for transferring deep learning operators into a non-Euclidean (graphical) domain, and have been successful in a variety of tasks such as graph classification, node identification, link prediction in protein interactions, knowledge graphs, and social network analysis, among others. Of particular interest here, they have been successfully applied to study neurological disorders including Alzheimer’s disease and autism [31, 32], and could similarly have utility in the study of stroke.

2 Stroke dataset analyzed

To demonstrate the utility of using TPC and GNN for stroke recovery research, we analyzed cases from the well-characterized NINDS tPA trial data set [33]. This study was a randomized, double-blinded, placebo-controlled trial that compared the effects of intravenous tPA (a thrombolytic agent used in ischemic stroke to dissolve blood clots) versus placebo treatment in 624 patients. The data set captured neurologic deficits on the NIH Stroke Scale (NIHSS) [34], which is the most widely used measurement scale for stroke neurologic deficits, and has well-defined clinimetric properties [35, 36]. Each item is scored on a scale (from 0–2 or 0–5), with higher values indicating greater stroke severity. The NINDS tPA trial captured NIHSS scores across 5 time points: at hospital admission, at 2 h, 24 h, 7–10 days, and 3 months post-stroke. Here, we examined symptom progression in 11 neurologic domains as assessed by 15 individual item subscores on the NIHSS. A description of the items/variables is given in Table 1. We excluded a total of 135 cases who had imputed data at any time point (134) and/or had died (118). We excluded these cases because the imputation approach that had been used could distort patterns of change in scores for individual patients (i.e., missing values were imputed as the worst score for each NIHSS item). After exclusions, there were 489 remaining cases for analysis. Further, we treated time as a series of discrete observations, 0–4, starting with the patients’ assessment at admission. We have to treat time discretely rather than continuously because of how the

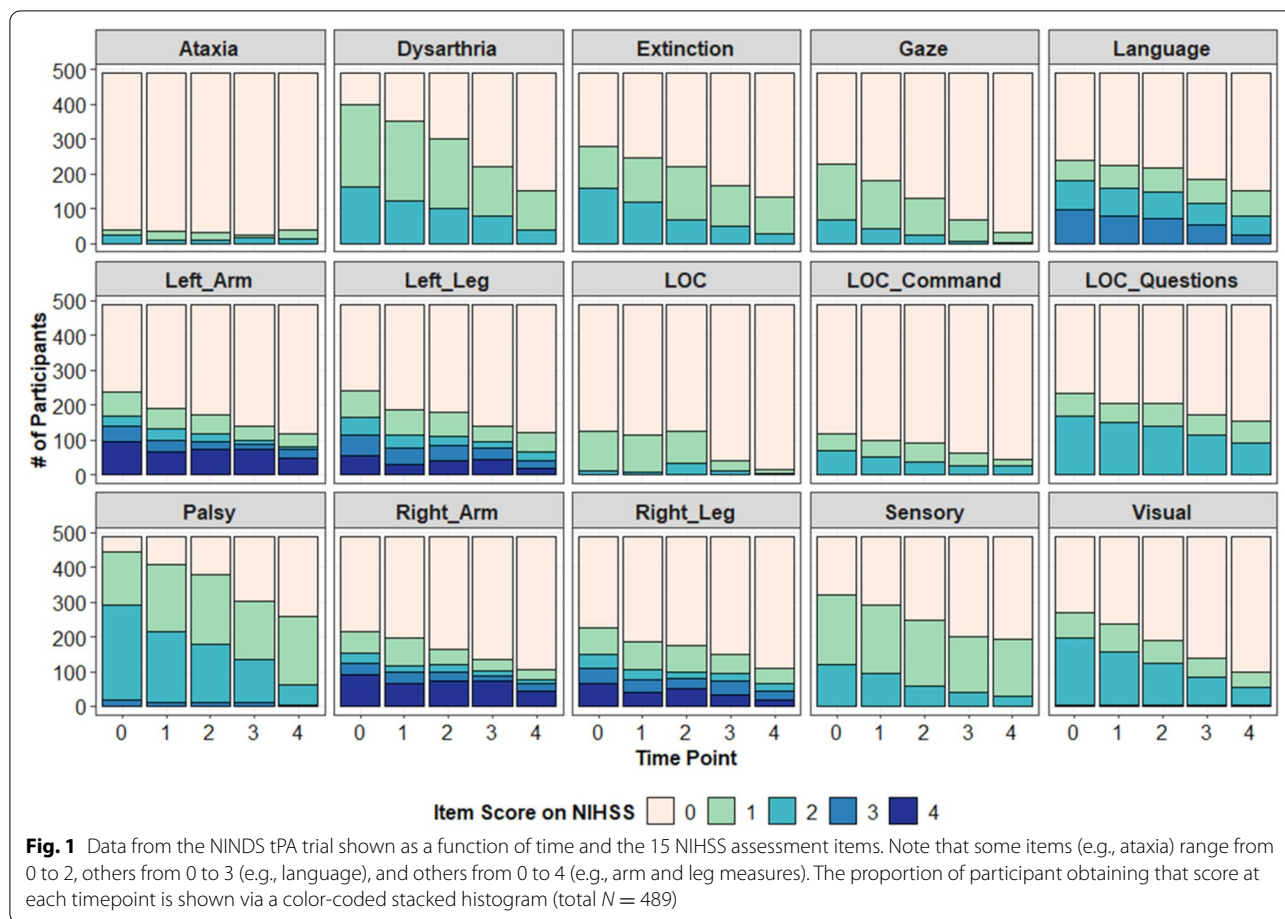
Table 1 NINDS tPA trial data, variable names and symptom descriptions

Variable name	Symptom description
ATAXIA	Coordination
CONSCIO (LOC)	Level of consciousness
DYSAR (dysarthria)	Speech (slurring)
EXTIN (extinction)	Spatial perception
GAZE	Eye movements
LANG (language)	Language
LOCCOM (LOC communication)	Command following
LOCQU (LOC questionnaire)	Question answering
MOTORLA (left arm)	Left arm strength
MOTORLL (left leg)	Left leg strength
MOTORRA (right arm)	Right arm strength
MOTORRL (right leg)	Right leg strength
PALSY	Facial weakness
SENSORY	Skin sensation (pain/pinprick)
VISUAL	Visual fields

data are coded in the NINDS tPA trial database. Ideally, we could measure time continuously in days or years, preserving the variability in assessment times [17–19], but that information was not available to us. Instead, in both the mixed-effect and TPC models, we fit trajectories based on discrete time. Although this transformation of the time variable means that absolute changes in time are arbitrary (i.e., 0–2 is 24 h, but 2–4 is potentially 3 months), relative changes in time are still meaningful (i.e., negative slopes mean that neurological deficits were improving over time, at the chosen timepoints) (Fig. 1).

2.1 Mixed-effects regression model

To obtain individual trajectories for each person on the different domains of the NIHSS, we fit a series of ordinal (cumulative link [21]) and Poisson (generalized linear [22]) mixed-models with random-intercepts and slopes for each subject. We focus on the ordinal models in the text because cumulative link models are designed to handle ordered, but non-continuous, response data. We present the Poisson models in the Additional file 1: Appendix S1. Although these models provide a reasonable fit to the data, it is less clear that the NIHSS items meet the assumption of treating the responses as counts (i.e., the language item is scored as 0 = normal, 1 = mild aphasia, 2 = severe aphasia, 3 = total/global aphasia, but severe aphasia is not necessarily twice mild aphasia). However, the Poisson model might be more familiar to many readers than the ordinal model, and the models do largely agree in correlations between trajectories over time as shown in the Additional file 1: Appendix S1. The



models do not completely agree, however, so we defer to the ordinal model as more appropriate given the scoring of the NIHSS.

The NINDS dataset contains repeated measures at 5 time points across 11 different neurological domains (as measured by 15 different NIHSS assessment items), resulting in 75 observations per patient. To understand how changes in different symptoms relate to each other over time, we extracted the random-effects from each model to get a unique trajectory for each individual in each domain. Ideally, this estimation could be done in a single multilevel model that nests time within each domain within each participant. However, because the NIHSS domains all have numerically different maxima, they cannot all be estimated in the same ordinal model. As such, we chose to fit a unique model for each domain, extract the slopes for each person, and then compare those slopes across domains. Thus, it is important to remember that *absolute* differences in the outcome are difficult (if not impossible) to interpret (i.e., is total sensory loss, a 2 on the sensory domain, equivalent

in severity to total gaze palsy, a 2 on the gaze domain?). However, *relative* changes across domains are still meaningful (i.e., negative slopes mean improvement over time for all domains and if the slopes are positively correlated, that means the symptoms tend to improve together).

Models also included fixed-effects of Time, Group (tPA versus placebo), and the Group \times Time interaction. Estimates, standard errors, and p -values for all models are presented in the Additional file 1: Appendix S1. Statistical significance was defined as ($\alpha = 0.05$) for all tests. Although the fixed-effects are presented in the results below, we want to emphasize that demonstrating the efficacy of tPA is not the goal of our analysis. Our goal is to describe how individuals change over time across domains and see which domains tend to be correlated with each other (doing this first with mixed-effect models and then with TPC). We present the effects of Group and the Group \times Time interactions as an internal validity

check so that these analysis can be compared to past work showing the efficacy of tPA [33].

2.2 Trajectory profile clustering

The Trajectory Profile Clustering algorithm [27] is designed to group together patients based on the similarities of their disease trajectories. In essence, it uses graphical tools to generate trajectory profiles for each individual that track their evolution of symptoms across time, then clusters them into communities of similarly behaving individuals that define a recovery subtype. The algorithm proceeds as follows:

1. *Model using bipartite networks* At time point t we construct a bipartite graph modeling connections between N individuals and V disease variables/symptoms. The connections between the individuals and symptoms are encoded through an adjacency matrix A_t of size $N \times V$. For M time points, we can represent the set of these bipartite graphs as an $N \times V \times M$ by stacking the A_t across time points to generate a tensor X where X_{ivt} gives the value of individual i 's disease symptom v at time t .
2. *Threshold and binarize to obtain trajectory profile* We threshold each symptom to set values less than a fixed fraction κ of the maximum score for the symptom to zero. For example, if symptom v takes score values in $(0, 1, 2, 3, 4, 5)$, if $\kappa = 0.5$, we binarize scores such that scores below $5 \times 0.5 = 2.5$ are set to 0, and above are set to 1. We call this thresholded matrix the trajectory profile matrix, T^i for patient i , that contains a representation of how the set of symptoms that a patient is *severely affected* by varies with time. The matrix entries of T^i are calculated as follows:

$$T_{vt}^i = 0 \text{ if } X_{ivt} \leq \max(v) \cdot \kappa, \tag{1}$$

$$= 1 \text{ if } X_{ivt} > \max(v) \cdot \kappa. \tag{2}$$

Since the range of values for each symptom represents the entire scale of severity, this thresholding ensures that patients are only considered 'connected' to symptoms that they severely express.

3. *Create a patient–patient network based on trajectory similarity* We create a patient–patient network P of all patients. The nodes of this network denote patients, and the strength of a link between patient i and patient j captures the similarity of their trajectory profiles. P has an adjacency matrix given by:

$$P_{ij} = \sum_{v,t} (T_{vt}^i \equiv T_{vt}^j). \tag{3}$$

In other words, P_{ij} gives the number of matrix entries for which trajectory profile T^i has the same value as T^j . This formulation implies that symptoms are *equally weighted*. While the approach is amenable to non-uniform weighting, there is little clinical consensus on the relative importance of different symptoms. Hence, in the interest of not introducing external bias, we choose uniform weighting, adopting an agnostic approach that assumes all symptoms/indicators are equally important. Other applications may require unequal weighting for symptoms and different time points, in which case one may calculate the patient–patient matrix as follows: $P_{ij} = \sum_{v,t} w_{vt} (T_{vt}^i \equiv T_{vt}^j)$ where w_{vt} is the weight of symptom v at time t .

4. *Cluster the network to identify subtypes* We then perform Louvain community detection [37] to maximize the Newman–Girvan modularity function [38] on the network defined by the adjacency matrix P . Such community detection allows us to identify 'communities' of patients, where individuals within a community have a relatively more similar stroke recovery profiles than patients between communities. As is common in network community detection approaches, the number of communities is not set a priori, but rather chosen so that the modularity is maximized. This process allows us to cluster trajectory profiles, and hence patients, into subtypes which have high intra-subtype similarity. The subtypes are denoted by C^1, C^2, \dots, C^L , where each C^l is a collection of trajectory profiles of the patients in that subtype, and L is the total number of subtypes.
5. *Construct aggregate profiles to characterize each subtype:* We average the trajectory profiles of all patients in each community C^l to obtain the 'community/subtype profile' S^l . The subtype profile is indicative of the symptom features that describe the subtype. More specifically, it is the normalized average of the trajectory profiles of all the patients in that subtype, i.e., S^l is a $V \times M$ matrix with elements defined by

$$S_{vt}^l = \frac{\sum_{i \in C^l} T_{vt}^i}{N_l}, \tag{4}$$

where N_l is the total number of individuals in subtype C^l .

2.3 Graph neural network

Graph neural networks (GNNs) [30] are a machine learning approach that captures the relationships represented in graphs through message passing between the nodes of those graphs. GNNs take a graph as input and pass them several layers of nodes, artificial 'neurons'. Here we use

graph neural networks to identify the timepoints that are most relevant in determining stroke recovery subtypes. Specifically, we train a graph neural network on symptom–symptom graphs generated at each timepoint, and test the accuracy of a GNN in its ability to classify an individual into the correct recovery subtype using data from a single timepoint. A higher accuracy at a given timepoint implies that the recovery subtypes attributed to the patients are strongly correlated with the symptom profiles at that timepoint.

We generate a symptom–symptom interaction graph G_t at each timepoint t where the nodes represent the disease symptoms. This graph is undirected (i.e., if node x is connected to node y , vice-versa is also true, i.e., the adjacency matrix of this graph is symmetric). The graph is generated as follows. First we generate a symptom–patient binary interaction network for a given timepoint as in step (3) of the previous section. We then project it to symptom space to obtain the symptom interaction profile of each patient at a timepoint. The corresponding adjacency matrix (of size $V \times V$) for the graph for patient i is given by

$$G_{it} = T_{ivt}^T \times T_{ivt}. \quad (5)$$

Lastly, We repeat this for each individual such that there exist N symptom–symptom interaction graphs at each timepoint.

We then separate the individual cases into a training data set (70% of total individuals) and a test data set (30% of total individuals) used to validate our approach. A convolutional graph neural network is trained on a graph classifying task for each time point, with labels provided by the subtypes/communities of that individual. The stratification of individuals into their recovery subtypes at each timepoint is then measured by testing the accuracy of the GNN on the test data for each timepoint.

The graph neural network takes as input symptom–symptom networks where we consider the 15 NIHSS assessment items as the nodes. The network consists of an input layer, a single hidden layer, and an output layer. The hidden layer comprises 64 artificial neurons. The input is processed through two graph convolutional layers with ReLU nonlinearities. We then calculate the graph representation by averaging all the neuron representations in the output layer, which contains an equal number of neurons to the number of subtypes. The output is passed through a softmax classifier that yields the probability of the graph belonging to a particular category/subtype. We use cross-entropy loss and the adaptive moment estimation (ADAM) optimizer.

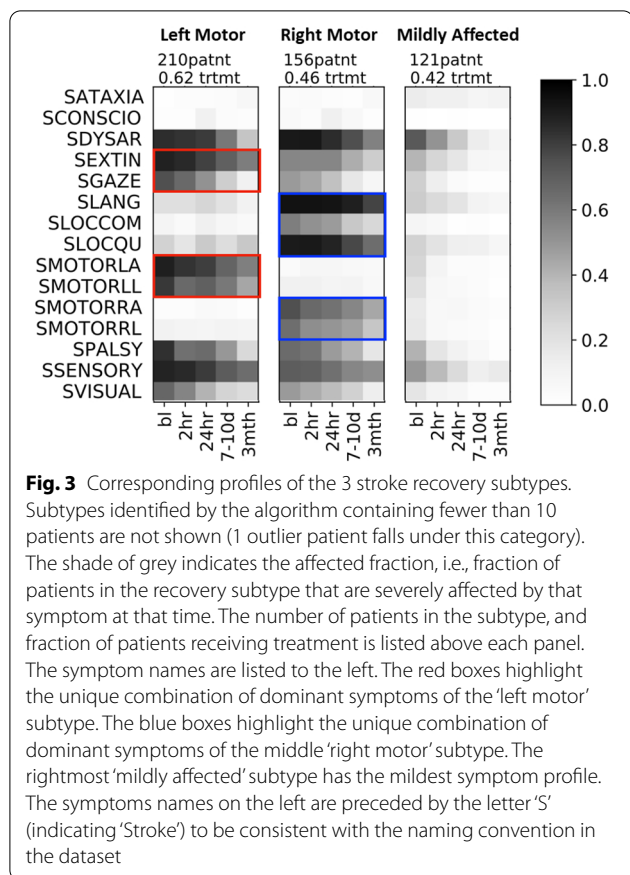
3 Results

3.1 Key patterns in slopes from the ordinal mixed-effect models

Full details of the models are presented in the Additional file 1: Appendix S1. In brief, however, there were statistically significant negative effects of Time for several different domains of the NIHSS: LOC ($b = -0.43$), dysarthria ($b = -0.76$), visual extinction ($b = -0.38$), gaze ($b = -0.68$), language (-0.39), the left arm ($b = -0.99$), the right arm ($b = -0.65$), the left leg ($b = -0.66$), the right leg ($b = -0.55$), palsy ($b = -0.67$), and sensation ($b = -0.39$), showing that neurological deficits generally improved over time. Consistent with prior analyses of these data [11], there were also statistically significant effects of Group for extinction ($b = +0.51$), the left leg ($b = +0.49$), the right leg ($b = +0.74$), and sensation ($b = +0.38$), and Group \times Time interactions for dysarthria ($b = +0.13$), the left arm ($b = +0.36$), and palsy ($b = +0.15$), which showed that the placebo group fared worse overall or improved more slowly over time in several domains compared to the tPA group.

Inspection of the person-level coefficients of this model also provides some insights relevant to the current goal of creating behavioral phenotypes. As shown in Fig. 2, and as would be expected given the common occurrence of post-stroke hemiplegia (weakening on one side of the body), some of the strongest correlations were for the ipsilateral arm and leg. The right arm and leg showed a similar timecourse of change in impairment ($r = 0.80$), as did the left arm and leg ($r = 0.81$). Second, there are also patterns in the correlation matrix consistent with lateralization of function as affected by unilateral stroke. For instance, the NIHSS item for extinction was positively associated with left arm/leg deficits (both from right hemisphere damage; $r = 0.36 - 0.40$), and much less associated with right arm/leg deficits (from left hemisphere damage; $r = 0.08 - 0.14$). Gaze was also positively associated with extinction $r = 0.53$, possibly reflecting the fact that gaze deviation is typically more pronounced in patients with hemineglect [39]. Lastly, language was positively associated with the right arm/leg (all left hemisphere effected; $r = 0.56 - 0.57$) and trivially associated with the left arm/leg (which are right hemisphere effected, $r = -0.07 - 0.04$).

In sum, ordinal mixed-effect regression provides us with analytical replication of past work (i.e., the superiority of tPA to placebo as shown in several different domains of the NIHSS) and new insights into how sets of symptoms co-evolve over time. However, these correlations between individual trajectories do not tell us *which* individuals cluster together nor do they provide us clear guidance on *where* cut-offs between different



to a left hemisphere syndrome with primarily expressive aphasia). These findings are in alignment with prior factor analysis on the clinimetric properties of the NIHSS [40] and principal component analysis (PCA) to define common behavioral clusters [41], as well as results from our mixed-effects model in Sect. 3.1. The fact that our results capture clinically relevant subtypes and corroborate these prior findings supports the content validity of this analytic approach. Additionally, our clustering approaches reveal subtype structure at a finer scale (both in terms of symptoms as well as longitudinal symptom evolution) than can be achieved with PCA, and results that are clinically consistent.

It is notable that all three identified trajectory subtypes included both tPA- and placebo-treated patients, suggesting that treatment effects were less defining characteristics of patient recovery profiles than were initial severity and stroke laterality. TPC also provides interesting insights into patterns of symptom prevalence over time across subtypes. Spatial perception deficits (hemineglect) are present in both the left-motor and right-motor subtypes, but tend to be milder and have better resolution in left hemisphere strokes. This observation reinforces the importance of targeted screening during rehabilitation

for hemineglect symptoms in both left- and right-hemisphere stroke, since persistent milder symptoms that could be amenable to treatment might otherwise be overlooked. Visual deficits are also present in both left- and right-hemisphere strokes as would be expected, but contrary to conventional understanding that visual deficits resolve less well than hemineglect, the overall prevalence of persistent visual symptoms at 3 months is lower than for hemineglect.

3.2.1 Effects of tPA treatment and time

A natural extension of our TPC subtyping is to study the timecourse of stratification into these subtypes. One might wonder whether subtype (and consequently the expected recovery profile of a patient) is largely driven by baseline symptom severity or by early stroke treatments. Machine learning is particularly well-posed to answer such questions. Since we operate in the graph domain, we use graph neural networks. We first extract trajectory profiles independently on patients who had received tPA within 3 h of stroke onset compared to patients who had received a placebo. In Fig. 4, we see that the identified subtypes retain the same symptom clusters identified in Fig. 3, but that overall symptom severity is lower in the tPA-treated population (Fig. 4A), particularly in symptoms that are dominant identifiers of the group. For instance, in the left-motor group, assessment items for gaze, left arm and leg strength, and pain/pinprick sensation showed higher recovery in the tPA-treated group (Fig. 4A) versus placebo group (Fig. 4B). Similarly, in the right-motor group, assessment items for right arm and leg strength, and command-following (SLOCCOM) showed higher recovery in the tPA-treated group (Fig. 4A) compared to the placebo group (Fig. 4B). As expected, the effect of tPA is less obvious in the minimally impaired group. We explore in Fig. 5 the accuracy of a neural network in predicting the subtype of an individual given data at a discrete timepoint. We generated a symptom-interaction network for each individual at each timepoint, and trained a convolutional-GNN to learn properties of interaction with neighbors. The convolutions are used for averaging over the neighborhood. If the learned properties for that timepoint are separated according to subtypes, then data at that timepoint is considered a good predictor of the subtype. In the training stage, we assume that the subtype is known to the neural network, which attempts to learn correlations between symptom-interaction patterns and the subtype. We then test to identify if the features learned by the neural network are consistent with the actual subtypes of the test patients.

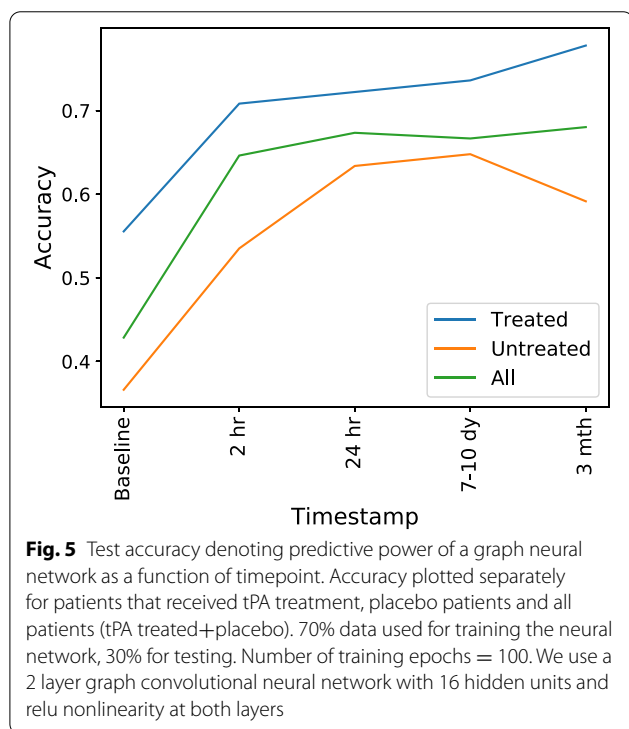
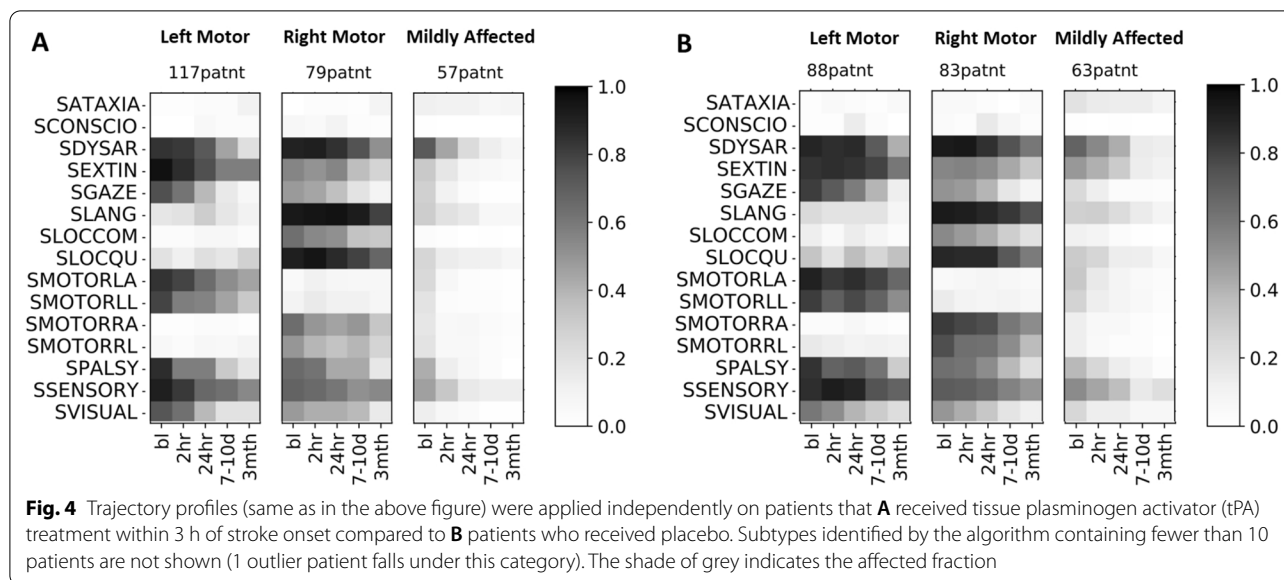


Figure 5 shows that there is a difference in predictive accuracy at baseline for the tPA vs. Placebo groups. The baseline timepoint is a more accurate predictor of subtype for patients who received tPA. This finding may seem unexpected if one posits that tPA ‘rescues’ an otherwise poor prognosis with severe baseline symptoms predicting poor outcomes. However, the predictive accuracy for the tPA group rapidly increases during the first

2 h, and given the expected timecourse for therapeutic effects of tPA, this finding provides additional validation of our approach. For the Placebo group, predictive accuracy grows at comparable rates (comparable slopes) up to the 24-h mark, showing peak predictive accuracy at the 7–10 day mark, with standard error on the order 10^{-2} . The tPA group showed a further uptick in predictive accuracy by the 3-month mark, suggesting that treatment continues to exert an effect on recovery subtype stratification even in the later stages of post-stroke. One might speculate that this is the result of tPA treatment salvaging a greater ‘reserve’ of neural tissue for later rehabilitation therapies to act upon. Our report on the rapid increase in predictive accuracy from 2 to 24 h post-stroke furthermore aligns with recent work by Heitsch et al. [42] who reported on the early change in NIHSS scores between 6 and 24 h as a dynamic phenotype associated with long-term outcomes.

4 Conclusion and future work

In this work we introduce a network-based, data-driven method for stroke recovery analysis. First, we analyze the NINDS tPA stroke dataset using conventional quantitative medicine methods including a ordinal mixed-effects regression model, examining the effects of time, group (tPA vs. Placebo), and their interactions across neurological domains. Further, to identify stroke recovery subtypes and examine their characteristics at a finer resolution, we use the Trajectory Profile Clustering method which accounts not only for symptom severity at different timepoints, but also symptom interactions and their temporal evolution. Of note, although the analytical approach is clinically agnostic, we identify

subtypes that are clinically relevant. In particular, we identify a mildly affected recovery subtype comprising a larger proportion of patients who received tPA. Additionally, we observed that the two other recovery subtypes stratify as left- versus right-sided hemiplegia. Additionally, we identified that left motor deficits are strongly correlated with deficits in gaze and extinction, whereas right motor deficits correlated with deficits in language. These results again are biologically relevant, and are further validated by convergent findings in the mixed-effects regression models. Lastly, we use graph neural networks to study how much of the stratification into subtypes is identifiable at different time points, and found that stroke recovery trajectories were largely defined within the first 24 h, consistent with the expected pharmacodynamics of tPA treatment delivered in the first 3 h after stroke.

This paper is the first work introducing network trajectory approaches for stroke recovery phenotyping, and is aimed at enhancing the translation of such novel computational approaches for practical clinical application. This work presents a data-driven method that is widely applicable to heterogenous neurological disorders such as stroke, and bridges the fields of predictive medicine and network informatics. Because our approach is uniquely adapted to accommodate input variables on multiple scales, future applications could include the integration of other types of data that may contribute to the heterogeneity of recovery, such as data on patient genotypes.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s40708-022-00160-w>.

Additional file 1: Appendix S1. Details of the ordinal (cumulative link) and Poisson (generalized linear) mixed-effect models.

Acknowledgements

We would like to acknowledge Michelle Girvan, Adam de Havenon, Michael M. Binkley, Laura Heitsch, and Alen Delic for helpful conversations and perspectives. We would also like to thank Steven C. Cramer and Arne G. Lindgren for their diligent critical review and comments on the manuscript.

Author contributions

SK developed and implemented the TPC algorithm. KL implemented and studied the mixed-effects model on the data. RB conceived, facilitated and guided the project. SK and KL analyzed the computational results and KL and RB interpreted the results in a medical context. All authors were involved in writing the manuscript. All authors read and approved the final manuscript.

Funding

This work was not externally funded.

Availability of data and materials

The NINDS dataset was first released in [33] and is publicly available. The code is available on [github/chimeraki/Stroke-Analysis](https://github.com/chimeraki/Stroke-Analysis).

Declarations

Competing interests

The authors declare that they have no competing interests.

Author details

¹Gatsby Computational Neuroscience Unit, University College London, London W1T 4JG, UK. ²Physical Therapy and Neurology, Washington University School of Medicine, 4444 Forest Park Ave., Suite 1101, St. Louis, MO 63108-2212, USA. ³Department of Neurology, University of Maryland School of Medicine, 655 W. Baltimore Street, Bressler Research Building, 12th Floor, Baltimore, MD 21201, USA, on behalf of the GPAS Collaboration, Phenotyping Core.

Received: 11 November 2021 Accepted: 23 April 2022

Published online: 19 June 2022

References

- Wang C, Winstein C, D'Argenio DZ, Schweighofer N (2020) The efficiency, efficacy, and retention of task practice in chronic stroke. *Neurorehab Neural Repair* 34(10):881–890
- Lohse K, Bland MD, Lang CE (2016) Quantifying change during outpatient stroke rehabilitation: a retrospective regression analysis. *Arch Phys Med Rehabil* 97(9):1423–1430
- Stinear CM, Barber PA, Petoe M, Anwar S, Byblow WD (2012) The PREP algorithm predicts potential for upper limb recovery after stroke. *Brain* 135(Pt 8):2527–2535. <https://doi.org/10.1093/brain/aws146>
- Wahl A-S, Schwab ME (2014) Finding an optimal rehabilitation paradigm after stroke: enhancing fiber growth and training of the brain at the right moment. *Front Hum Neurosci* 8:381. <https://doi.org/10.3389/fnhum.2014.00381>
- Simpkins AN, Janowski M, Oz HS, Roberts J, Bix G, Doré S, Stowe AM (2020) Biomarker application for precision medicine in stroke. *Transl Stroke Res* 11(4):615–627. <https://doi.org/10.1007/s12975-019-00762-3>
- Braun RG, Heitsch L, Cole JW, Lindgren AG, de Havenon A, Dude JA, Lohse KR, Cramer SC, Worrall BB, Core P et al (2021) Domain-specific outcomes for stroke clinical trials: what the modified rankin isn't ranking. *Neurology* 97(8):367–377
- Shelton FDN, Reding MJ (2001) Effect of lesion location on upper limb motor recovery after stroke. *Stroke* 32(1):107–112
- Dromerick AW, Edwardson MA, Edwards DF, Giannetti ML, Barth J, Brady KP, Chan E, Tan MT, Tamboli I, Chia R et al (2015) Critical periods after stroke study: translating animal stroke recovery experiments into a clinical trial. *Front Hum Neurosci* 9:231
- Cramer SC, Koroshetz WJ, Finklestein SP (2007) The case for modality-specific outcome measures in clinical trials of stroke recovery-promoting agents. *Stroke* 38(4):1393–1395
- Cramer SC, Wolf SL, Saver JL, Johnston KC, Mocco J, Lansberg MG, Savitz SJ, Liebeskind DS, Smith W, Wintermark M et al (2021) The utility of domain-specific end points in acute stroke trials. *Stroke* 52(3):1154–1161
- Felberg RA, Okon NJ, El-Mitwalli A, Burgin WS, Grotta JC, Alexandrov AV (2002) Early dramatic recovery during intravenous tissue plasminogen activator infusion: clinical pattern and outcome in acute middle cerebral artery stroke. *Stroke* 33(5):1301–1307
- van der Vliet R, Selles RW, Andrinopoulou E-R, Nijland R, Ribbers GM, Frens MA, Meskers C, Kwakkel G (2020) Predicting upper limb motor impairment recovery after stroke: a mixture model. *Ann Neurol* 87(3):383–393
- Wang W, Kiik M, Peek N, Curcin V, Marshall IJ, Rudd AG, Wang Y, Douiri A, Wolfe CD, Bray B (2020) A systematic review of machine learning models for predicting outcomes of stroke with structured data. *PLoS ONE* 15(6):0234722
- Stinear CM, Byblow WD, Ackerley SJ, Smith M-C, Borges VM, Barber PA (2017) Prep2: a biomarker-based algorithm for predicting upper limb function after stroke. *Ann Clin Transl Neurol* 4(11):811–820
- Sucharew H, Khoury J, Moomaw CJ, Alwell K, Kissela BM, Belagaje S, Adeyoye O, Khatri P, Woo D, Flaherty ML, Ferioli S, Heitsch L, Broderick JP,

- Kleindorfer D (2013) Profiles of the national institutes of health stroke scale items as a predictor of patient outcome. *Stroke* 44(8):2182–2187. <https://doi.org/10.1161/STROKEAHA.113.001255>
16. Hommel M, Detante O, Favre I, Touzé E, Jaillard A (2016) How to measure recovery? Revisiting concepts and methods for stroke studies. *Transl Stroke Res* 7(5):388–394. <https://doi.org/10.1007/s12975-016-0488-0>
 17. Singer JD, Willett JB, Willett JB et al (2003) *Applied longitudinal data analysis: modeling change and event occurrence*. Oxford University Press, Oxford
 18. Garcia TP, Marder K (2017) Statistical approaches to longitudinal data analysis in neurodegenerative diseases: Huntington's disease as a model. *Curr Neurol Neurosci Rep* 17(2):14
 19. Lohse KR, Shen J, Kozlowski AJ (2020) Modeling longitudinal outcomes: a contrast of two methods. *J Motor Learn Dev* 8(1):145–165
 20. Bates D, Mächler M, Bolker B, Walker S (2015) Fitting linear mixed-effects models using lme4. *J Stat Softw* 67(1):1–48. <https://doi.org/10.18637/jss.v067.i01>
 21. Christensen RHB (2015) ordinal—regression models for ordinal data. R package version 28:2015. <http://www2.uaem.mx/r-mirror/web/packages/ordinal/>
 22. Rizopoulos D (2022) Glmadaptive: generalized linear mixed models using adaptive gaussian quadrature. R package version 0.8-5
 23. Kozlowski AJ, Heinemann AW (2013) Using individual growth curve models to predict recovery and activities of daily living after spinal cord injury: an scirehab project study. *Arch Phys Med Rehab* 94(4):154–164
 24. Pawson T, Linding R (2008) Network medicine. *FEBS Lett* 582(8):1266–1270
 25. Huang S (2004) Back to the biology in systems biology: what can we learn from biomolecular networks? *Brief Funct Genomics* 2(4):279–297
 26. Barabási A-L, Gulbahce N, Loscalzo J (2011) Network medicine: a network-based approach to human disease. *Nat Rev Genet* 12(1):56–68
 27. Krishnagopal S, Coelln Rv, Shulman LM, Girvan M (2020) Identifying and predicting parkinson's disease subtypes through trajectory clustering via bipartite networks. *PLoS ONE* 15(6):0233296
 28. Krishnagopal S (2020) Multi-layer trajectory clustering: a network algorithm for disease subtyping. *Biomed Phys Eng Express* 6(6):065003
 29. Albawi S, Mohammed TA, Al-Zawi S (2017) Understanding of a convolutional neural network. In: 2017 international conference on engineering and technology (ICET), pp. 1–6. IEEE
 30. Scarselli F, Gori M, Tsoi AC, Hagenbuchner M, Monfardini G (2008) The graph neural network model. *IEEE Trans Neural Netw* 20(1):61–80
 31. Song T-A, Chowdhury SR, Yang F, Jacobs H, El Fakhri G, Li Q, Johnson K, Dutta J (2019) Graph convolutional neural networks for Alzheimer's disease classification. In: 2019 IEEE 16th international symposium on biomedical imaging (ISBI 2019), pp. 414–417. IEEE
 32. Parisot S, Ktena SJ, Ferrante E, Lee M, Guerrero R, Glocker B, Rueckert D (2018) Disease prediction using graph convolutional networks: application to autism spectrum disorder and Alzheimer's disease. *Med Image Anal* 48:117–130
 33. National Institute of Neurological Disorders and Stroke rt-PA Stroke Study Group (1995) Tissue plasminogen activator for acute ischemic stroke. *N Engl J Med* 333(24):1581–1587. <https://doi.org/10.1056/NEJM199512143332401>
 34. Brott T, Adams HP, Olinger CP, Marler JR, Barsan WG, Biller J, Spilker J, Holleran R, Eberle R, Hertzberg V (1989) Measurements of acute cerebral infarction: a clinical examination scale. *Stroke* 20(7):864–870. <https://doi.org/10.1161/01.str.20.7.864>
 35. Millis SR, Straube D, Iramaneerat C, Smith EV, Lyden P (2007) Measurement properties of the National Institutes of Health Stroke Scale for people with right- and left-hemisphere lesions: further analysis of the clomethiazole for acute stroke study-ischemic (class-I) trial. *Arch Phys Med Rehab* 88(3):302–308. <https://doi.org/10.1016/j.apmr.2006.12.027>
 36. Goldstein LB, Samsa GP (1997) Reliability of the national institutes of health stroke scale: extension to non-neurologists in the context of a clinical trial. *Stroke* 28(2):307–310
 37. Blondel VD, Guillaume J-L, Lambiotte R, Lefebvre E (2008) Fast unfolding of communities in large networks. *J Stat Mech Theo Exp* 2008(10):10008
 38. Girvan M, Newman ME (2002) Community structure in social and biological networks. *Proc Natl Acad Sci* 99(12):7821–7826
 39. Berger MF, Pross R, Ilg U, Karnath H (2006) Deviation of eyes and head in acute cerebral stroke. *BMC Neurol* 6(1):1–8
 40. Lyden P, Lu M, Jackson C, Marler J, Kothari R, Brott T, Zivin J (1999) Underlying structure of the national institutes of health stroke scale: results of a factor analysis. *Stroke* 30(11):2347–2354
 41. Corbetta M, Ramsey L, Callejas A, Baldassarre A, Hacker CD, Siegel JS, Astafiev SV, Rengachary J, Zinn K, Lang CE et al (2015) Common behavioral clusters and subcortical anatomy in stroke. *Neuron* 85(5):927–941
 42. Heitsch L, Ibanez L, Carrera C, Binkley MM, Strbian D, Tatlisumak T, Bustamante A, Ribó M, Molina C, Dávalos A et al (2021) Early neurological change after ischemic stroke is associated with 90-day outcome. *Stroke* 52(1):132–141

Submit your manuscript to a SpringerOpen® journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)
