

A patient distance metric for neurology

Daniel B. Hier^a, Jonathan Kopel^b, Steven U. Brint^a, Donald Wunsch II^c, Gayla R. Olbricht^d, Sima Azizi^c, Blaine Allen^c

^a Department of Neurology and Rehabilitation, University of Illinois at Chicago, Chicago, IL

^b Department of Internal Medicine, Texas Tech University Health Sciences Center, Lubbock TX

^c Department of Electrical and Computer Engineering, Missouri University of Science and Technology, Rolla MO 65401

^d Department of Mathematics and Statistics, Missouri University of Science and Technology, Rolla MO 65401

Corresponding Author:

Daniel B. Hier MD

912 S. Wood Street

University of Illinois at Chicago (MC 796)

Chicago IL 60612

Email: dhier@uic.edu

Telephone: (312) 622-2776

Daniel B. Hier
ORCID ID: <https://orcid.org/0000-0001-9603-0797>

Donald C. Wunsch II
ORCID ID: <https://orcid.org/0000-0002-9726-9051>

Jonathan Kopel
Orchid ID: <https://orcid.org/0000-0001-5934-2695>

Blaine Allen
Orchid ID: <https://orcid.org/0000-0001-9603-0797>

Abstract

Objective:

Neurologists lack a metric for measuring the distance between neurological patients. When neurological signs and symptoms are represented as neurological concepts from a hierarchical ontology and neurological patients are represented as sets of concepts, distances between patients can be represented as inter-set distances.

Methods:

We converted the neurological signs and symptoms from 721 published neurology cases into sets of concepts with corresponding machine-readable codes. We calculated inter-concept distances based a hierarchical ontology and we calculated inter-patient distances by semantic weighted bipartite matching. We evaluated the accuracy of a k-nearest neighbor classifier to allocate patients into 40 diagnostic classes.

Results:

Within a given diagnosis, mean patient distance differed by diagnosis, suggesting that across diagnoses there are differences in how similar patients are to other patients with the same diagnosis. The mean distance from one diagnosis to another diagnosis differed by diagnosis, suggesting that diagnoses differ in their proximity to other diagnoses. Utilizing a k-nearest neighbor classifier and inter-patient distances, the risk of misclassification differed by diagnosis.

Conclusion:

If signs and symptoms are converted to machine-readable codes and patients are represented as sets of these codes, patient distances can be computed as an inter-set distance. These patient distances given insights into how homogeneous patients are within a diagnosis (stereotypy), the distance between different diagnoses (proximity), and the risk of diagnosis misclassification (diagnostic error).

Keywords: ontology, patient similarity, distance metric, neurological findings, signs and symptoms, UMLS, SNOMED CT, classification, diagnostic error, proximity

Background and Related Work

Neurological diagnosis is difficult [1-2]. Neurologists rely on symptoms (patient complaints) and signs (examination findings) to come to a preliminary diagnosis (sometimes called a *working diagnosis*). Neurologists use testing to make a *definitive diagnosis*. Errors in diagnosis are inevitable and reflect the complexity of neurologic diagnosis. Most studies indicate an error rate of 20 to 35% when the preliminary diagnosis is compared to a definitive diagnosis after testing [3-5].

Neurologists think about neurological diagnosis in three ways: *diagnosis stereotypy*, *diagnosis proximity*, and *diagnosis difficulty*. Diagnosis stereotypy is the idea that some diagnostic conditions are more stereotyped in presentation than others. For example, Bell palsy presents reliably with unilateral facial weakness, decreased taste on half the tongue, and hyperacusis. Similarly, Parkinson disease presents reliably with bradykinesia, cogwheel rigidity, and resting tremor. On the other hand, non-stereotyped diseases such as multiple sclerosis present variably with inconstant combinations of weakness, hyperreflexia, sensory loss, visual symptoms, cognitive symptoms, and ataxia.

Diagnosis proximity is the idea that for a given diagnosis, some diagnoses are closer while others are more distant. For example, most neurologists place basal ganglia disorders such as striatonigral degeneration, progressive supranuclear palsy, and Parkinson disease close to each other and remote from cranial nerve disorders such as Bell palsy, trigeminal neuralgia, and Ramsay Hunt syndrome.

Diagnosis difficulty is the idea that diagnostic error is more likely for some diagnoses than others. One measure of diagnostic difficulty is how often a patient with a given diagnosis is misclassified as having another diagnosis. For example, studies have found that multiple sclerosis, Guillain Barré syndrome, subdural hematoma, and myasthenia gravis are often misclassified at the bedside and are especially prone to diagnostic error [6].

Neurologists lack a mathematical representation of diagnosis stereotypy, diagnosis proximity, and diagnosis difficulty. A distance metric for neurological patients is a step towards representing these ideas in mathematical form.

The creation of a distance metric for neurological patients based on signs and symptoms is challenging. First, neurological symptoms and neurological signs are recorded in the electronic health record as unstructured free text. Second, examiners use a variety of equivalent terms to represent the same meaning: *hyperreflexia* is equivalent to *increased reflexes*; *Babinski sign* is equivalent to *extensor plantar response*; and so on. Third, the number of signs and symptoms may vary from patient to patient. Some patients may have as few as one or two signs, while other more complex patients may have as many as 10 or 20 different signs. Fourth, converting unstructured text from electronic health records into machine-readable codes is difficult [7-8].

The SNOMED CT ontology and the UMLS Metathesaurus allow the consolidation of multiple synonymous terms under the same concept [9-10]. Both terminologies assign unique machine-readable codes to a concept. We have identified 1167 core concepts from the UMLS Metathesaurus as a basis for capturing the signs and symptoms of the neurological examination [11].

When signs and symptoms are converted to concepts and represented as machine-readable codes, patients can be instantiated mathematically as a set or as a vector. If sets are used to represent a patient, each sign or symptom is added to the set as a unique element. The cardinality of the set (number of set elements) is equal to the number of signs and symptoms. When a patient is represented as a vector, each sign or symptom can be represented as a dimension of the vector so that the number of dimensions is equal to the number of signs and symptoms. The magnitude of the vector in each dimension indicates whether the finding is present or absent. A variety of distance metrics are available to calculate the distances between vectors and sets including the Manhattan distance, cosine distance, Euclidean distance, and

Mahalanobis distance. When patients are represented as sets of concepts (Figure 1), a commonly used metric is the Jaccard similarity [Equation 1].

$$J_{sim}(A, B) = \frac{|A \cap B|}{|A \cup B|} \quad [1]$$

In Equation [1], the Jaccard similarity [12] between Set A and Set B is the intersection of Set A with Set B divided by the union of Set A with Set B. Haase et al. [13] introduced a bipartite graph metric of set similarity.

$$Sim(A, B) = \frac{1}{|A|} * \sum_{a \in A} \max_{b \in B} (sim(a, b)) \quad [2]$$

In Equation [2], the sum of the maximum similarity between each element in set A with an element in Set B is divided by |A|, the number of elements in set A. Girardi et al. [14] and Mabotuwana [15] have suggested that the accuracy of inter-set distances can be enhanced when inter-concept distances from a concept hierarchy [16-26] are included in the calculation of inter-set distance. Inter-concept measures of distance can be combined with inter-set measures of distance to compute inter-patient distances [27]. Currently, no single concept-based inter-set measure of patient distance has proven superior to others.

Methods

Case abstraction.

We abstracted 4595 signs and symptoms from 721 published neurological patient studies [28-39]; mean 6.4 ± 3.3 findings per patient. We mapped findings (signs and symptoms) to one of 1163 UMLS CUI (concept unique identifier) codes in the neurological examination ontology [11]. We represented each patient as an unordered set of unique CUI codes. For example, a patient with Alzheimer disease was converted to the set:

{“getting lost”, “forgetful”, “disorientation”, “irritable”, “impaired memory”}

and coded as

{C0424015, C0542476, C0233407, C0022107, C0233794}

Distance between signs and symptoms (inter-concept distances).

We calculated distances between signs and symptoms by the method of Wu and Palmer [16] utilizing the subsumption hierarchy in the neuro-ontology [11]. Distances between concepts were normalized with a minimum of 0.0 (closest) and a maximum of 1.0 (most distant).

$$dist(a, b) = 1 - \frac{2 * depth(LCS)}{depth(a) + depth(b)} \quad [3]$$

In Equation [3] $dist(a, b)$ is the semantic distance between concept a and concept b, LCS is the lowest common subsumer in the hierarchical ontology for both a and b, $depth(a)$ is number of levels from the root concept to concept a, $depth(b)$ is the number of levels from the root concept to concept b, and $depth(LCS)$ is the number of levels from the root concept to the LCS.

Distances between patients (inter-set distances).

We calculated distances between by *semantic weighted bipartite matching* [27].

$$distance(A, B) = \left(\frac{1}{|A| + |B|} \right) * \left(\sum_{a \in A} \min_{b \in B} (dist(a, b)) + \sum_{b \in B} \min_{a \in A} (dist(a, b)) \right) \quad [4]$$

In Equation [4], $distance(A, B)$ is the distance between Set A and Set B, $|A|$ is the number of elements in Set A, $|B|$ is the number of elements in Set B, and $dist(a, b)$ is the semantic distance between concept a and concept b based on a hierarchical ontology.

Statistical methods.

We used Prism 8.4.0 for t-tests, ANOVA, scatter plots, and box plots. We used Orange 3.23.0 for k-nearest neighbor classification and multidimensional scaling (MDS).

Stratification of the dataset by diagnosis.

We stratified the dataset of 721 patients into four diagnostic tranches: *movement disorders*, *nerve and muscle disorders*, *cranial nerve disorders*, and *brain and spinal cord disorders*. Each of the four tranches consisted of 10 diagnoses. Each diagnosis occurred 3 or more times in the tranche (Table 1).

Results

Mean inter-patient distances are lower with semantic weighted bipartite matching than with the Jaccard distance (unpaired t-test, $df = 1440$, $p < .0001$). Because of the uncommonness of exact concept matches between patients, the mean inter-patient distance by the Jaccard distance metric is 0.98 ± 0.01 , very close to the maximum of 1.00. By using the semantic weighted bipartite matching, the mean inter-patient distance for the 721 patients drops to 0.59 ± 0.04 . The greater dispersion of inter-patient distances with semantic weighted bipartite matching is illustrated when comparing inter-patient distances between a single patient with multiple sclerosis with the other 720 patients in the dataset (Figure 2). The histogram illustrates that the semantic weighted bipartite metric results in smaller inter-patient distances and greater dispersion of the distances than the Jaccard distance.

Based on patient diagnosis, we divided the dataset into four diagnostic tranches, each with ten diagnosis classes (Table 1). Each of the 40 diagnoses in the four diagnostic tranches occurred at least 3 times in the dataset. We calculated the mean intra-class patient distance for the 20 diagnostic classes in the brain and spinal cord tranche and the movement disorder tranche (Figure 3, means differ one-way ANOVA, $dF = 19$, $p < 0.0001$). We also calculated mean intra-class patient distance for the 20 diagnoses in the cranial nerve and nerve and muscle tranches (Figure 4, means differ, one-way ANOVA, $dF = 19$, $p < 0.0001$).

Mean inter-class distances were calculated for all 40 of the diagnosis classes shown in Table 1 (see data availability). Using inter-patient distances, we created a multidimensional scale (MDS) diagram for each of the four diagnostic tranches. Each point on the MDS represents a unique patient (Figures 5-8). Representative inter-class

distances are shown in Figure 10 (class distances from the class of ALS patients, means differ, one-way ANOVA, $dF = 9$, $p < 0.0001$)) and Figure 11 (class distances from the class of Parkinson disease patients, means differ, one-way ANOVA, $dF = 9$, $p < 0.0001$). Distance measures also allow the identification of the nearest neighbors to any specific patient in the dataset. Figure 12 shows the 20 nearest neighbors to a patient with amyotrophic lateral sclerosis, and Figure 13 shows the 20 nearest neighbors to a patient with Parkinson disease.

We used a k-nearest neighbor classifier to classify each patient in the four diagnostic tranches by diagnosis (Table 2). Based on performance of the classifier across a variety of values of k, we selected the value of $k = 3$ with uniform distance weighting (Figure 9) for all the tranches. F1, accuracy, recall, and precision for each of the four tranches is shown in Table 3 (means differ, one-way ANOVA, $dF = 3$, $p < 0.0001$).

Discussion

We have represented neurological patients as sets of signs and symptoms. We have demonstrated a reduction in inter-patient distances when concept similarity based on an ontology hierarchy is considered in calculating inter-patient distances. For the 721 patients in our dataset, the mean inter-patient distance went from 0.98 ± 0.01 to 0.59 ± 0.14 when concept similarity was considered in the distance metric. Furthermore, dispersion of inter-patient distances is improved when concept similarity is considered (Figure 2). The large reduction in mean inter-patient distances with the semantic weighted bipartite matching metric (0.39), reflects the high frequency of zero similarity between patients in the dataset by Jaccard similarity. For the 519,841 inter-patient comparisons made, 86.5% showed zero Jaccard similarity (i.e. that the two patients shared no concepts in common). Mabotuwana et al. [12] in a study comparing the similarity of radiology reports based on a semantic vector found that when concept similarity was not considered, 77% to 86% of the report comparisons shared no exact concept matches (zero similarity). By considering semantic similarity between concepts

based on the SNOMED CT hierarchy, they found a distance gain of approximately 0.10 by using concept similarity to enhance the calculation of inter-report distances.

The problem of stereotypy. Neurologists hold views that within certain diagnosis classes, patients are more similar to each other, i.e. more stereotyped in presentation, than in other diagnosis classes. Within a diagnosis class of patients, the inter-patient distance can be calculated between each patient and the other patients in the same class (Figures 3 and 4). We suggest that a low intra-class mean inter-patient distance implies greater intra-class similarity and that the presentation is more stereotyped in that class. We also suggest that a large mean inter-patient distance in a diagnosis class corresponds to more variability in presentation between patients in that class. The twenty diagnoses in Figure 3 are from the *movement disorder tranche* and the *brain and spinal cord tranche*. The twenty diagnoses from Figure 4 are from the *cranial nerve tranche* and *muscle and nerve tranche*. We suggest that diagnosis classes with smaller intra-class means (essential tremor, hemiballismus, and dystonia in Figure 3 and sixth nerve palsy, meralgia paresthetica, and Bell's Palsy in Figure 4) are more stereotyped in the presentation. On the other hand, we suggest that conditions such as subdural hematoma, multiple sclerosis, lumbar root disease, myelopathy, and polyneuropathy with large intra-class mean patient distances are more variable in presentation.

The problem of proximity. Diagnostic proximity is the idea that some diagnoses are similar to certain diagnoses and dissimilar to other diagnoses. The calculation of mean inter-patient distances across diagnoses allows the determination of which diagnoses are closer and which diagnoses are more distant from a given diagnosis. We calculated all inter-class distances for the 40 diagnoses in Table 1 (see data availability). Examples of calculation of inter-class distances is shown in Figures 10 and 11. For amyotrophic lateral sclerosis, the closest diagnosis classes are myelopathy, subdural hematoma, and multiple sclerosis. For Parkinson disease, the closest diagnosis classes are striatonigral degeneration, essential tremor, and Lewy body dementia. The MDS diagrams (Figures 5-8) offer an alternate way to visualize diagnosis proximity based on inter-patient distances. For example, Figure 5 illustrates the proximity of median nerve

neuropathy to ulnar neuropathy and the proximity of myasthenia gravis to myopathy. Figure 6 illustrates the proximity of Alzheimer disease to frontotemporal dementia as well as the proximity of amyotrophic lateral sclerosis to myelopathy. Figure 7 illustrates the proximity of Parkinson disease to its close neighbors progressive supranuclear palsy and striatonigral degeneration.

The problem of misclassification.

Errors in diagnosis (misclassification) are a major challenge to medicine in general and neurology in particular [2-5, 40]. The reasons for diagnostic error among physicians are complex [41]. Although some patients present to physicians with stereotyped patterns of signs and symptoms that lead to a rapid diagnosis through matching to stored patterns, other patients present with signs and symptoms that do not fit a stored pattern and require hypothesis generation and testing. These types of disease presentations pose a greater diagnostic challenge to the physician and a greater risk of diagnostic error [40,41]. We examined the performance of the k-nearest neighbor classifier for 40 different diagnostic conditions in 4 different diagnosis tranches (Table 2). Mean accuracy of diagnosis varied between 76% and 87% in the four tranches (Tables 2 and 3). These accuracies are comparable to the bedside diagnostic accuracy of human experts [3-5]. The diagnoses most often misclassified in each of the four tranches were polyneuropathy, B₁₂ deficiency, Wilson disease, striatonigral degeneration, multiple sclerosis, herpes simplex encephalitis, Devic disease, and Ramsay Hunt syndrome. Factors that may make these diseases more prone to misclassification include overlap with other diagnoses (for example Ramsay Hunt syndrome overlaps with Bell Palsy; striatonigral degeneration with Parkinson disease, etc. and the lack of a highly stereotyped presentation as with B₁₂ deficiency and multiple sclerosis).

Implications for the neurologist and neurological plausibility.

Any proposed neurology inter-patient distance metric needs to be neurologically plausible. The calculated distances between individual neurology patients and the distances between diagnosis classes of neurology patients should conform with the thinking of neurologists about such proximities. These proximities should be compatible

with neurological knowledge as catalogued in neurology textbooks and the scientific literature. For example, the proximities between diagnoses illustrated in Figures 5-8 should make sense to neurologists who routinely see these types of patients. A distance metric between neurological patients based on signs and symptoms can allow neurologists to think quantitatively about how similar the presentation of one patient is to another, how similar the signs and symptoms of one diagnosis class is to another, and the risk of misclassification for certain diagnosis classes.

Limitations

One limitation of the proposed distance metric is that we did not consider the severity of deficits such as weakness or ataxia. When deficits were present, they were binarized as either present or absent and not graded in severity. Another limitation is that some of the diagnosis classes were narrower than others. Although some of the diagnosis classes were specific (Huntington disease, Alzheimer disease, and Parkinson disease), others were more general such as polyneuropathy, myopathy, and meningitis. This decision to use more general categories for some diagnosis classes reflects the reality that signs and symptoms alone are unlikely to distinguish specific causes of meningitis, polyneuropathy, or myopathy without additional ancillary testing. Studies are needed to compare the computed inter-patient distances to expert neurologic opinion. The validity of the results would be improved by a larger dataset of patients than the 721 in our sample. We accepted the signs and symptoms from these published studies as the ground truth for distance calculations. Studies have shown that when neurologists examine patients, they may disagree as to the ground truth [42]. A further limitation of the study is that we utilized published cases from the textbooks of neurology rather than de-identified patient records from electronic medical records. Furthermore, we abstracted concepts from the case histories rather than by an NLP algorithm. This process can be automated by natural language processing (NLP) methods [43-45]. We chose to use manual abstraction rather than NLP because we wanted to carefully curate a database of test patients with minimal coding errors. Our experience with MetaMap [44-45] is that as many as 25% of concepts will need post-processing to improve accuracy. In the future, we plan to extend our methods to a larger dataset of

patients and utilize signs and symptoms from de-identified patients in electronic health records. Future advances in NLP could make the conversion of signs and symptoms in EHR text to machine-readable codes more accurate and efficient [46-48].

Inter-rater reliability for abstracting clinical cases into UMLS codes or SNOMED CT codes is another concern [6-7]. Finally, we have not compared the semantic weighted bipartite distance metric to other available distance metrics [12, 27]. Additional work will need to be done to establish the reliability of the abstraction process and identify the best metrics to calculate patient distances between neurological patients.

Conclusions

Neurological signs and symptoms that reside in the electronic health record as unstructured data can be represented as sets of UMLS concepts and stored as UMLS CUI machine readable codes. Inter-concept distances for signs and symptoms can be calculated based on a concept hierarchy [11]. Using a semantic weighted bipartite matching metric, inter-patient distances can be calculated for neurological cases. When distances for groups of patients in the same diagnosis class are known, inter-patient distances can give important insights into variability within a given diagnosis class, the proximity between different diagnosis classes, and the risk of misclassification by diagnosis.

Abbreviations

CUI: UMLS concept unique identifier

UMLS: Unified Medical Language System

SCTID: SNOMED CT concept identifier

MDS: multidimensional scaling

Human Studies

The Institutional Review Board of the University of Illinois at Chicago approved this work.

Funding

None to report.

Conflicts of Interest

None to report.

Authors contributions.

Research design by DBH. Data collection by SUB, DBH, and JK. Data analysis by all. Manuscript writing and editing by all.

Data Availability

Neurology cases are available at <http://dx.doi.org/10.17632/z3d6hwrdrmh.2>

Inter-concept distances are available at <http://dx.doi.org/10.17632/svrx3wgc4.3>

Inter-patient distances are available at <http://dx.doi.org/10.17632/svrx3wgc4.4>

Inter-class distances for 40 diagnoses available at <http://dx.doi.org/10.17632/n9vbwc6v6s.2>

References

- [1] H.S. Barrows, The Diagnostic (Problem Solving) Skill of the Neurologist, *Arch. Neurol.* 1972; 26: 273. doi:10.1001/archneur.1972.00490090099009.
- [2] R. Govindarajan, The Implication of Diagnostic Errors, *Contin. Lifelong Learn. Neurol.* 2017; 23: 1458–1466. doi:10.1212/CON.0000000000000516.
- [3] M.I. Chimowitz, E.L. Logigian, L.R. Caplan, The accuracy of bedside neurological diagnoses, *Ann. Neurol.* 1990; 28: 78–85. doi:10.1002/ana.410280114.
- [4] Segal MM, Williams MS, Gropman AL, et al. Evidence-based decision support for neurological diagnosis reduces errors and unnecessary workup. *J Child Neurol.* 2014; 29(4):487–492. doi:10.1177/0883073813483365
- [5] Heuer JF, Gruschka D, Crozier TA, et al. Accuracy of prehospital diagnoses by emergency physicians: comparison with discharge diagnosis. *Eur J Emerg Med.* 2012; 19(5):292–296. doi:10.1097/MEJ.0b013e32834ce104
- [6] Campbell WW. Diagnosis and localization of neurologic disease, Chapter 53. In Dejong's *The neurologic examination*. 7th edition. Lippincott Williams and Wilkins, Philadelphia, 2013, pp 769-795.
- [7] Andrews JE, Richesson RL, Krischer J. Variation of SNOMED CT coding of clinical research concepts among coding experts. *J Am Med Inform Assoc.* 2007; Jul-Aug;14(4):497-506.
- [8] Chiang MF, Hwang JC, Yu AC, Casper DS, Cimino JJ, Starren J. Reliability of SNOMED-CT Coding by Three Physicians using Two Terminology Browsers, *AMIA Annu Symp Proc.* 2006: 131–135.
- [9] Bhattacharyya SB. *Introduction to SNOMED CT*. Singapore: Springer, 2016.
- [10] Bodenreider O. The Unified Medical Language System (UMLS): integrating biomedical terminology, *Nucleic Acids Research*, 2004; 32, Issue suppl_1, Pages D267–D270, <https://doi.org/10.1093/nar/gkh061>.
- [11] Hier, D.B., Brint, S.U. A Neuro-ontology for the neurological examination. *BMC Med Inform Decis Mak* 2020; **20**: 47. <https://doi.org/10.1186/s12911-020-1066-7>
- [12] Jaccard P. The Distribution of the flora in the alpine zone, *New Phytologist*, 1912; 11: 37–50, doi:10.1111/j.1469-8137.1912.tb05611.x

- [13] Haase P., Siebes R., van Harmelen F. Peer Selection in Peer-to-Peer Networks with Semantic Topologies. In: Bouzeghoub M., Goble C., Kashyap V., Spaccapietra S. (eds) *Semantics of a Networked World. Semantics for Grid Databases. ICSNW 2004. Lecture Notes in Computer Science*, 2004; vol 3226. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-30145-5_7
- [14] Girardi D, Wartner S, Halmerbauer G, Ehrenmüller M, Kosorus H, Dreiseitl S. Using concept hierarchies to improve calculation of patient similarity, *J. Biomed. Inform.* 2016; 63: 66–73. doi:10.1016/j.jbi.2016.07.021.
- [15] Mabotuwana T, Lee MC, Cohen-Solal EV. An ontology-based similarity measure for biomedical data-application to radiology reports. *J Biomed Inform.* 2013;46(5):857–868. doi:10.1016/j.jbi.2013.06.013
- [16] Wu Z, Palmer M. Verb semantics and lexical selection. In *Proceedings of the 32nd Annual Meeting of the Associations for Computational Linguistics*, 1994; pp 133-138.
- [17] Rada R, Hafedh M, Bicknell E, Blettner M. Development and Application of a Metric on Semantic Nets. *IEEE Transactions on Systems, Man, and Cybernetics.* 1989; 19(1): 17-30.
- [18] Leacock C, Chodorow M. Combining Local Context and WordNet Similarity for Word Sense Identification, *WordNet.* 1998; doi:10.7551/mitpress/7287.003.0018.
- [19] Resnik P. Using Information Content to Evaluate Semantic Similarity in a Taxonomy. 1995; <http://arxiv.org/abs/cmp-lg/9511007>.
- [20] Jiang JJ, Conrath DW. Semantic Similarity Based on Corpus Statistics and Lexical Taxonomy. In *Proceedings of International Conference Research on Computational Linguistics (ROCLING X)*, 1997; Taiwan, pp 19-33, <https://www.aclweb.org/anthology/O97-1002>.
- [21] Lin D. An Information-Theoretic Definition of Similarity, *ICML 1998 Proceedings of the Fifteenth International Conference on Machine Learning.* Pages 296-304, July 24 - 27, 1998.
- [22] Lee W, Shah N, Sundlass K, Musen M. Comparison of Ontology-based Semantic-Similarity Measures. *Medical College of Wisconsin, Milwaukee, WI, Symp. A Q. J. Mod. Foreign Lit.* 2008; 384–388.

- [23] McInnes B, Pedersen T, Pakhomov S. UMLS-interface and UMLS-similarity: open source software for measuring paths and semantic similarity. In: Proceedings of the American Medical Informatics Association (AMIA) Symposium, San Francisco, CA. 2009; p. 431–5.
- [24] Pedersen T, Pakhomov S, Patwardhan C, Chute C. Measures of semantic similarity and relatedness in the biomedical domain. *J Biomed Inform*, 2007; 40 (3), pp. 288-299.
- [25] Caviedes JE, Cimino JJ. Towards the development of a conceptual distance metric for the UMLS, *J. Biomed. Inform.* 2004; 37: 77–85. doi:10.1016/j.jbi.2004.02.001.
- [26] McInnes BT, Pedersen T. Evaluating semantic similarity and relatedness over the semantic grouping of clinical term pairs *J. Biomed. Inform.* 2015; 54: 329–336. doi:10.1016/j.jbi.2014.11.014.
- [27] Jia Z, Lu X, Duan H, Li H. Using the distance between sets of hierarchical taxonomic clinical concepts to measure patient similarity, *BMC Med. Inform. Decis. Mak.* 2019; 19: 1–11. doi:10.1186/s12911-019-0807-y.
- [28] Blumenfeld H. *Neuroanatomy through Clinical Cases*, 2nd Edition. Sunderland, MA: Sinauer Associates, 2010.
- [29] Macleod M, Simpson M, Pal S. *Clinical Cases Uncovered: Neurology*. Wiley-Blackwell, West Sussex UK, 2011.
- [30] Noseworthy JH. *Fifty Neurologic Cases from Mayo Clinic*. Oxford University Press, Oxford UK, 2004.
- [31] Pendlebury ST, Anslow P, Rothwell PM. *Neurological case histories*, Oxford University Press, Oxford UK, 2007.
- [32] Toy EC, Simpson E, Mancias P, Furr-Stimming EE. *Case Files Neurology*, 3rd edition. McGraw-Hill, New York, 2018.
- [33] Waxman SG. *Clinical Neuroanatomy*. 28th Edition. McGraw Hill, New York, 2017.
- [34] Hauser SL, Levitt LP, Weiner HL. *Case Studies in Neurology for the House Officer*. Williams and Wilkins, Baltimore, 1986.
- [35] Liveson JA, Spielholz N. *Peripheral neurology: case studies in electrodiagnosis*. Philadelphia, FA Davis Company. 1979.

- [36] Gauthier SG, Rosa-Netto P. Case studies in dementia. Cambridge, Cambridge University Press, 2011.
- [37] Erro R, Stamelou M, Bhatia K. Case studies in movement disorders. Cambridge, Cambridge University Press, 2017.
- [38] Solomon T, Michael BD, Miller A, Kneen R. Case studies in neurological infections of adults and children. Cambridge, Cambridge University Press, 2019.
- [39] Howard J, Singh A. Neurology image-based clinical review. New York, Demos Publishing, 2017.
- [40] Committee on Diagnostic Error in Health Care; Board on Health Care Services; Institute of Medicine; The National Academies of Sciences, Engineering, and Medicine; Balogh EP, Miller BT, Ball JR, editors. Improving Diagnosis in Health Care. Washington (DC): National Academies Press (US); 2015 Dec 29. Available from: <https://www.ncbi.nlm.nih.gov/books/NBK338596/> doi: 10.17226/21794
- [41] Schiff GD, Hasan O, Kim S, et al. Diagnostic Error in Medicine: Analysis of 583 Physician-Reported Errors. *Arch Intern Med.* 2009;169(20):1881–1887. doi:10.1001/archinternmed.2009.333
- [42] Shinar D, Gross CR, Mohr JP, et al. Interobserver variability in the assessment of neurologic history and examination in the Stroke Data Bank. *Arch Neurol.* 1985;42(6):557–565. doi:10.1001/archneur.1985.04060060059010
- [43] Savova GK, Masanz JJ, Ogren PV, Zheng J, Sohn S, Kipper-Schuler KC, Chute CG, Mayo clinical Text Analysis and Knowledge Extraction System (cTAKES): Architecture, component evaluation and applications. *J. Am. Med. Informatics Assoc.* 2010; 17: 507–513. doi:10.1136/jamia.2009.001560.
- [44] Aronson AR, Lang FM, An overview of MetaMap: Historical perspective and recent advances. *J. Am. Med. Informatics Assoc.* 2010; 17: 229–236. doi:10.1136/jamia.2009.002733.
- [45] Reátegui R, Ratté S. Comparison of MetaMap and cTAKES for entity extraction in clinical notes, *BMC Med. Inform. Decis. Mak.* 2018; 18. doi:10.1186/s12911-018-0654-2.
- [46] Zheng K, Vydiswaran VGV, Liu Y, Wang Y, Stubbs A, Uzuner Ö, Gururaj AE, Bayer S, Aberdeen J, Rumshisky A, Pakhomov S, Liu H, Xu H. Ease of adoption of

clinical natural language processing software: An evaluation of five systems, *J. Biomed. Inform.* 2015; 58: S189–S196. doi:10.1016/j.jbi.2015.07.008.

[47] Kreimeyer K, Foster M, Pandey A, Arya N, Halford G, Jones SF, Forshee R, Walderhaug M, Botsis T. Natural language processing systems for capturing and standardizing unstructured clinical information: A systematic review, *J. Biomed. Inform.* 2017; 73: 14–29. doi:10.1016/j.jbi.2017.07.012.

[48] Velupillai S, Suominen H, Liakata M, Roberts A, Shah AD, Morley K, Osborn D, Hayes J, Stewart R, Downs J, Chapman W, Dutta R. Using clinical Natural Language Processing for health outcomes research: Overview and actionable suggestions for future advances, *J. Biomed. Inform.* 2018; 88: 11–19. doi:10.1016/j.jbi.2018.10.005.

[49] Jana N, Barik S, Arora N. Current use of medical eponyms--a need for global uniformity in scientific publications. *BMC Med Res Methodol.* 2009;9:18. Published 2009 Mar 9. doi:10.1186/1471-2288-9-18

Table 1. Four diagnosis tranches and forty diagnosis classes used in analyses. The first column is an abbreviation used in Tables and Figures. Typical findings are listed illustratively for non-neurologists and are not meant to be a definitive reference on each condition.

	Tranche*	Typical Findings
	Nerve and muscle tranche	
ULN	ulnar neuropathy	weakness, pain, sensory loss
MED	median neuropathy	pain, sensory loss
GBS	Guillain Barré syndrome	weakness, areflexia, sensory loss, paresthesias
B12	B ₁₂ deficiency	paresthesias, confusion, weakness, sensory loss
MYO	myopathy	weakness
MG	myasthenia gravis	weakness, diplopia, ptosis
CR	cervical radiculopathy	weakness, sensory loss, reflex change, pain
LR	lumbar radiculopathy	weakness, sensory loss, reflex change, pain
PN	polyneuropathy	weakness, sensory loss, hyporeflexia
MP	meralgia paresthetica	pain, sensory loss
	Movement disorder tranche	
HD	Huntington disease	chorea, personality change
PAR	Parkinson disease	tremor, bradykinesia, rigidity
PSP	progressive supranuclear palsy	bradykinesia, rigidity, gaze palsies
SND	striatonigral degeneration	bradykinesia, rigidity
NPH	normal pressure hydrocephalus	urinary incontinence, dementia, gait apraxia
ET	essential tremor	tremor
HB	hemiballismus	hemiballismus
DYS	dystonia	dystonia
LBD	Lewy body dementia	dementia, bradykinesia, hallucinations
WIL	Wilson disease	tremor, ataxia, dystonia, bradykinesia, personality change
	Brain and spinal cord tranche	
ALS	amyotrophic lateral sclerosis	weakness, hyperreflexia, fasciculations
ALZ	Alzheimer disease	amnesia, dementia
FTD	frontotemporal dementia	aphasia, dementia, executive dysfunction
HSE	herpes simplex encephalitis	fever, headache, stiff neck, confusion
MEN	meningitis	fever, headache, stiff neck
MS	multiple sclerosis	weakness, sensory changes, hyperreflexia, diplopia
MYL	myelopathy	weakness, sensory level, urinary retention, hyperreflexia
SAH	subarachnoid hemorrhage	headache, stiff neck
SDH	subdural hematoma	headache, lethargy, weakness
	Cranial nerve tranche	
BPV	Benign positional vertigo	vertigo
RH	Ramsay Hunt syndrome	facial weakness, hearing loss
BEL	Bell palsy	facial weakness
DV	Devic disease	blurred vision, weakness, numbness, spasticity, hyperreflexia
IIH	idiopathic intracranial hypertension	blurred vision, papilledema, headache
SIX	sixth nerve palsy	diplopia
THD	third nerve palsy	diplopia, ptosis
AN	acoustic neuroma	tinnitus, hearing loss, nystagmus
ON	optic neuritis	blurred vision, papilledema
TN	trigeminal neuralgia	face pain

*The non-possessive form of eponymous diseases has been used uniformly [49]

Table 2. Performance of k-nearest neighbor classifier by diagnosis

Tranche*			
	Nerve and muscle tranche	N	Recall**
MP	meralgia paresthetica	4	100
ULN	ulnar neuropathy	8	100
MED	median neuropathy	10	90
CR	cervical radiculopathy	10	90
LR	lumbar radiculopathy	15	87
GBS	Guillain Barré syndrome	16	81
B12	B ₁₂ deficiency	5	80
MG	myasthenia gravis	13	77
MYO	myopathy	12	75
PN	polyneuropathy	14	64
Movement disorder tranche			
HB	hemiballismus	4	100
PAR	Parkinson disease	16	100
NPH	normal pressure hydrocephalus	11	100
HD	Huntington disease	12	92
PSP	progressive supranuclear palsy	6	83
ET	essential tremor	5	80
DYS	dystonia	5	80
LBD	Lewy body dementia	5	40
WIL	Wilson disease	5	40
SND	striatonigral degeneration	4	25
Brain and spinal cord tranche			
MYL	myelopathy	21	100
ALS	amyotrophic lateral sclerosis	16	88
SAH	subarachnoid hemorrhage	10	80
SDH	subdural hematoma	10	70
ALZ	Alzheimer disease	13	69
CJD	Creutzfeldt-Jacob disease	9	67
MEN	meningitis	17	65
FTD	frontotemporal dementia	11	64
MS	multiple sclerosis	14	50
HSE	herpes simplex encephalitis	10	40
Cranial nerve tranche			
AN	acoustic neuroma	8	100
BEL	Bell palsy	7	100
BPV	Benign positional vertigo	8	100
IIH	idiopathic intracranial hypertension	9	100
SIX	sixth nerve palsy	4	100
TN	trigeminal neuralgia	7	100
ON	optic neuritis	4	75
THD	third nerve palsy	7	71
DV	Devic disease	3	67
RH	Ramsay Hunt syndrome	4	0

*The non-possessive form of eponymous diseases has been used uniformly [49]

**Recall = true positives/(true positives + false negatives)

Table 3. Performance of k-nearest neighbor classifier by tranche with stratified 10-fold cross-validation, k = 3, and distance metric set to uniform weighting.

Tranche	Accuracy	F1	Recall	Precision	N
brain and spinal cord	0.71	0.71	0.71	0.71	131
cranial nerve	0.87	0.84	0.87	0.83	61
movement disorder	0.81	0.78	0.81	0.84	74
nerve and muscle	0.81	0.81	0.81	0.83	107

Recall = true positives/(true positives + false negatives)

Precision = true positives/ (true positives + false positives)

Accuracy = (true positives + true negatives) / (true positives + true negatives + false negatives +false positives)

F1 = harmonic mean of precision and recall

Figure 1. Representation of patient A with absent reflex and aphasia and patient B with absent ankle reflex, absent knee reflex, and alexia as two sets. The only finding in common is absent ankle reflex.

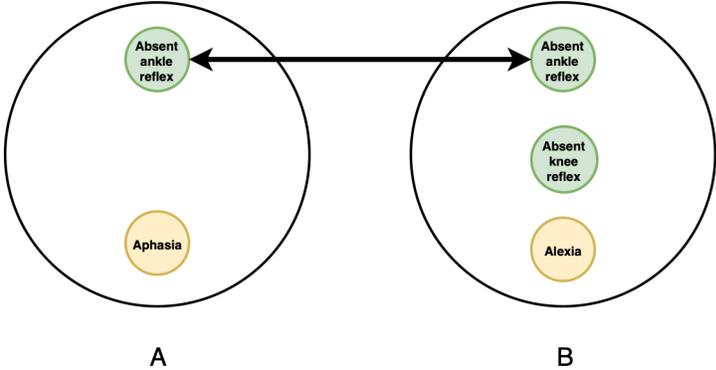


Figure 2. Inter-patient distances between a single patient with multiple sclerosis compared to the other 720 patients in the dataset when calculated by either Jaccard distance or semantic weighted bipartite distance. Each filled circle or square represents a single patient. The thin bar is group mean. The dark bar at the top of the Jaccard scatter plot reflects the observation that for 566 of the 720 patient distances, the value was 1.0 for the Jaccard distance. Better dispersion of distances is noted with semantic weighted distance (blue filled circles). Means differ, two tailed t-test, $dF = 1365$, $p < .0001$

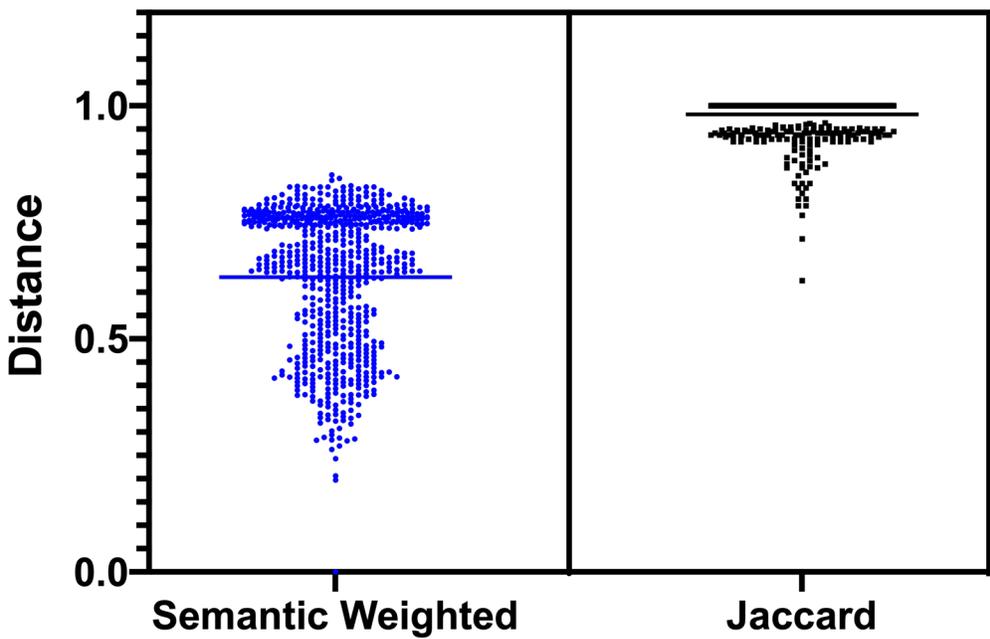


Figure 3. Box plots (minimum, mean, and maximum) for intra-class distances for patients from movement disorder tranche and brain and spinal cord tranche. Intra-class means are arranged in ascending order. See Table 1 for abbreviations. Means differ, one-way ANOVA, $dF=19$, $p < 0.0001$

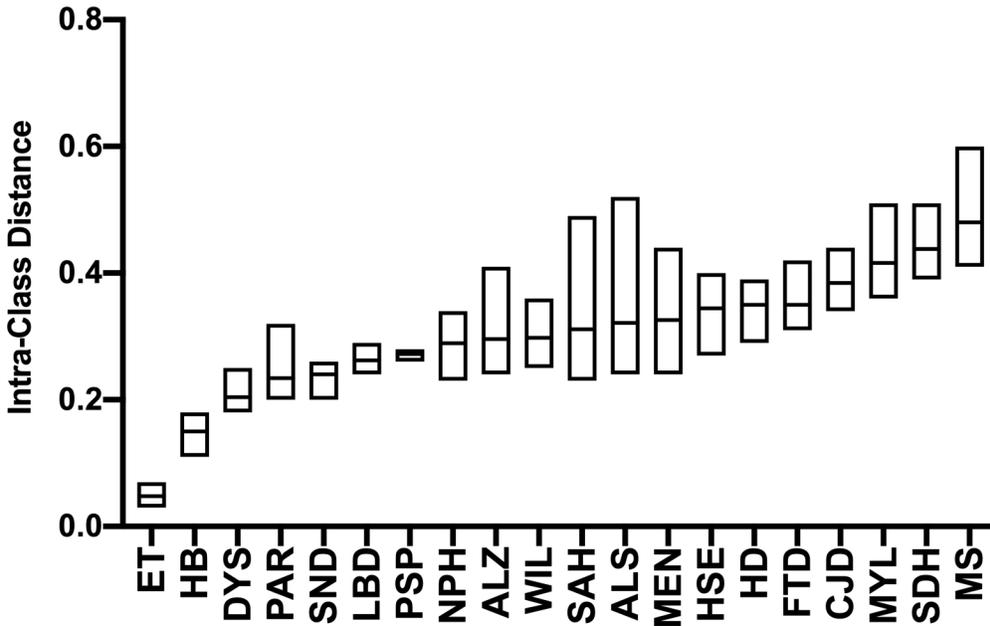


Figure 4. Box plots (minimum, mean, and maximum) for intra-class distances for patients from cranial nerve tranche and muscle and nerve tranche. Intra-class means are arranged in ascending order. See Table 1 for abbreviations. Means differ, one-way ANOVA, $dF = 19$, $p < 0.0001$

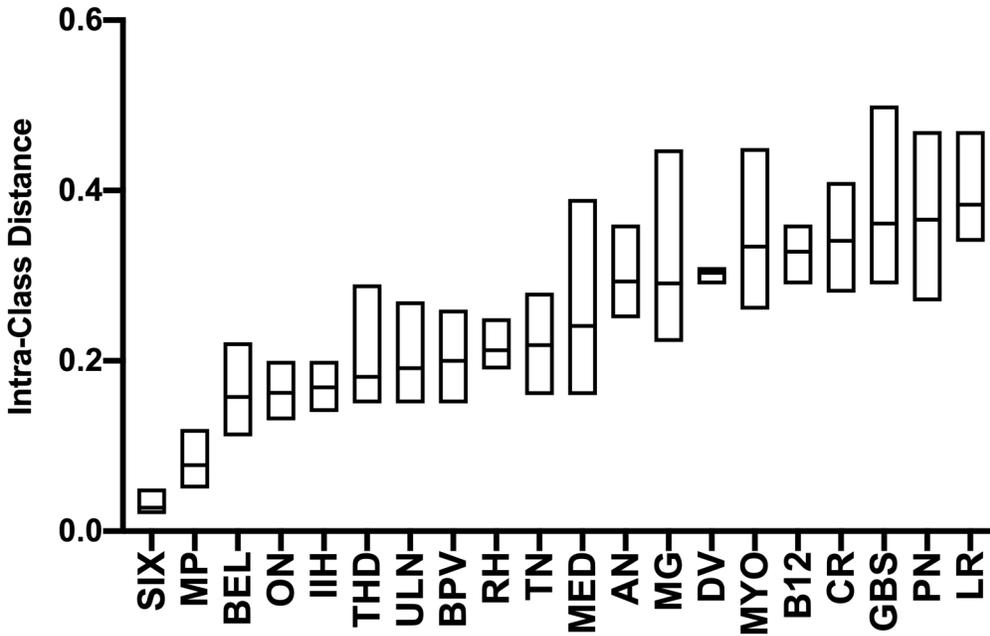


Figure 5. Nerve and muscle tranche MDS. Note that meralgia paresthetica (MP), which is entirely sensory, is to the far right and myasthenia gravis (MG), which is entirely motor is to far left. See Table 1 for abbreviations.

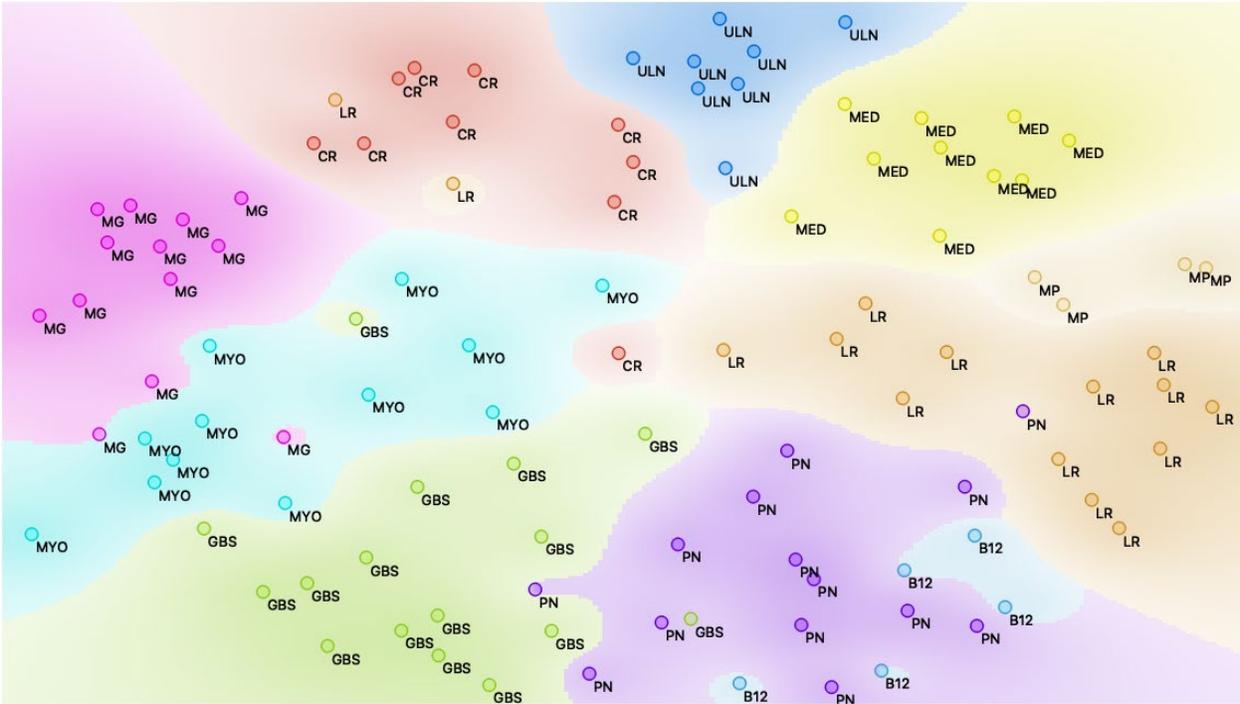


Figure 6. Brain and spinal cord tranche MDS. Note proximity of frontal-temporal dementia (FTD) to Alzheimer disease (ALZ) and proximity of amyotrophic lateral sclerosis (ALS) to multiple sclerosis (MS) and myelopathy (MYL). See Table 1 for abbreviations.

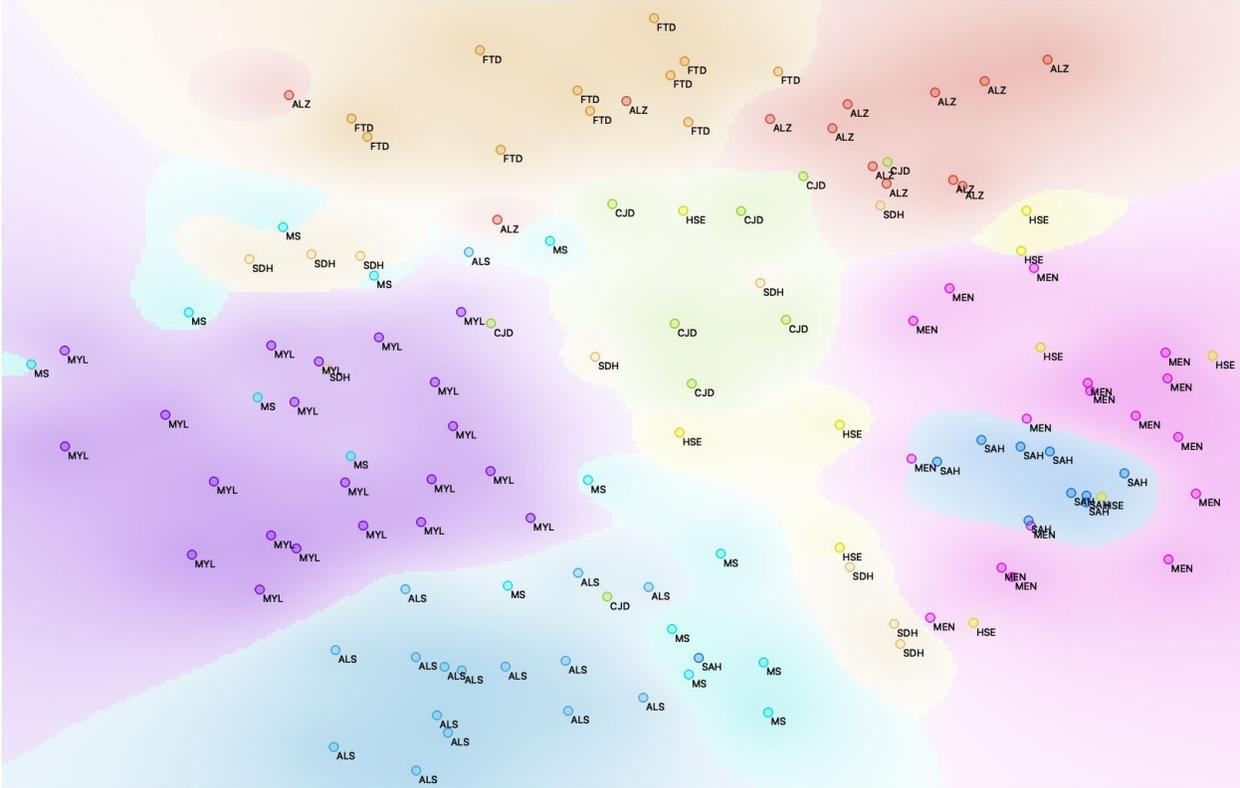


Figure 7. Movement disorder tranche MDS. Note that bradykinetic movement disorders (PAR, PSP, and SND) are to right on the diagram and that hyperkinetic movement disorders (HB, HD, and DYS) are to the left on the diagram. See Table 1 for abbreviations.

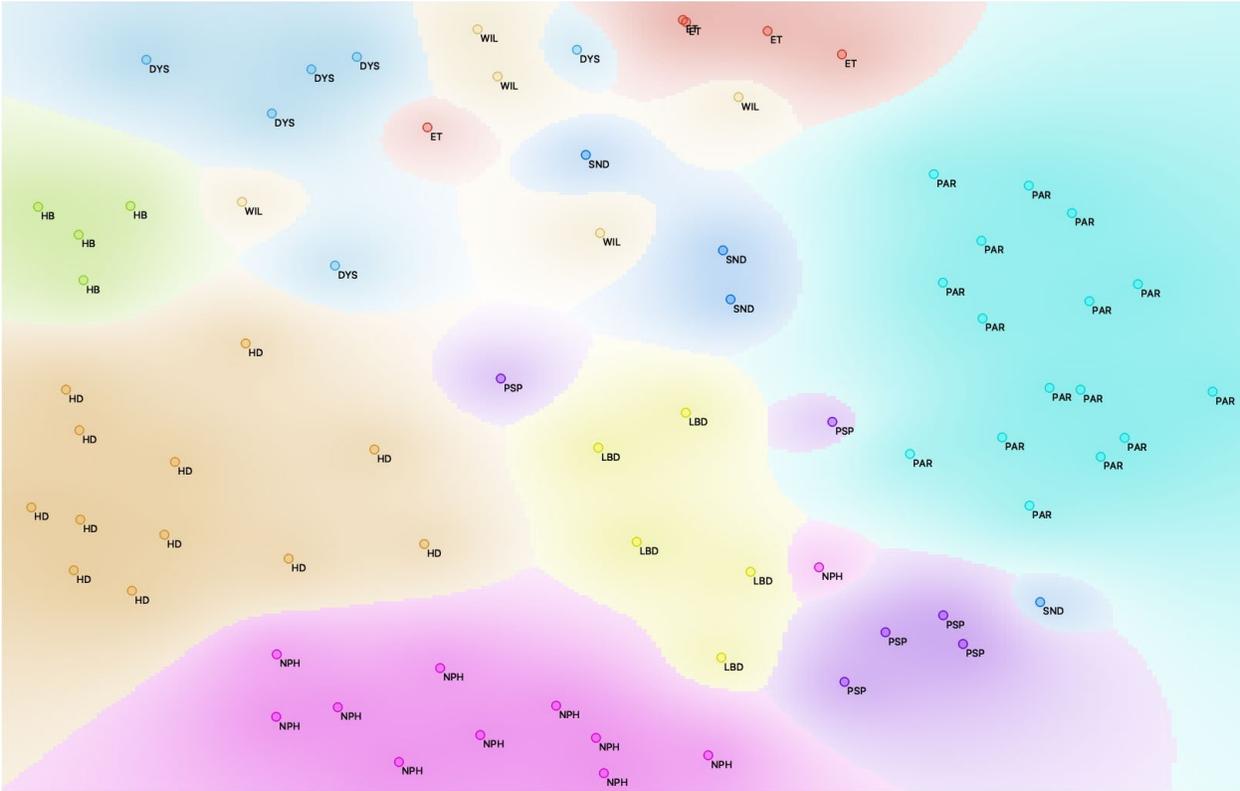


Figure 8. MDS for the cranial nerve tranche. Note that disorders of the visual system (IIH, THD, SIX, ON) are clustered on the right side of the diagram; disorders of acoutico-vestibular nerve (AN, RH, BEL BPV) are clustered on the left side of the diagram. See Table 1 for abbreviations.

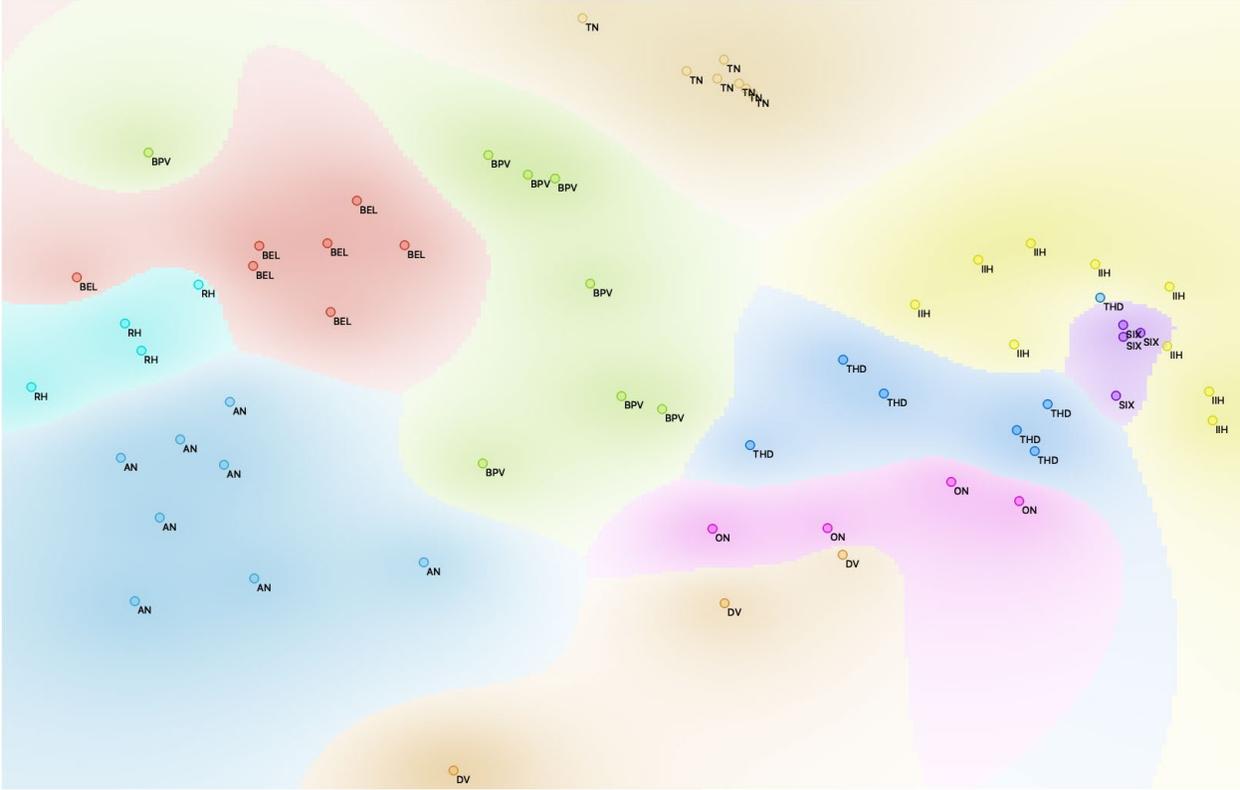


Figure 9. Performance of k-nearest neighbor classifier related to k and stratified with 10-fold cross-validation with uniform distance metric. Accuracy is shown separately by tranche. Means differ, one-way ANOVA, $dF=3$, $p < 0.0001$

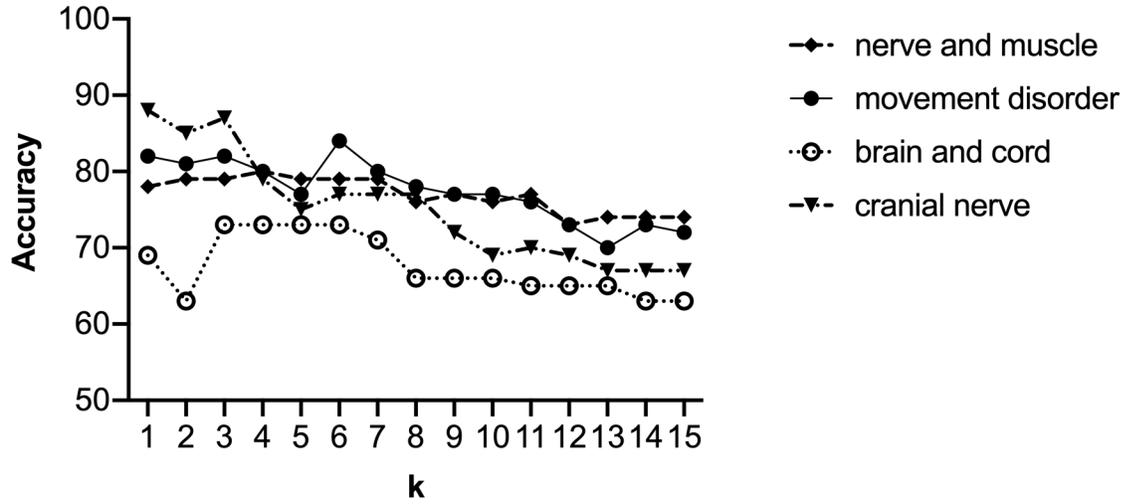


Figure 10. Distance from amyotrophic lateral sclerosis (ALS) class to other classes in the brain and spinal cord tranche. Distances are from patients in the ALS class. Solid bars are class means. Each filled circle is a single patient observation. All filled black circles are distances to the filled blue circles (ALS). MYL (myelopathy) and SDH (subdural hematoma) are the two closest classes to ALS, followed by MS (multiple sclerosis).

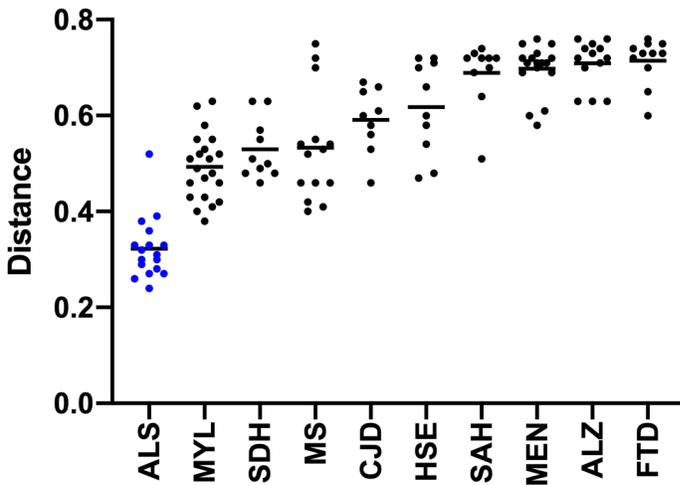


Figure 11. Distance from Parkinson (PAR) class to other classes in movement disorder tranche. Distances are from patients in the PAR class. Solid bars are class means. Each filled circle is a single patient observation. All filled black circles are distances to the filled blue circles (PAR). SND (striatonigral degeneration) and ET (essential tremor) are the two closest classes to Parkinson's disease, followed by LBD (Lewy body dementia). Means differ, one-way ANOVA, $dF=9$, $p < 0.0001$

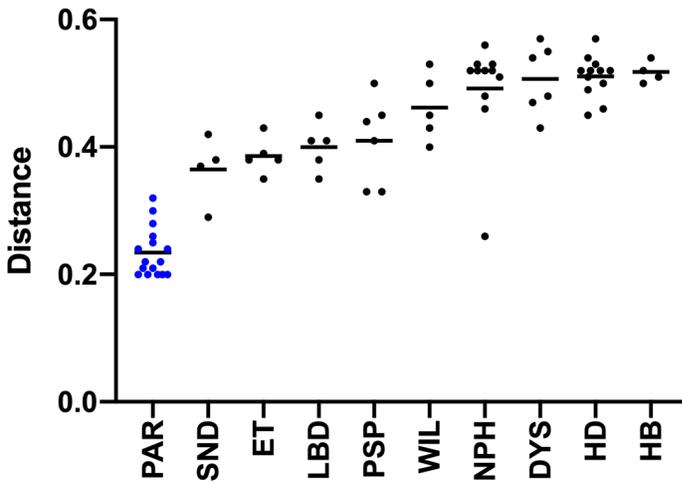


Figure 12. Twenty nearest neighbors to a single patient with amyotrophic lateral sclerosis (ALS). The first patient is the ALS patient, and the distance to self is 0.0. Most of the nearest neighbors have ALS, but a few have myelopathy (MYL), multiple sclerosis (MS), or myopathy (MYO).

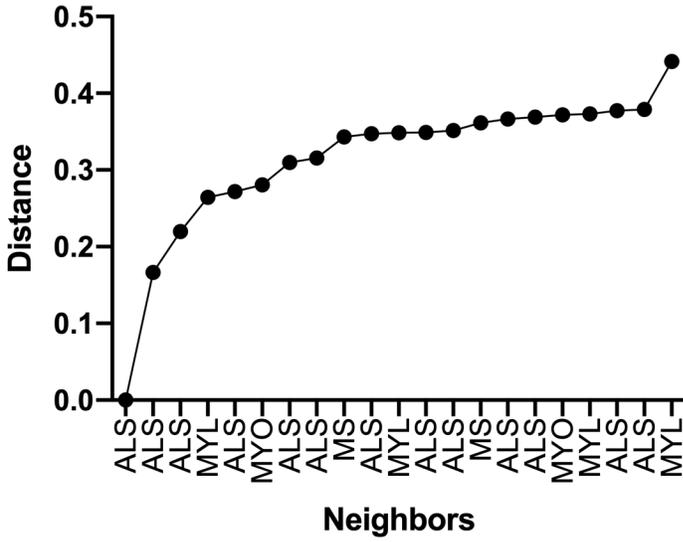


Figure 13. Twenty nearest neighbors to a single patient with Parkinson's disease (PAR). The first patient is the Parkinson disease patient, and distance to self is 0.0. Most of the nearest neighbors have Parkinson disease, but one has normal pressure hydrocephalus (NPH), two have Lewy body dementia (LBD), and two have striatonigral degeneration (SND).

