10 years ESJ
Special edition

Subsumption is a Novel Feature Reduction Strategy for High Dimensionality Datasets

Donald C. Wunsch III, MS
Daniel B. Hier, MD
Missouri University of Science and Technology, USA

Doi:10.19044/esj.2022.v18n4p20

Submitted: 10 November 2021
Accepted: 20 January 2022
Published: 08 February 2022

Cite As:
Wunsch D.C. & Hier D.B. (2022). *Subsumption is a Novel Feature Reduction Strategy for High Dimensionality Datasets.* European Scientific Journal, ESJ, 18 (4), 20. https://doi.org/10.19044/esj.2022.v18n4p20

Abstract

High dataset dimensionality poses challenges for machine learning classifiers because of high computational costs and the adverse consequences of redundant features. Feature reduction is an attractive remedy to high dimensionality. Three different feature reduction strategies (subsumption, Relief F, and principal component analysis) were evaluated using four machine learning classifiers on a high dimension dataset with 474 unique features, 20 diagnoses, and 364 instances. All three feature reduction strategies proved capable of significant feature reduction while maintaining classification accuracy. At high levels of feature reduction, the principal components strategy outperformed Relief F and subsumption. Subsumption is a novel strategy for feature reduction if features are organized in a hierarchical ontology.

Keywords: Machine Learning, Feature Reduction, Neurology, Ontology, Principal Components, Relief, subsumption

Introduction And Previous Work

Electronic health records (EHRs) hold vast clinical data (Esteva et al., 2019). Some of the value of this data can be unlocked by machine learning (Xiao, Choi, & Sun, 2018; Miotto, Wang, Wang, Jiang, & Dudley, 2018). Healthcare datasets are of high dimensionality with hundreds or thousands of unique features. Although much hospital data is numerical (e.g., laboratory results), admission notes, progress notes, and discharge summaries are challenging to convert to a computable form. One approach to making the signs and symptoms of patients computable is deep phenotyping. With deep phenotyping, signs and symptoms are converted to concepts from an ontology such as the Human Phenotype Ontology (HPO) (Kohler et al., 2017, 2019; Groza et al., 2015). One application of deep phenotyping is classifying patients into disease categories based on signs and symptoms (Xiao et al., 2018).

Feature selection (dimension reduction) is important to machine learning. Feature selection can improve model accuracy, reduce over-fitting, eliminate irrelevant features, reduce computational costs, and enhance model interpretability (Kuhn, Johnson, et al., 2013; Kuhn & Johnson, 2019). Feature reduction strategies include filter methods, wrapper methods, ensemble methods, principal components analysis, and genetic algorithms (Kuhn et al., 2013; Visalakshi & Radha, 2014; Kuhn & Johnson, 2019; Al-Jabery, Obafemi-Ajayi, Olbricht, & Wunsch II, 2020). Ontologies offer an opportunity for feature reduction due to their hierarchical structure. Most medical ontologies are based on a subsumptive containment hierarchy with classes organized from the more specific to the more general. Each child class inherits properties from its parent class. This inheritance of properties is called subsumption. For example, the child concepts micrographia, masked face, impaired turns, decreased arms swing, reduced blink rate inherit the concept of slowness of movement from their parent concept *bradykinesia* (Fig. 1). Fine tremor, resting tremor, action tremor, postural tremor, voice tremor, senile tremor inherit tremor from their parent concept *tremor*. The hierarchical structure of ontologies makes an ontology well-suited for feature reduction. We use the term subsumption to describe this novel feature reduction strategy.

Table 1. Codes and counts for neurological diagnoses with their common findings

Code	N	Disease	Common Findings
ALS	23	amyotrophic lateral sclerosis	weakness, spasticity, hyperreflexia, muscle atrophy
ALZ	17	Alzheimer disease	dementia, memory loss, impaired insight, forgetfulness, disorientation
CJD	12	Creutzfeldt Jacob disease	dementia, myoclonus, personality change, confusion, ataxia
FTD	13	fronto-temporal dementia	aphasia, personality change, disinhibition, socially inappropriate behavior
GBS	22	Guillain Barre syndrome	ascending weakness, hyporeflexia, dysautonomia, facial weakness
HD	17	Huntington disease	personality change, chorea, athetosis, confusion, memory loss
HSE	16	herpes simplex encephalitis	confusion, fever, stiff neck, aphasia, disorientation
IIH	14	intracranial hypertension	headache, blurred vision, transient visual obscurations

LR	16	lumbar radiculopathy	foot weakness, pain, absent ankle reflex, sensory loss over foot
MED	16	median nerve neuropathy	sensory loss in hand, pain
MEN	24	meningitis	stiff neck, fever, confusion, headache
MG	18	myasthenia gravis	diplopia, weakness, muscle fatigue, eyelid ptosis
MS	24	multiple sclerosis	spasticity, hyperreflexia, weakness, optic neuritis
MYL	35	myelopathy	sensory level, weakness, sphincter dysfunction
MYO	18	myopathy	proximal muscle weakness
NPH	14	normal pressure hydrocephalus	incontinence, dementia, gait apraxia
PAR	20	Parkinson disease	bradykinesia, cogwheel rigidity, resting tremor
PN	19	polyneuropathy	sensory loss, hyporeflexia, distal weakness
PSP	9	progressive supranuclear palsy	bradykinesia, poor upgaze, confusion
SAH	17	subarachnoid hemorrhage	headache, stiff neck, confusion, nausea, vomiting

This paper examines the ability of subsumption to reduce the number of features in a high dimension dataset and compares it to Relief F and principal components analysis. Subsumption uses the hierarchical structure of an ontology to collapse narrowly defined features into broadly defined features. Relief F uses a distance metric to identify the best features that discriminate between cases of different classes. Principal components analysis creates new features from a linear weighted combination of existing features.

Methods

Overview

The effects of different feature selection strategies on the accuracy of several machine learning classifiers were studied. We specifically examined subsumption, Relief F, and principal components analysis as feature reduction strategies. Four classifiers were tested (trees, SVM, kNN, and a multilayer perceptron neural network). Six lower dimension datasets were constructed by feature selection (ranging from 11 to 464 features). Each dataset was split 80:20 into training and test sets. Validation accuracy was calculated by 5-fold cross-validation on the training set. Test accuracy was based on the withheld test dataset. Mean classification accuracy ± SD of 10 trials per classifier was calculated.

Test dataset

The test dataset consisted of 364 cases distributed between 20 different neurological diseases (Table 1). All cases were derived from 11 standard textbooks of neurology (Bhatia and Erro, and Stamelou, 2017; Noseworthy, 2004; Gauthier & Rosa-Neto, 2011; Blumenfeld, 2010; Howard & Singh, 2016; Toy, Simpson, & Tintner, 2012; Liveson, 2000; Hauser, Weiner, & Levitt, 1986; Solomon, Michael, Miller, & Kneen, 2019; Waxman, 2009; Pendlebury, Anslow, & Rothwell, 2007). For each entry into the dataset, the disease diagnosis was entered as the machine learning label. Symptoms (what the patient complains of) and signs (examination findings by the physician) were abstracted from the case histories and then mapped to one of 1435 concepts in a neuro-ontology by previously described methods (Hier & Brint, 2020; Hier et al., 2020). To capture all the signs and symptoms of the 364 cases, 474 unique concepts were needed. Cases were represented as 475-dimension vectors. The first element of the vector was the label (disease diagnosis), followed by 474 features (signs and symptoms). Features were binarized. The test dataset was a 364 (cases) x 475 (label + features) matrix. There was an average of 11.2 ± 3.5 features per case.

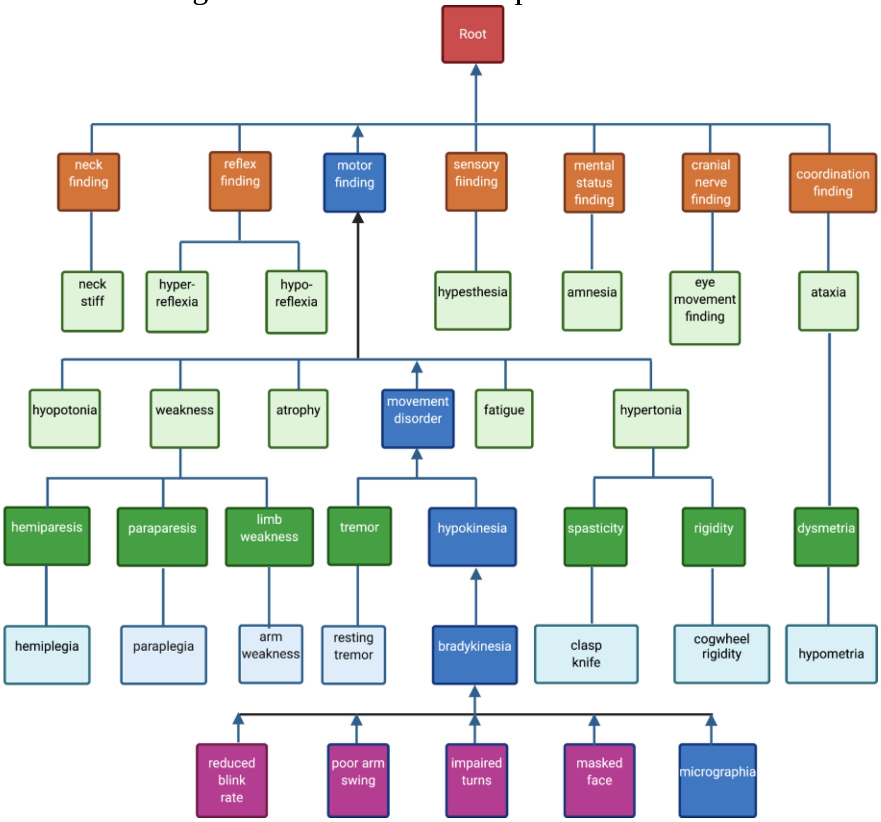

Fig. 1. A small excerpt from the neuro-ontology. The neuro-ontology11 major branches below the root (seven are shown). Concepts in the ontology become increasingly specific at lower levels going from coarsest (least specific) to most granular (most specific) at the lowest level. The concept micrographia (dark blue) is most specific and subsumed by bradykinesia, then hypokinesia, then movement disorder, and finally by the least specific concept motor finding. Each color represents a different level in the concept hierarchy.

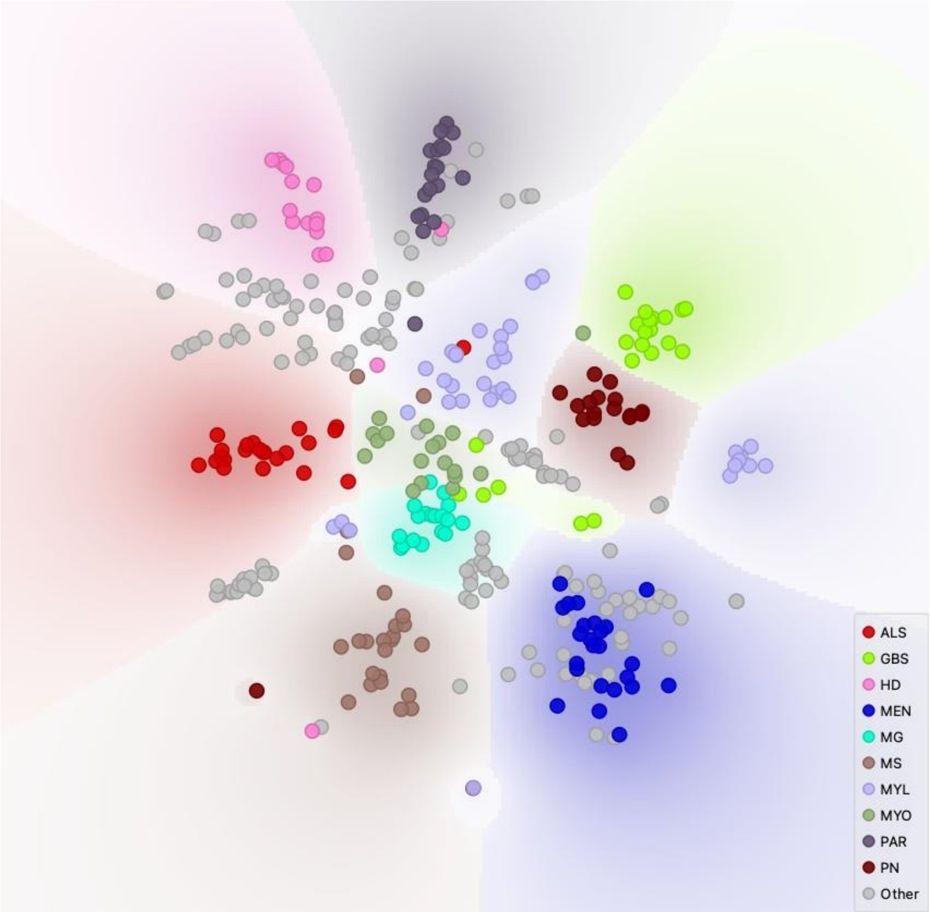

Fig. 2. A t-SNE map (Demsar et al., 2013) illustrates the complexity of the diagnosis classification task. The t-SNE is based on the 20 diagnoses, with each diagnosis shown as a different color. All 364 cases were mapped based on the entire 474 feature set. Although distinctive clusters by disease label are seen, note that there is overlap between diagnoses. For codes to abbreviations and common symptoms of each disease, see Table 1.

Feature reduction by subsumption

Features in the test dataset are concepts from a neuro-ontology (Hier Brint, 2020). The neuro-ontology is a subsumptive hierarchy that supports *IS-A* relationships. Subsumption can repeatedly reduce the number of features by consolidating child concepts with parent concepts. Since the neuro-ontology is at most eight levels deep, we had a potential of eight steps of subsumption for feature reduction. However, some branches of the neuro-ontology were only 3 or 4 levels deep. We used Python to traverse the neuro-ontology (Hier & Brint, 2020) from each of its 1435 terminal nodes to the root node (Fig. 1). We created 1435 ordered lists (one for each concept) of length n=8 where the last element in the list was the penultimate concept (last node before root node), and the first element in the list was the terminal concept. If the list was less than eight elements long, it was back-filled to 8 elements by repeating the first element (terminal node) until all lists were equal in length. Using the ordered lists as a reference, eight new datasets were created by the serial selection of features from the lists eight times. Two of the new datasets showed a minimal reduction in features and were eliminated from the analysis. The remaining six datasets had 11, 76, 245, 360, 424, and 464 features respectively.

Feature reduction by principal components

Principal components analysis (PCA) is a popular multivariate statistical technique for feature reduction that creates new linear combinations of existing variables (Hotelling, 1933). The goal of PCA is to reduce the number of features while retaining as much information as possible (Ringner, 2008; Abdi & Williams, 2010). With PCA, the original variables are replaced with a smaller number of variables called factor scores (weighted linear combinations of the actual variables). Factor scores were calculated by the factor analysis module of SPSS 28 (IBM Corporation, Chicago IL) with extraction by principal components, rotation by Varimax, and Kaiser normalization to create new datasets with 11, 76, 245, 360, 424, and 464 features to parallel the dimensionality of the subsumption datasets.

Feature reduction by Relief F

The Relief algorithm for feature selection was described by Kira & Kendell (1992) and later modified as Relief F by Kononenko et al. (1997). Relief F is a filter-based method that evaluates each feature independently of other features. The algorithm is based on finding index cases in the dataset and then examining matching nearest neighbors (hits) and non-matching neighbors (misses). It uses a difference function to find features that best distinguish the hits from the misses. The Relief F ranking widget from Orange data mining (Demsar et al., 2013) was used to create six subsets of 11, 76, 245, 360, 424, and 464 features respectively.

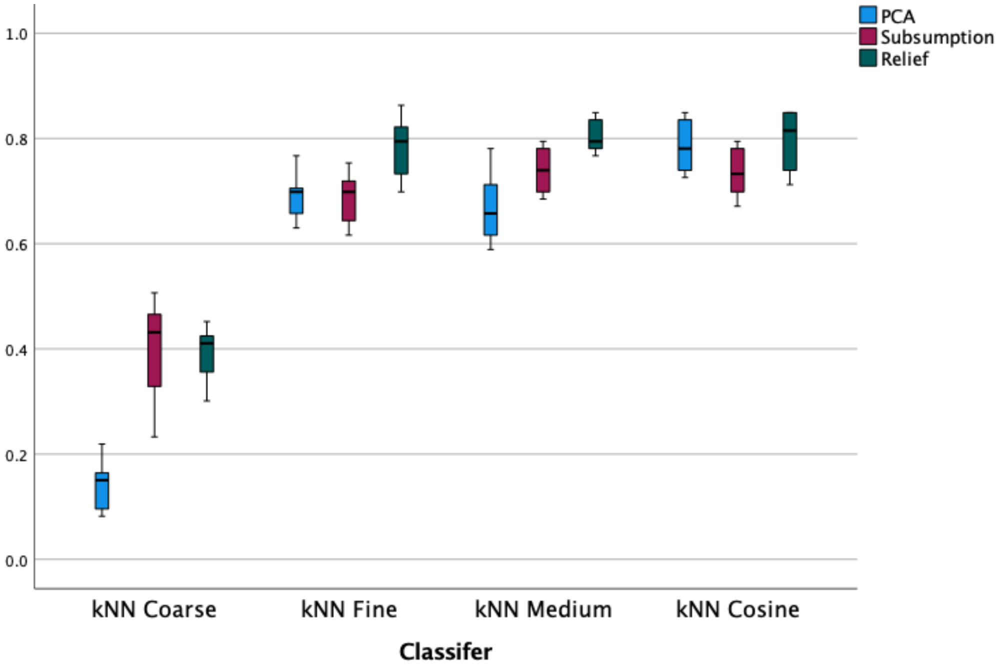

Fig. 3. Comparative accuracy of four different kNN classifiers utilizing 76 features. The four kNN classifiers performed similarly except for the coarse kNN, which performed significantly worse than the other three for all three feature reduction strategies. (One-way ANOVA with post hoc Bonferroni test, p < .05. The cosine kNN classifier was used for subsequent analyses.)

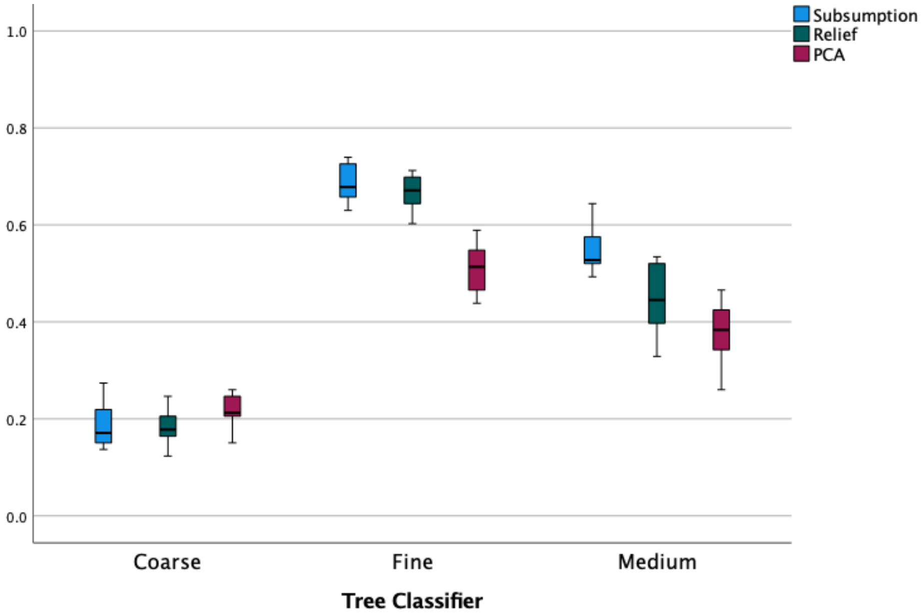

Fig. 4. Comparative accuracy of three different tree classifiers utilizing 76 features. For all three feature reduction strategies, Fine outperformed medium; medium outperformed coarse, one-way ANOVA with post hoc Bonferroni test, p<.05. For additional analyses, the fine tree classifier was used.

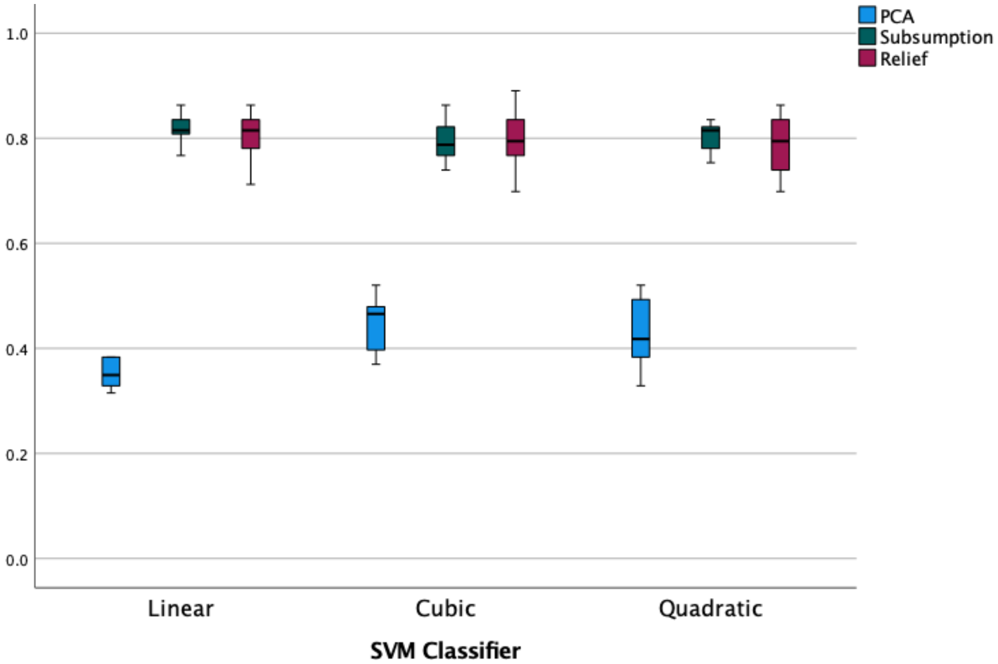

Fig. 5. Comparative accuracy of three different SVM classifiers utilizing 76 features. All three SVM classifiers performed similarly, although performance was lower with the PCA feature reduction strategy (One-way ANOVA, p <0.05). For additional analyses, a linear SVM classifier was selected.)

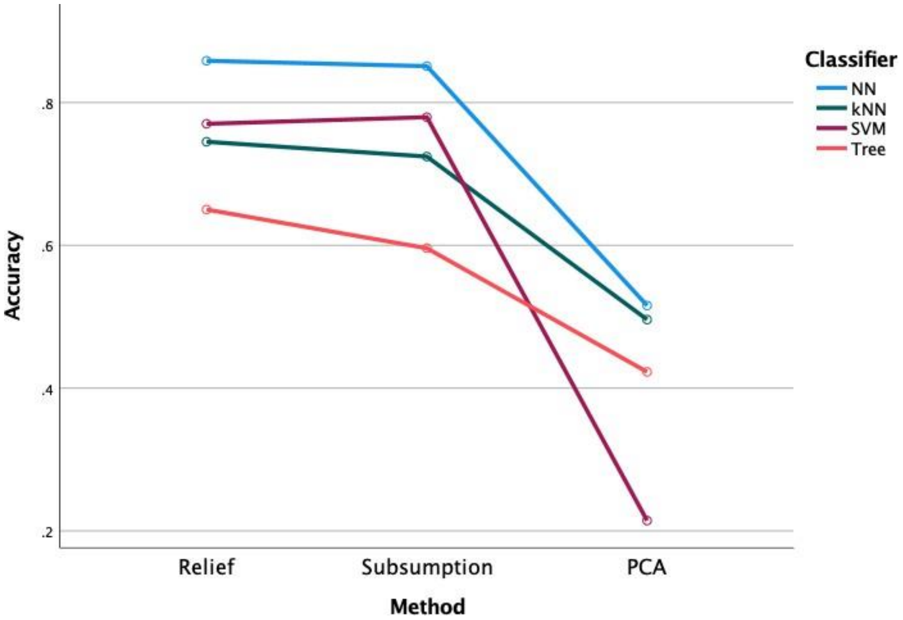

Fig. 6. The average across all feature levels shows that the NN classifier performed best for all three feature reduction strategies (One-way ANOVA, post hoc Bonferroni test, p< 0.05). The low average accuracy for PCA for all classifiers reflects the pooling of high accuracy at low dataset dimensionality with low accuracy at high dimensionality).

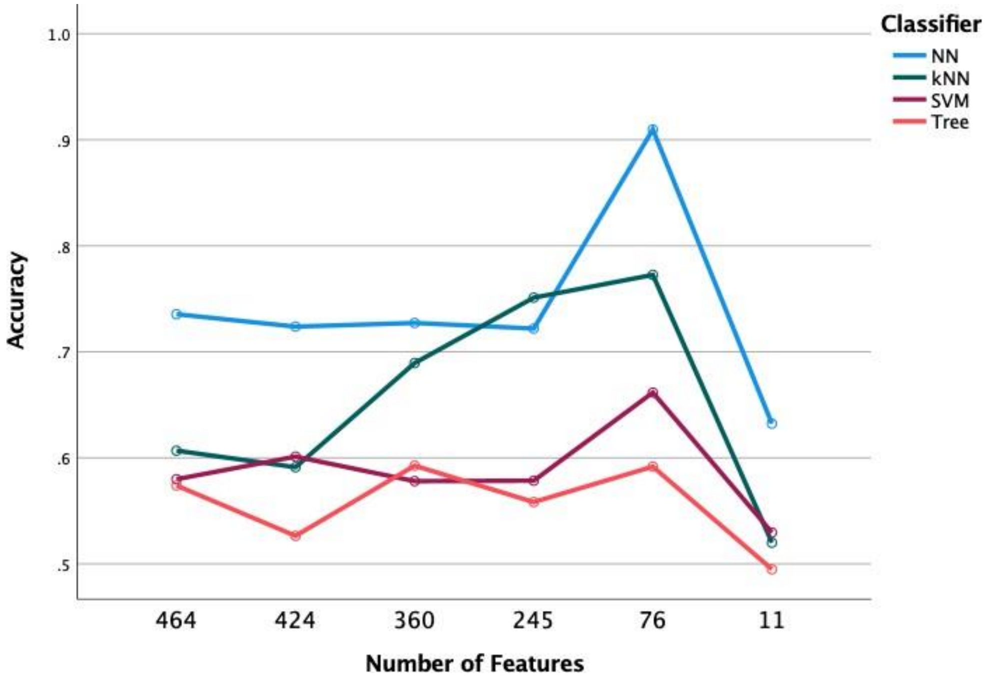

Fig. 7. The NN classifier outperforms the other classifiers at all levels of dataset dimensionality, performing best near 76 features. Results are pooled across all three feature reduction strategies.

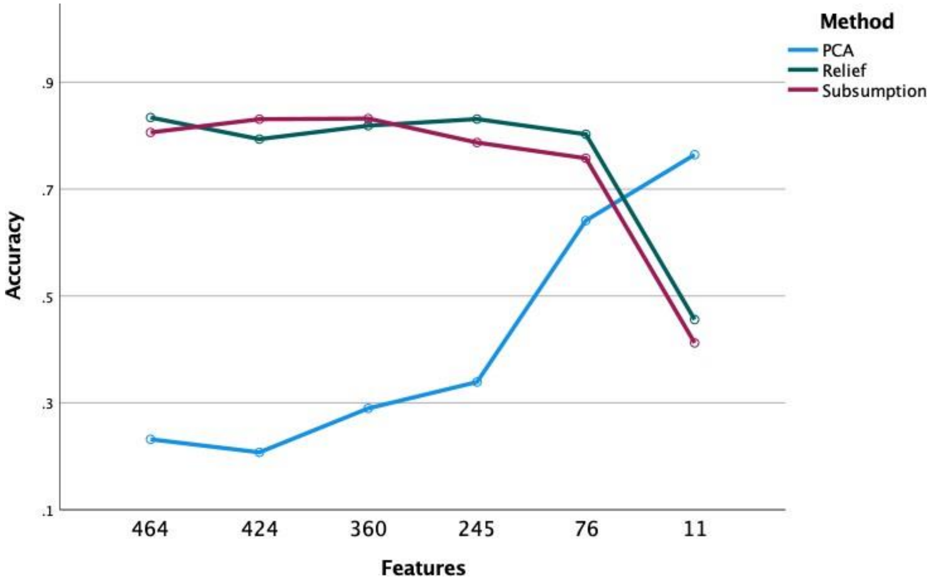

Fig. 8. Results are pooled across all four classifiers. At lower feature levels, the PCA dimension reduction strategy performed best.

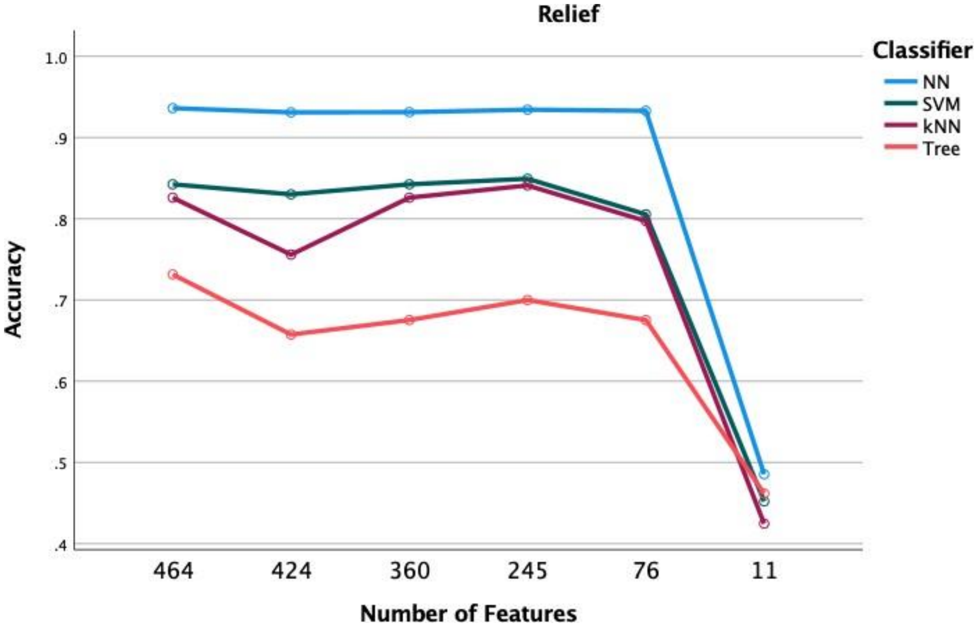

Fig. 9. With feature reduction by Relief accuracy dropped below 76 features. The NN classifier performed best; the tree classifier performed worst.

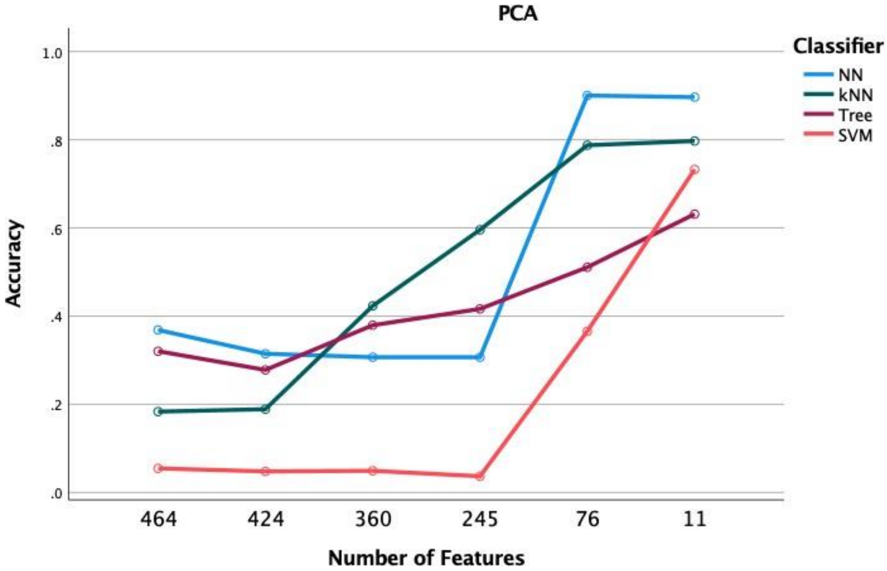

Fig. 10. With the PCA feature reduction strategy, all classifiers performed better at 11 features than a higher number of features. The NN classifier performed best, and SVM performed worst with the PCA strategy.

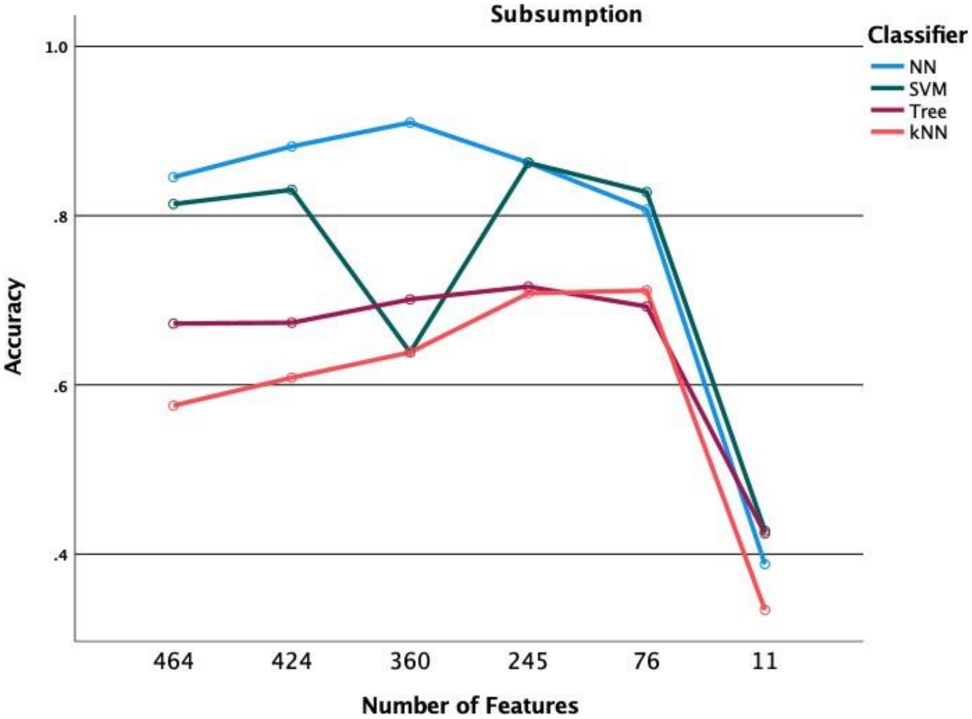

Fig. 11. NN classifier performs best, and tree classifier performs worst with the subsumption strategy. With feature reduction by subsumption, accuracy begins to drop below 76 features.

Machine learning classifiers

MATLAB (MathWorks, Natick MA) was used to construct the k-nearest neighbor (kNN), support virtual machine (SVM), tree, and multilayer perceptron (NN) classifiers. Fine, medium and coarse kNN classifiers with Euclidean distance metrics corresponding to k=1, k=10, and k=100 nearest neighbors were evaluated. A cosine distance kNN classifier with k=10 nearest neighbors was also evaluated. Linear, quadratic, and cubic support vector machine (SVM) classifiers with hyperplane boundaries of orders 1, 2, and 3, respectively, were evaluated (Smola & Scholkopf, 2004). The SVM classifiers use a one-vs-one multi-class classification strategy and standardize predictor data by default. Fine, medium, and coarse classification trees with corresponding thresholds for the maximum number of splits set to 100, 20, and 4, respectively, were evaluated. The default splitting criterion was the Gini Diversity Index (Kingsford & Salzberg, 2008). A multilayer perceptron (NN) of 3 hidden layers was evaluated, each with 500 neurons. Each neuron utilized a hyperbolic tangent transfer function. Output layers used a SoftMax transfer function. The learning rate was set at 0.01 with a momentum constant of 0.1. Each trial of the NN was constrained to a maximum of 300 epochs as a precautionary measure (most trials ran for fewer than 60 epochs). Training ceased after six successive increases in a validation error. Training performance was evaluated by cross-entropy, which consistently yielded higher classification accuracy than a mean-squared error performance metric (De Boer, Kroese, Mannor, & Rubinstein, 2005).

Statistical testing

Based on the 20 possible diagnoses, a 20 x 20 confusion matrix was constructed for each classifier run. Accuracy was the number of cases on the diagonal (match between actual and predicted diagnosis) divided by the total number of cases. When accuracy is calculated based on the entire multi-class confusion matrix, precision, recall, and accuracy are equivalent so that only accuracy is reported (Mohajon, 2020). Average classification accuracy (mean ± SD) was based on ten runs. Group comparisons were made across classifiers, feature reduction strategies, and ontology levels by one-way ANOVA with the F test and a significance level of $p < 0.05$ (SPSS 28, SPSS Inc, Chicago IL). Post hoc means comparisons of individual group means were by the Bonferroni method (Chen, Feng, & Yi, 2017).

Results and Discussion

The features of the dataset are the signs and symptoms of patients with neurological diseases. The labels of the dataset are the disease diagnoses. All features were binarized. The dataset had high dimensionality (474 signs and symptoms) for 364 cases (Table 1). Each classifier was evaluated on a multi-class classification task that involved assigning each of the 364 cases to one of 20 classes (diagnoses) based on the available features. The task was repeated on all 6 data subsets of varying dimensionality. The performance of variations of the kNN, tree, and SVM classifiers were evaluated. All variations of the kNN classifier performed similarly (across all feature reduction strategies) except for the coarse kNN classifier, which performed significantly worse than the others (Fig. 3). This is likely due to the smaller value of k for the coarse classifier. We selected the cosine kNN classifier for subsequent analyses, as it performed best. Fine tree classifiers performed better than medium tree classifiers, and medium trees performed better than coarse trees (Fig. 4). The fine tree classifier was chosen for subsequent analyses. The linear, cubic, and quadratic versions of the SVM classifier performed similarly (Fig. 5). The linear SVM classifier was chosen for subsequent studies.

The neural network (NN) outperformed the other classifiers (Fig. 6, 7, 10). As the number of features in the dataset was reduced from 464 to 11 features, each feature reduction strategy behaved differently. Relief F maintained a high level of classification accuracy until the number of features reached 11, where accuracy dropped significantly (Fig. 8 and Fig. 9). Like Relief F, subsumption maintained a high level of accuracy until the number of features dropped to 11 (Fig. 8 and Fig. 11). PCA classification accuracy improved steadily as features were reduced from 424 to 11 (Fig. 8 and Fig. 10). For both the Relief F strategy and the subsumption strategy, the tree classifier performed least well at all levels of feature reduction compared to the other classifiers (Fig. 9 and Fig. 11).

Validation accuracy was compared to test set accuracy across all classifiers, features, and strategies. Although test set accuracy was lower than validation set accuracy across classifiers, methods, and feature levels, the fall-off was not dramatic, suggesting that significant model overfitting did not occur.

The classification task was to assign one of 20 different labels (diagnoses) to each of the 364 cases based on the features. The features of the dataset were based on a subsumptive containment hierarchy (Hier & Brint, 2020). Subsumption allowed for the successive reduction of the number of features in the dataset from 474 to 11, substituting more general concepts for more specific concepts. Feature reduction by subsumption was compared to Relief F and principal components (PCA).

The goal of feature reduction is to find the minimal subset of features that maintains classifier accuracy and retains predicted class sizes reflective of the class sizes in the ground truth dataset (Dash & Liu, 1997; Tang, Alelyani, & Liu, 2014; Koller & Sahami, 1996). Two commonly used strategies to reduce dataset dimensionality include feature selection and feature extraction. Feature selection (filter methods, wrapper methods) emphasizes algorithms that reduce the number of features into the smallest subset that accurately predicts class membership. Feature extraction methods (principal components, linear discriminant analysis, etc.) emphasize collapsing many features into a smaller number of highly predictive features. The use of subsumption to collapse many features into a smaller number of features bears more resemblance to a feature extraction strategy than a feature selection strategy. Others have suggested using knowledge embedded in a hierarchical ontology as a feature reduction strategy (Corrales, Lasso, Ledezma, & Corrales, 2018). Our results indicate that Relief F, subsumption, and PCA are useful feature reduction strategies across various classifiers. In general, we found that NN, kNN, and SVM classifiers outperformed the tree classifiers (Fig. 7). Importantly, when very high levels of feature reduction are desired, the results suggest that PCA outperforms both Relief F and subsumption (Fig. 8).

This work has limitations. First, the dataset was small, and future testing utilizing a larger dataset will be advantageous. Second, cases were based on textbooks examples rather than actual patient data from electronic health records. Third, class sizes were not perfectly balanced (Table 1). Fourth, due to asymmetries in the depth of the ontology, the subsumption strategy yielded only six different levels of feature reduction (464, 424, 360, 245, 76, and 11 features). Classification accuracy was evaluated only at six dataset dimensions to make fair comparisons between feature reduction strategies. The performance of Relief F or PCA at other levels of dimensionality was not examined, although these strategies could have created additional datasets of varying dimensionality. Other studies have found that when different feature reduction strategies are compared, classifier performance depends on the nature of the dataset, the classifier utilized, and the feature reduction algorithm (Janecek, Gansterer, Demel, & Ecker, 2008). Lastly, we used the default settings for the hyperparameters for the MLP neural network. Additional investigation into fine-tuning the hyperparameters for the machine learning algorithms might improve classification accuracy.

Conclusions

Several conclusions can be drawn from these results.

1) For all classifiers, PCA worked best at lower levels of dimensionality (Fig 8 and Fig. 10). Performance was best at 11 features and dropped at 76 features for tree and SVM and at 245 features for NN and kNN.
2) Classification accuracy using Relief F (Fig. 9) and subsumption (Fig. 11) did not decline until the number of features was reduced below 76. For all classifiers, accuracy was lower for subsumption and Relief F than for PCA at the 11 feature level.
3) Test accuracy was close to validation accuracy across all experiments suggesting that classification models were robust and generalizable.
4) When averaged across all feature reduction strategies, all classifiers performed best at 76 features (Fig. 7).
5) When averaged across all feature reduction strategies, the NN classifier outperformed the SVM, kNN, and tree classifiers (Fig. 6).
6) The results suggest that when features in a high dimension dataset are derived from a subsumptive hierarchical ontology, subsumption is a novel method for feature reduction that does not sacrifice classification accuracy until the highest levels of feature reduction are reached.

Funding

None to report.

Conflicts of interests

None to report.

Human studies

Not applicable.

References:

1. Abdi, H., & Williams, L. J. (2010). Principal component analysis. Wiley interdisciplinary reviews: computational statistics, 2(4), 433–459.
2. Al-Jabery, K. K., Obafemi-Ajayi, T., Olbricht, G. R., & Wunsch II, D. C. (2020). 9 - data analysis and machine learning tools in MATLAB and Python. In K. K. Al-Jabery, T. Obafemi-Ajayi, G. R. Olbricht, & D. C. Wunsch II (Eds.), Computational learning approaches to data analytics in biomedical applications (p. 231-290). Academic Press. DOI: https://doi.org/10.1016/B978-0-12-814482-4.00009-7
3. Bhatia, K.P.and Erro, R. and Stamelou, M. (2017). Case studies in movement disorders. Cambridge University Press.
4. Blumenfeld, H. (2010). Neuroanatomy through clinical cases. Sinauer Associates.
5. Chen, S. Y., Feng, Z., & Yi, X. (2017, Jun). A general introduction to adjustment for multiple comparisons. J Thorac Dis, 9(6), 1725–1729.
6. Corrales, D. C., Lasso, E., Ledezma, A., & Corrales, J. C. (2018). Feature selection for classification tasks: Expert knowledge or traditional methods? Journal of Intelligent & Fuzzy Systems, 34(5), 2825–2835.
7. Dash, M., & Liu, H. (1997). Feature selection for classification. Intelligent data analysis, 1(3), 131–156.
8. De Boer, P.-T., Kroese, D. P., Mannor, S., & Rubinstein, R. Y. (2005). A tutorial on the cross-entropy method. Annals of operations research, 134(1), 19–67.
9. Demsˇar, J., Curk, T., Erjavec, A., Cˇ rt Gorup, Hocˇevar, T., Milutinovicˇ, M., & Zupan, B. (2013). Orange: Data mining toolbox in Python. Journal of Machine Learning Research, 14, 2349-2353. Retrieved from http://jmlr.org/papers/v14/demsar13a.html
10. Esteva, A., Robicquet, A., Ramsundar, B., Kuleshov, V., DePristo, M., Chou, K., .& Dean, J. (2019). A guide to deep learning in healthcare. Nature medicine, 25(1), 24–29.
11. Gauthier, S., & Rosa-Neto, P. (2011). Case studies in dementia: Volume 1: Common and uncommon presentations. Cambridge University Press.
12. Groza, T., Ko¨hler, S., Moldenhauer, D., Vasilevsky, N., Baynam, G., Zemojtel, T., et al. (2015). The human phenotype ontology: semantic unification of common and rare disease. The American Journal of Human Genetics, 97(1), 111–124.
13. Hauser, S., Weiner, H., & Levitt, L. (1986). Case studies in clinical neurology for the house officer. Williams & Wilkins.

14. Hier, D. B., & Brint, S. U. (2020). A neuro-ontology for the neurological examination. BMC Medical Informatics and Decision Making, 20(1), 1–9.
15. Hier, D. B., Kopel, J., Brint, S. U., Wunsch, D. C., Olbricht, G. R., Azizi, S., & Allen, B. (2020). Evaluation of standard and semantically-augmented distance metrics for neurology patients. BMC Medical Informatics and Decision Making, 20(1), 1–15.
16. Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. Journal of educational psychology, 24(6), 417.
17. Howard, J., & Singh, A. (2016). Neurology image-based clinical review. Springer Publishing Company.
18. Janecek, A., Gansterer, W., Demel, M., & Ecker, G. (2008). On the relationship between feature selection and classification accuracy. In New challenges for feature selection in data mining and knowledge discovery (pp. 90–105).
19. Kingsford, C., & Salzberg, S. L. (2008). What are decision trees? Nature biotechnology, 26(9), 1011–1013.
20. Kira, K., & Rendell, L. A. (1992). The feature selection problem: Traditional methods and a new algorithm. In AAAI (Vol. 2, pp. 129–134).
21. Ko¨hler, S., Øien, N. C., Buske, O. J., Groza, T., Jacobsen, J. O., McNamara, C., and others (2019). Encoding clinical data with the human phenotype ontology for computational differential diagnostics. Current protocols in human genetics, 103(1), e92.
22. Ko¨hler, S., Vasilevsky, N. A., Engelstad, M., Foster, E., McMurry, J., Ayme', S. et al. (2017). The human phenotype ontology in 2017. Nucleic acids research, 45(D1), D865 -D876.
23. Koller, D., & Sahami, M. (1996). Toward optimal feature selection (Tech. Rep.). Stanford InfoLab.
24. Kononenko, I., Sˇimec, E., & Robnik-Sˇikonja, M. (1997). Overcoming the myopia of inductive learning algorithms with Relief F. Applied Intelligence, 7(1), 39–55.
25. Kuhn, M., & Johnson, K. (2019). Feature engineering and selection: A practical approach for predictive models. CRC Press.
26. Kuhn, M., Johnson, K., et al. (2013). Applied predictive modeling (Vol. 26). Springer. Liveson, S. A. (2000). Peripheral neurology: case studies. Oxford University Press.
27. Miotto, R., Wang, F., Wang, S., Jiang, X., & Dudley, J. T. (2018). Deep learning for healthcare: review, opportunities, and challenges. Briefings in bioinformatics, 19(6), 1236–1246.

28. Mohajon, J. (2020). Confusion Matrix for Your Multi-Class Machine Learning Model. Towards data science. May 28, 2020. Retrieved at https://towardsdatascience.com/confusion-matrix-for-your-multi-class-machine-learning-model-ff9aa3bf7826
29. Noseworthy, J. H. (2004). Fifty neurologic cases from Mayo Clinic. Oxford University Press.
30. Pendlebury, S., Anslow, P., & Rothwell, P. (2007). Neurological case histories: Case histories in acute neurology and the neurology of general medicine. Oxford University Press.
31. Ringner, M. (2008). What is principal component analysis? Nature Biotechnology, 26(3), 303–304.
32. Smola, A. J., & Scho¨lkopf, B. (2004). A tutorial on support vector regression. Statistics and Computing, 14(3), 199–222.
33. Solomon, T., Michael, B., Miller, A., & Kneen, R. (2019). Case studies in neurological infection. Cambridge University Press.
34. Tang, J., Alelyani, S., & Liu, H. (2014). Feature selection for classification: A review. Data classification: Algorithms and applications, 37.
35. The National Center for Biomedical Ontology. (2021). The Human Phenotype Ontology. https://bioportal.bioontology.org/ontologies/HP. (Uploaded: 2020-12-07)
36. Toy, E. C., Simpson, E., & Tintner, R. (2012). Case Files Neurology, the second edition. McGraw Hill.
37. Visalakshi, S., & Radha, V. (2014). A literature review of feature selection techniques and applications: Review of feature selection in data mining. In 2014 IEEE international conference on computational intelligence and computing research (pp. 1–6).
38. Waxman, S. (2009). Clinical neuroanatomy, 26th edition. McGraw-Hill Education.
39. Xiao, C., Choi, E., & Sun, J. (2018). Opportunities and challenges in developing deep learning models using electronic health records data: a systematic review. Journal of the American Medical Informatics Association, 25(10), 1419–1428.