



Predicting Ischemic Stroke Outcome Using Deep Learning Approaches

Gang Fang ^{*}, Zhennan Huang and Zhongrui Wang

Institute of Computing Science and Technology, Guangzhou University, Guangzhou, China

Predicting functional outcomes after an Ischemic Stroke (IS) is highly valuable for patients and desirable for physicians. This facilitates physicians to set reasonable goals for patients and cooperate with patients and relatives effectively, and furthermore to reach common after-stroke care decisions for recovery and make exercise plans to facilitate rehabilitation. The objective of this research is to apply three current Deep Learning (DL) approaches for 6-month IS outcome predictions, using the openly accessible International Stroke Trial (IST) dataset. Furthermore, another objective of this research is to compare these DL approaches with machine learning (ML) for performing in clinical prediction. After comparing various ML methods (Deep Forest, Random Forest, Support Vector Machine, etc.) with current DL frameworks (CNN, LSTM, Resnet), the results show that DL doesn't outperform ML significantly. DL methods and reporting used for analyzing structured medical data should be developed and improved.

OPEN ACCESS

Edited by:

Quan Zou,

University of Electronic Science and Technology of China, China

Reviewed by:

Yansen Su,

Anhui University, China

Cheng Zhang,

Peking University, China

*Correspondence:

Gang Fang

gangf@gzhu.edu.cn

Specialty section:

This article was submitted to Computational Genomics, a section of the journal Frontiers in Genetics

Received: 02 December 2021

Accepted: 15 December 2021

Published: 24 January 2022

Citation:

Fang G, Huang Z and Wang Z (2022)

Predicting Ischemic Stroke Outcome Using Deep Learning Approaches.

Front. Genet. 12:827522.

doi: 10.3389/fgene.2021.827522

INTRODUCTION

Stroke is one of the leading causes of death and permanent disability in the last 20 years globally (Global Burden of Disease Collaborative Network, 2018; World Health Organization, 2018). In China, the number of patients diagnosed with stroke each year is approximately 2 million, and the mortality rate is 11.48% (Chen et al., 2017). Stroke is mainly subtyped into ischemic (85%) and hemorrhagic types (15%) (Caplan, 2016). IS occurs when a cerebral artery is blocked (Park, 2017). Long-term physical disabilities after IS can create enormous mental and financial burdens for families and society. Proper exercise and early rehabilitation definitely improve recovery of patients and reduce disabilities (Veerbeek et al., 2011). Predicting a patient's functional outcomes precisely after a stroke will help physicians in managing an appropriate long-term plan for early rehabilitation. In addition, it guides clinicians in setting realistic goals, provides accurate information to patients and their caregivers, and facilitates the creation of an early discharge plan (Veerbeek et al., 2011). Now, endovascular treatment (EVT) is widely used for IS. Accurate prediction of functional outcomes and reperfusion may potentially improve stroke care, as it can guide selecting the most beneficial treatment option for the individual patient: to perform or to refuse EVT. Recently, clinical variables and radiological image biomarkers are utilized in studies on outcome prediction strategies in ischemic stroke patients after EVT (Venema et al., 2017; Van Os, 2018). More works have been devoted to predicting functional outcomes after stroke (Stinear, 2010; Meyer et al., 2015; Lin et al., 2020). Several medical communities have created and developed scores and methods that can predict the patient's functional outcomes after a stroke effectively by only using data readily collected at admission (Ntaios et al., 2012; Hilbert et al., 2019). The score statistically analyzes the data and identifies the most relevant predictors from a set of covariates selected by domain experts. The method uses deep learning to predict the functional outcome of patients with acute IS after EVT.

Recently, machine learning methods have been ubiquitously used to solve complex problems in many scientific fields, especially in medical science. Medical diagnosis and prognosis prediction are fulfilled in this way (Lin et al., 2018; Van Os, 2018; Debs et al., 2020; Fang et al., 2020).

Recently, DL frameworks have attained great success in various applications, particularly in image processing and natural language processing (NLP) (Hinton et al., 2012; Krizhevsky et al., 2012), leading to the hot wave of DL (Goodfellow et al., 2016). Though DL frameworks are powerful, they have apparent deficiencies. For example, large scale training data is always required for training, restricting the direct application of DL to tasks with smaller scale data. It is well known that DL is a supervised learning. But nowadays the data of many real tasks are still not sufficiently and correctly labeled due to the high cost of labeling. Because of this, DL frameworks always perform inferiorly in tasks with poor quality data. DL frameworks, especially modern deep neural networks, always possess too many hyper-parameters, and careful tuning of them directly can mainly influence the learning performance of DL. Recently, it is used to diagnose and predict prognosis in the clinical medical field (Ge et al., 2019; Hilbert et al., 2019; Debs et al., 2020). But it is seldomly used to analyze structured clinical medical data. In this paper, currently used DL frameworks are tested to predict stroke outcomes. Furthermore, several ML methods, especially Deep Forest (DF) (Zhou and Feng, 2019), are used to analyze IST dataset and are compared with several DL frameworks. The DF is proposed based on gcForest (multi-Grained Cascade Forest), which is a novel ensemble method of decision tree. This method generates a deep forest ensemble, with a cascade structure which enables gcForest to do representation learning (Zhou and Feng, 2019). Its representational learning ability can be further enhanced by multi-grained scanning when the inputs are with high dimensionality, potentially enabling gcForest to be contextual or structural aware (Zhou and Feng, 2019). In their experiments, the training time cost of DF is smaller than that of DL; even so, DF attains highly superior performance to DL. Herein, the DF and other ML methods are compared with DL to analyze structured clinical medical data. The results show that there are no evidences of superior performance of DL over ML.

MATERIALS AND METHODS

Data

The data used in this paper is The International Stroke Trial (IST) dataset. The IST, including the pilot phase between 1991 and 1993, was conducted between 1991 and 1996 and is a large, prospective, randomized controlled trial, with 100% complete baseline data and over 99% complete follow-up data. The objective of the trial is to know whether early administration of aspirin, heparin, both, or neither influenced the clinical course and outcome of acute IS (Sandercock et al., 2011). The dataset analyzed in this study is downloaded from the IST website. Patients in this trial are identified only by an anonymous code. They were treated more than 20 years ago, and many

have died. Hospitals are also identified by an anonymous code. There are no identifying data such as name, address, or social security numbers appearing. Patient age has been rounded to the nearest whole number. Thus, usage of the dataset definitely can't present material risk to confidentiality of patients.

The following baseline data: time from onset to randomization, gender, age, aspirin administration within 3 days prior to randomization, systolic blood pressure at randomization, presence or absence of atrial fibrillation (AF), level of consciousness, and neurological deficit, are all included in the dataset. Neurological deficits are classified as one of the Oxfordshire Community Stroke Project (OCSP) categories: posterior circulation syndrome (POCS), partial anterior circulation syndrome (PACS), total anterior circulation syndrome (TACS), and lacunar syndrome (LACS). A total of 19,435 patients from 467 hospitals in 36 countries are randomized within 48 h of symptoms onset, of whom 13,020 take a CT scan before randomization, 5,569 are first scanned after randomization, and 846 were not scanned at all. Entries with missing data are deleted, with 18,128 entries left. We exclude patients who are not finally diagnosed as IS. The variable of 6-month outcome is taken as a target. It is represented as 1-dead, 2-dependent, 3-not recovered, 4-recovered, and 8 or 9-missing status. The entries of 6-month outcome with missing status are also deleted. Six-month outcome of 2-dependent and 3-not recovered are merged as one category (not recovered) due to their similarity, and then the target includes three categories (0-dead, 1-not recovered, 2-recovered). At last, 16,403 patients are left. The data of these 16,403 patients finally diagnosed as IS are used to predict the outcome of IS using ML and DL.

METHODS

This paper investigates the ability of some supervised ML methods to predict IS outcomes. Classic ML methods such as support vector machine (SVM) (Cristianini and Shawe-Taylor, 2000), random forest (RF) (Liaw and Wiener, 2002), and deep forest (DF) (Zhou and Feng, 2019) are explored for comparison to DL frameworks such as convolutional neural network (CNN) (LeCun et al., 1998), long- and short-term memory network (LSTM) (Hochreiter and Schmidhuber, 1997), and residual neural network (Resnet) (He et al., 2016). Developing logistic regression models is the usual approach to analyze the stroke outcomes; however, an alternative of ML methods has been proposed, particularly for large-scale and multi-institutional data. The prominent advantage of ML is that it can easily incorporate newly available data and improve prediction performance (Hamed et al., 2014). Nowadays, DL frameworks are prevalent and succeed in the field of image processing and natural language processing (NLP). In this paper, classical ML methods are compared to popular DL frameworks to exhibit their respective performances.

The workflow of the study consists of three sections. Firstly, features collected at the beginning of and on 14 days of randomization in the refined IST dataset (including 16,403 patients) are used. Features, such as date and comments, are

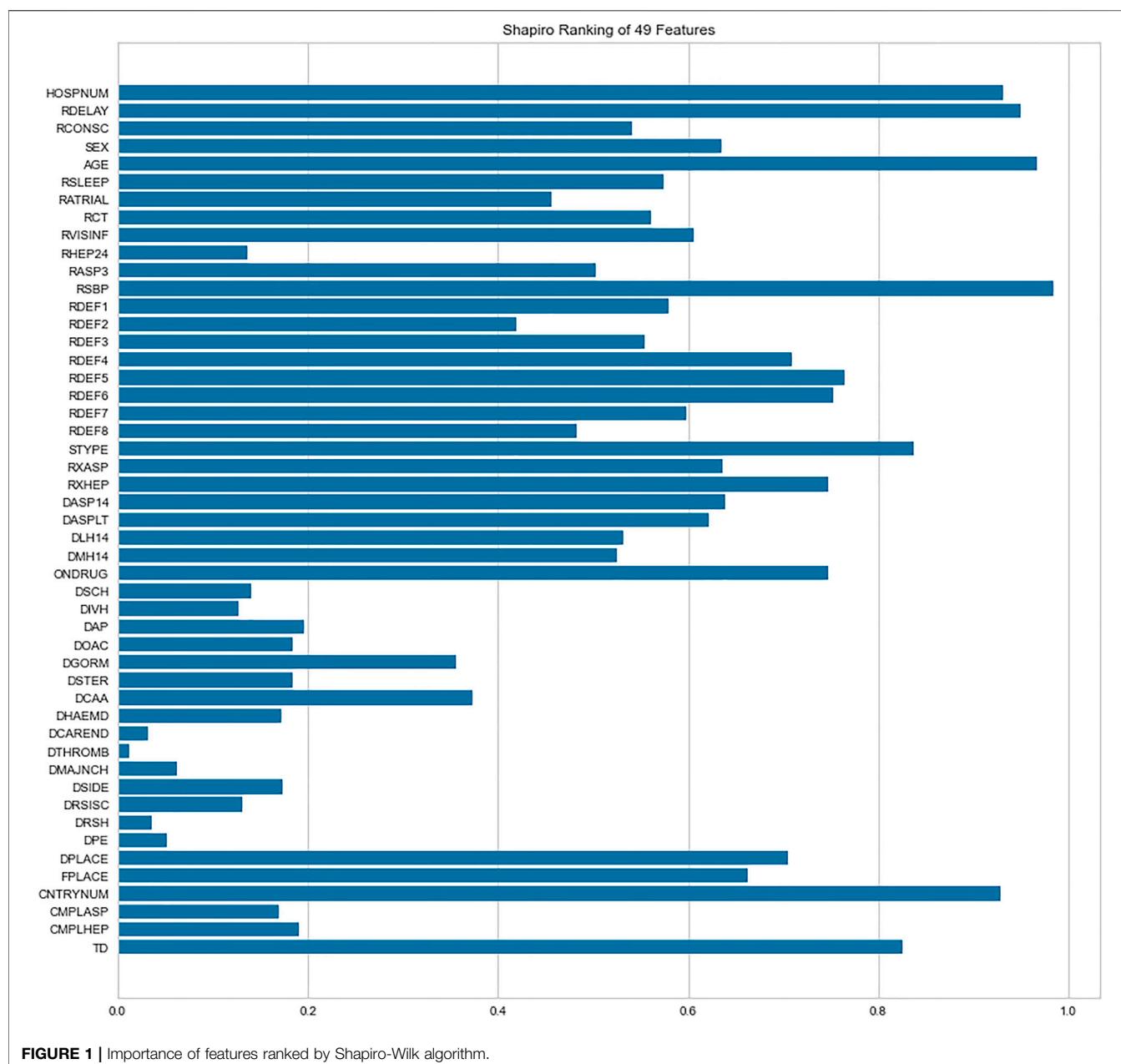


FIGURE 1 | Importance of features ranked by Shapiro-Wilk algorithm.

removed manually (features are definitely not related to IS outcome). Six-month outcome is kept as the target feature in the dataset. Features that overlap with 6-month outcome are deleted manually. Then, 50 features are kept and used. These features are utilized to predict long-term prognosis (6-month outcome) of acute IS. Based on previous research (Fang et al., 2020), feature selection carried out using recursive feature elimination with cross-validation (RFECV) don't eliminate explicitly less important features in the whole IST dataset. Thus, all initially chosen features are used to predict. Secondly, a simple CNN framework which consists of three convolutional layers and two fully connected layers is built, and the first convolutional layer is one dimensional convolution. The used

LSTM framework is a two-layer LSTM with just one direction and added into a one-dimensional convolutional layer before it. The last layer of the LSTM framework is a fully connected layer. The Resnet lacking bottleneck blocks which consists of eight residual blocks is also added into a one-dimensional convolutional layer as the first layer. This manipulation allows these DL frameworks to accept and process structured clinical medical data, such as IST. ML methods (SVM, RF, Multinomial-Naïve-Bayes, AdaBoost, and DF) are carried out immediately to compare with these DL frameworks. The SVM classifier use linear kernel (with the parameter `max_iter = 10,000`), and the other ML methods are carried out with default parameters. To implement these methods for this study, we use the libraries of scikit-learn 1.0.1 (Pedregosa

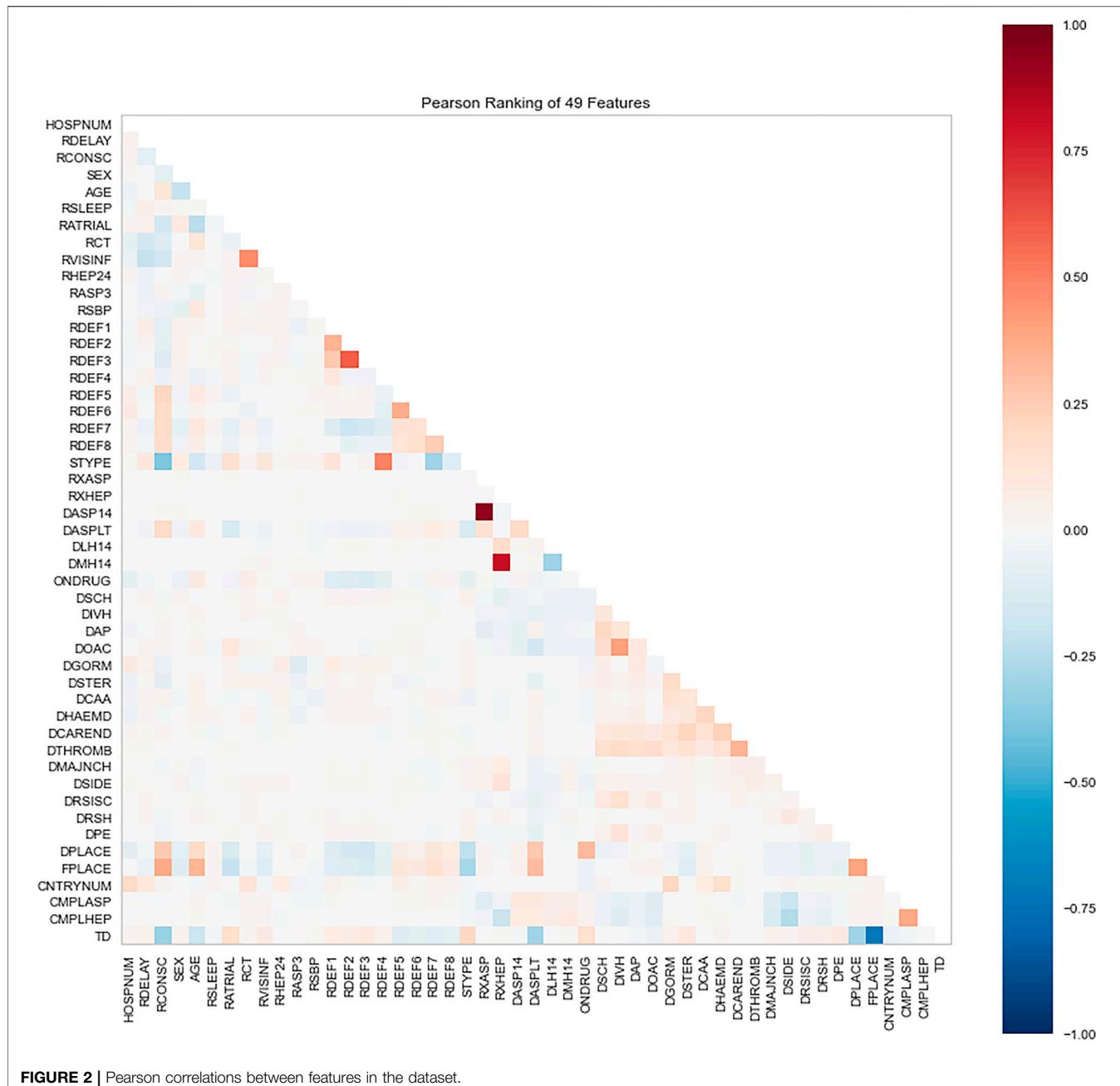


FIGURE 2 | Pearson correlations between features in the dataset.

et al., 2011) and PyTorch neural networks API (PyTorch., 2021). Thirdly, all these methods are implemented for comparison in predicting accuracy and other metrics.

RESULTS AND ANALYSIS

Because we only consider IS, 50 features are initially selected in the data of all kept 16,403 patients. The feature of 6-month outcome (OCCODE) is kept as target (including 3 categories: 0-dead, 1-not recovered, 2-recovered). The other 49 features include CNTRYNUM, HOSNUM, SEX, AGE, DPLACE, FPLACE, RDELAY, RCONSC,

RTRIAL, RSLEEP, RASP3, RSBP, RXASP, RXHEP, DASP14, DASPLT, RCT, RVISINF, DLH14, DMH14, neurological deficit symptom (RDEF1, RDEF2, RDEF3 ... etc.), STYPE, ONDRUG, DCAA, DOAC, TD, etc. Readers can be referred to **Supplementary Materials** for the detailed explanation of these features. Shapiro-Wilk algorithm is used to rank the importance of these features, and Pearson correlations between features are analyzed too. Shapiro-Wilk algorithm is a normal distribution assessing algorithm that regard the instances with respect to the feature, which is improved by Royston to process large data (Shapiro and Wilk, 1965; Royston, 1982). Except OCCODE, the other 49 features are ranked by the algorithm (**Figures 1, 2**).

	precision	recall	f1-score	support
0	0.99	1.00	0.99	849
1	0.78	0.99	0.88	2540
2	0.59	0.03	0.06	712
accuracy			0.83	4101
macro avg	0.79	0.67	0.64	4101
weighted avg	0.79	0.83	0.76	4101

FIGURE 3 | Performance of CNN after 5 epochs of training.

	precision	recall	f1-score	support
0	1.00	1.00	1.00	849
1	0.80	0.95	0.87	2540
2	0.49	0.18	0.27	712
accuracy			0.82	4101
macro avg	0.77	0.71	0.71	4101
weighted avg	0.79	0.82	0.79	4101

FIGURE 6 | Performance of DF.

	precision	recall	f1-score	support
0	0.98	0.99	0.99	849
1	0.78	0.99	0.87	2540
2	0.54	0.02	0.04	712
accuracy			0.82	4101
macro avg	0.77	0.67	0.63	4101
weighted avg	0.78	0.82	0.75	4101

FIGURE 4 | Performance of LSTM after 5 epochs of training.

	precision	recall	f1-score	support
0	1.00	1.00	1.00	849
1	0.80	0.96	0.87	2540
2	0.49	0.15	0.23	712
accuracy			0.83	4101
macro avg	0.76	0.70	0.70	4101
weighted avg	0.79	0.83	0.79	4101

FIGURE 7 | Performance of RF.

	precision	recall	f1-score	support
0	0.99	0.99	0.99	849
1	0.78	1.00	0.87	2540
2	0.25	0.00	0.01	712
accuracy			0.82	4101
macro avg	0.67	0.66	0.62	4101
weighted avg	0.73	0.82	0.75	4101

FIGURE 5 | Performance of Resnet after 5 epochs of training.

	precision	recall	f1-score	support
0	0.98	1.00	0.99	849
1	0.79	0.98	0.87	2540
2	0.46	0.05	0.08	712
accuracy			0.82	4101
macro avg	0.74	0.68	0.65	4101
weighted avg	0.77	0.82	0.76	4101

FIGURE 8 | Performance of SVM.

The Shapiro-Wilk results show that DTHROMB (Thrombolysis) and DCAREN (Carotid surgery) are the two least important features. The reason for this is that these therapies were seldom carried out in the 1990s. The Pearson analysis shows that high correlations between features are not common in the dataset. The highest correlated features are DASP14 (Aspirin given for 14 days or till death or discharge) and RXASP (Trial aspirin allocated) which are related to aspirin usage. The second highest correlated features are RXHEP (Trial heparin allocated) and DMH14 (Medium dose heparin given for 14 days or till death/discharge) which are related to heparin usage. After this all these 49 features (except OCCODE) are adopted to predict the outcome of IS using ML and DL.

Firstly, all selected 49 features of the IST dataset which consists of 16,403 patients are processed by DL frameworks. The dataset is divided into training set including 12,302 patients and test set including 4,101 patients randomly. When processed by CNN, 5 epochs of training are carried out and attain an accuracy of 0.826 in test set. Other metrics including precision, recall, and f1-score are

also considered (**Figure 3**). When processed by LSTM, 5 epochs of training are also carried out and attain an accuracy of 0.821 in test set. Other metrics are shown in **Figure 4**. When processed by Resnet, 5 epochs of training are carried out and attain an accuracy of 0.821 in test set. Other metrics are shown in **Figure 5**. In this study, all DL frameworks are trained with fewer epochs because more epochs of training lead to overfitting.

After this, test sets including all 4,101 patients are processed by ML approaches. First by DF, it attains an accuracy of 0.824 in test set. Other metrics including precision, recall, and f1-score are also considered (**Figure 6**). The performances of SVM and RF are showed in **Figures 7, 8**. For performances of other ML methods, readers can be referred to **Supplementary Materials** for more details.

The results show that DL frameworks don't outperform ML methods in any aspects when predicting IS outcomes in IST dataset. On the contrary, ML methods, especially DF, outperform DL in predicting IS outcomes of *recovered*. It attains a higher precision,

	precision	recall	f1-score	support
0	1.00	0.96	0.98	849
1	0.82	0.91	0.86	2540
2	0.49	0.31	0.38	712
accuracy			0.82	4101
macro avg	0.77	0.73	0.74	4101
weighted avg	0.80	0.82	0.80	4101

FIGURE 9 | Performance of CNN with 44 features after 20 epochs of training.

	precision	recall	f1-score	support
0	1.00	0.99	1.00	849
1	0.78	0.99	0.88	2540
2	0.62	0.04	0.07	712
accuracy			0.83	4101
macro avg	0.80	0.67	0.65	4101
weighted avg	0.80	0.83	0.76	4101

FIGURE 11 | Performance of Resnet with 44 features after 20 epochs of training.

	precision	recall	f1-score	support
0	0.99	1.00	0.99	849
1	0.80	0.96	0.87	2540
2	0.49	0.12	0.20	712
accuracy			0.82	4101
macro avg	0.76	0.69	0.69	4101
weighted avg	0.78	0.82	0.78	4101

FIGURE 10 | Performance of LSTM with 44 features after 20 epochs of training.

	precision	recall	f1-score	support
0	1.00	1.00	1.00	849
1	0.81	0.95	0.87	2540
2	0.52	0.19	0.28	712
accuracy			0.83	4101
macro avg	0.77	0.71	0.72	4101
weighted avg	0.80	0.83	0.80	4101

FIGURE 12 | Performance of DF with 44 features.

recall, and f1-score in predicting the outcomes of *recovered* (represented as 2, **Figure 6**). All methods, especially DL, don't work well in predicting the outcomes of *recovered*. The reason of this lies in the heterogeneity of data in this category. In other words, there are more variables that can exert influence on the recovery of IS patients.

Based on previous Shapiro-Wilk analysis, the less important features whose Shapiro-Wilk ranking value is less than 0.1 are eliminated. These features include DTHROMB (Thrombolysis), DCAREND (Carotid surgery), DRSH (Recurrent stroke within 14 days, Haemorrhagic stroke), DPE (Other events within 14 days, Pulmonary embolism), and DMAJNCH (Major non-cerebral haemorrhage). Then 44 features are left for predicting the outcomes of IS. The predicting performances of DL frameworks are compared to ML methods with these features. When processed by CNN, after 20 epochs of training it attains an accuracy of 0.817 in test set. Other metrics including precision, recall, and f1-score are also considered (**Figure 9**). After 20 epochs of training LSTM attains an accuracy of 0.823 in test set. Other metrics are shown in **Figure 10**. After 20 epochs of training Resnet attains an accuracy of 0.827 in test set. The accuracy doesn't decrease because the eliminated 5 features are less important and not related to the 6th outcome. Other metrics are shown in **Figure 11**.

Subsequently, test sets including all 4,101 patients with 44 features are processed by ML approaches. DF attains an accuracy of 0.828 in the test set. Other metrics including precision, recall, and f1-score are considered (**Figure 12**). The performances of SVM and RF are shown in **Figures 13, 14**. For performances of other ML methods, readers can be referred to **Supplementary Materials** for more.

	precision	recall	f1-score	support
0	1.00	1.00	1.00	849
1	0.80	0.96	0.87	2540
2	0.51	0.15	0.23	712
accuracy			0.83	4101
macro avg	0.77	0.70	0.70	4101
weighted avg	0.79	0.83	0.79	4101

FIGURE 13 | Performance of RF with 44 features.

	precision	recall	f1-score	support
0	1.00	0.96	0.98	849
1	0.77	1.00	0.87	2540
2	0.43	0.01	0.02	712
accuracy			0.82	4101
macro avg	0.73	0.65	0.62	4101
weighted avg	0.76	0.82	0.75	4101

FIGURE 14 | Performance of SVM with 44 features.

The results show that there is no decrease in predicting performance using both ML and DL after eliminating the five least important features. But compared to previous results, after 20 epochs of training Resnet attains a higher precision than before in

	precision	recall	f1-score	support
0	1.00	1.00	1.00	849
1	0.85	0.80	0.83	2540
2	0.42	0.50	0.45	712
accuracy			0.79	4101
macro avg	0.75	0.77	0.76	4101
weighted avg	0.81	0.79	0.80	4101

FIGURE 15 | Performance of Resnet with 44 features after 100 epochs of training.

predicting the outcomes of *recovered* with 44 features (represented as 2, **Figure 11**). Considering this observation, more epochs of training are carried out to attempt to explain this. After 100 epochs of training, the predicting accuracy of Resnet is 0.791 in test set. But it gets higher recall and f1-score than before in predicting the outcomes of *recovered* (**Figure 15**). After 500 epochs of training, the predicting accuracy of Resnet is 0.794 in test set and the other two DL frameworks overfit (accuracy of CNN and LSTM is 0.740 and 0.769 respectively). But there are no increases in recall and f1-score when predicting the outcomes of *recovered* (represented as 2, **Supplementary Figure S1**). After 100 epochs training of Resnet, the overall predicting accuracy decreased. But macro and weighted average f1-score increased (**Figure 15**) and are better than before (**Figures 3–5, Figures 9–11**). Macro and weighted average f1-score are an important index for performance of multi-classification tasks. It is suggested that Resnet will work better when trained appropriately, but it doesn't outperform ML methods especially DF significantly in this case (**Figures 12, 15**). When trained 500 epochs, it starts overfitting (**Supplementary Figure S1**). For more information readers can be referred to **Supplementary Materials**.

DISCUSSION

In this study, classic ML algorithms and current DL frameworks are adopted to predict the outcomes of IS in IST dataset. Both methods attain considerable accuracy. The performances of ML and DL are also compared. The results show that adapted DL frameworks don't outperform ML in predicting capability, although Resnet raised the weighted average f1-score after trained by 100 epochs (**Figure 15**). The main reason of this lies in that the used DL frameworks are developed and employed for processing image and serial data. They are seldom used in censored and structured medical clinical data. In this study, three DL frameworks, CNN, LSTM, and Resnet, are adapted to process this sort of data and predicting the outcomes of IS. The structure of the adapted CNN is similar to LeNet-5 (LeCun et al., 1998) with an added one-dimensional convolutional layer as the first convolutional layer. The used LSTM and Resnet are also added to a one-dimensional convolutional layer as the first layer. In this way, these DL frameworks can admit and process tabulated data, such as structured medical data. CNN attains the accuracy of 0.83 when trained with less epochs, but it gets less f1-score (**Figure 3**). This suggested that it doesn't work well in multiclassification task, so

does LSTM (**Figure 4**). After eliminating the 5 least important features and after trained with more epochs (100 epochs), Resnet gets a higher weighted average f1-score (**Figure 15**). The first reason is that the left 44 features are more important to the outcomes of IS. The second reason is that Resnet is a fairly complex DL framework. It adopts residual shortcut connection to overcome degradation problems. When trained appropriately Resnet can capture some intrinsic qualities of the tabulated data and work better in a multiclassification task. In this study, the used Resnet is similar to Resnet18 which possesses fewer layers. Next, deeper Resnet framework and more powerful computing workstations will be adopted to study this issue.

To investigate the predicting capability of DL in the IST dataset, the performances of classic ML algorithms are compared to them. The results show that DL doesn't surpass ML. Resnet raises f1-score after 100 epochs training with the selected 44 features. After eliminating the 5 least important features, the DF and RF raise the f1-score a little and both attain the accuracy of 0.83 (**Figures 12, 13**). And moderate f1-scores are attained in previous training and test (**Figures 6, 7**). This means the left 44 features are more important and the 2 ML classifiers are robust to be used in this sort of data. The newly proposed DF is used to be compared with DL frameworks. In our experiments, DF doesn't achieve highly competitive performance to deep neural networks, although the training time cost of DF is smaller than that of deep learning. The reason of this lies in that the used features were collected in the early 1990s. Some important features may be neglected, and this reduces the predicting ability of ML and DL. Next, deeper DL frameworks will be adopted to investigate the performance of them. Furthermore, some new features and variables will be collected to enhance the performance of the machine learning and deep learning approaches.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, and further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

GF wrote the manuscript and designed the study. ZH and ZW wrote some codes.

FUNDING

The study is supported by National Natural Science Foundation of China with grant number 61972107.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2021.827522/full#supplementary-material>

REFERENCES

- Caplan, L. R. (2016). *Caplan's Stroke*: Massachusetts. Cambridge University Press.
- Chen, Z., Jiang, B., Ru, X., Sun, H., Sun, D., Liu, X., et al. (2017). Mortality of Stroke and its Subtypes in China: Results from a Nationwide Population-Based Survey. *Neuroepidemiology* 48, 95–102. doi:10.1159/000477494
- Cristianini, N., and Shawe-Taylor, J. (2000). *An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods*. London: Cambridge University Press.
- Debs, N., Rasti, P., Victor, L., Cho, T.-H., Frindel, C., and Rousseau, D. (2020). Simulated Perfusion MRI Data to Boost Training of Convolutional Neural Networks for Lesion Fate Prediction in Acute Stroke. *Comput. Biol. Med.* 116, 103579. doi:10.1016/j.combiomed.2019.103579
- Fang, G., Liu, W., and Wang, L. (2020). A Machine Learning Approach to Select Features Important to Stroke Prognosis. *Comput. Biol. Chem.* 88, 107316. doi:10.1016/j.combiolchem.2020.107316
- Ge, Y., Wang, Q., Wang, L., Wu, H., Peng, C., Wang, J., et al. (2019). Predicting post-stroke Pneumonia Using Deep Neural Network Approaches. *Int. J. Med. Inform.* 132, 103986. doi:10.1016/j.ijmedinf.2019.103986
- Global Burden of Disease Collaborative Network (2018). *Global Burden of Disease Study 2017*. USA: IHME, 1–7.
- Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep Learning*. Cambridge, MA: MIT Press.
- Hamed, A., Richard, D., Bernard, Y., and Peter, M. (2014). Machine Learning for Outcome Prediction of Acute Ischemic Stroke Post Intra-arterial Therapy. *PLOS ONE* 9 (2), e88225. 10.1371/journal.pone.0088225.
- He, K., Zhang, X., Ren, S., and Sun, J. (2016). “Deep Residual Learning for Image Recognition[CI],” in 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, June 27–30, 2016, 770–778. doi:10.1109/CVPR.2016.90
- Hilbert, A., Ramos, L. A., van Os, H. J. A., Olabarriaga, S. D., Tolhuisen, M. L., Wermer, M. J. H., et al. (2019). Data-efficient Deep Learning of Radiological Image Data for Outcome Prediction after Endovascular Treatment of Patients with Acute Ischemic Stroke. *Comput. Biol. Med.* 115, 103516. doi:10.1016/j.combiomed.2019.103516
- Hinton, G., Deng, L., Yu, D., Dahl, G., Mohamed, A.-r., Jaitly, N., et al. (2012). Deep Neural Networks for Acoustic Modeling in Speech Recognition: The Shared Views of Four Research Groups. *IEEE Signal. Process. Mag.* 29 (6), 82–97. doi:10.1109/msp.2012.2205597
- Hochreiter, S., and Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Comput.* 9 (8), 1735–1780. doi:10.1162/neco.1997.9.8.1735
- Krizhevsky, A., Sutskever, I., and Hinton, G. (2012). “ImageNet Classification with Deep Convolutional Neural Networks,” in NIPS, Lake Tahoe, NV, December 3–6, 2012, 1097–1105.
- LeCun, Y., Bottou, L., Bengio, Y., and Haffner, P. (1998). Gradient-based Learning Applied to Document Recognition. *Proc. IEEE* 86 (11), 2278–2324. doi:10.1109/5.726791
- Liaw, A., and Wiener, M. (2002). Classification and Regression by randomForest. *R News* 2, 18–22. <https://www.researchgate.net/publication/228451484>
- Lin, W.-Y., Chen, C.-H., Tseng, Y.-J., Tsai, Y.-T., Chang, C.-Y., Wang, H.-Y., et al. (2018). Predicting post-stroke Activities of Daily Living through a Machine Learning-Based Approach on Initiating Rehabilitation. *Int. J. Med. Inform.* 111, 159–164. doi:10.1016/j.ijmedinf.2018.01.002
- Lin, C.-H., Hsu, K.-C., Johnson, K. R., Fann, Y. C., Tsai, C.-H., Sun, Y., et al. (2020). Evaluation of Machine Learning Methods to Stroke Outcome Prediction Using a Nationwide Disease Registry. *Comp. Methods Programs Biomed.* 190, 105381. doi:10.1016/j.cmpb.2020.105381
- Meyer, M. J., Pereira, S., McClure, A., Teasell, R., Thind, A., Koval, J., et al. (2015). A Systematic Review of Studies Reporting Multivariable Models to Predict Functional Outcomes after post-stroke Inpatient Rehabilitation. *Disabil. Rehabil.* 37, 1316–1323. doi:10.3109/09638288.2014.963706
- Ntaios, G., Faouzi, M., Ferrari, J., Lang, W., Michel, P., and Vemmos, K. (2012). An Integer-Based Score to Predict Functional Outcome in Acute Ischemic Stroke: The ASTRAL Score. *Neurology* 79, 2293–2294. doi:10.1212/WNL.0b013e31827a3c0a
- Park, J. (2017). *Acute Ischemic Stroke*. Unitedstates: Springer.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. (2011). Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* 12, 2825–2830. <https://arxiv.org/pdf/1201.0490v4.pdf>
- PyTorch (2021). Available at: <https://pytorch.org> (Accessed Feb 24, 2021)
- Royston, J. P. (1982). An Extension of Shapiro and Wilk's W Tests for Normality to Large Samples. *Appl. Statist.* 31 (1), 115–124. doi:10.2307/2347973
- Sandercock, P. A., Niewada, M., and Czlonkowska, A. (2011). The International Stroke Trial Database. *Trials* 12 (12), 101. doi:10.1186/1745-6215-12-101
- Shapiro, S. S., and Wilk, M. B. (1965). An Analysis of Variance Test for Normality (Complete Samples). *Biometrika* 52, 591–611. doi:10.1093/biomet/52.3-4.591
- Stinear, C. (2010). Prediction of Recovery of Motor Function after Stroke. *Lancet Neurol.* 9, 1228–1232. doi:10.1016/s1474-4422(10)70247-7
- Van Os, H. J. A., Ramos, L. A., Hilbert, A., van Leeuwen, M., van Walderveen, M. A. A., Kruijt, N. D., et al. (2018). Predicting Outcome of Endovascular Treatment for Acute Ischemic Stroke: Potential Value of Machine Learning Algorithms. *Front. Neurol.* 9 (9), 1–8. doi:10.3389/fneur.2018.00784
- Veerbeek, J. M., Kwakkel, G., van Wegen, E. E. H., Ket, J. C. F., and Heymans, M. W. (2011). Early Prediction of Outcome of Activities of Daily Living after Stroke. *Stroke* 42 (5), 1482–1488. doi:10.1161/strokeaha.110.604090
- Venema, E., Mulder, M. J. H. L., Roozenbeek, B., Broderick, J. P., Yeatts, S. D., Khatri, P., et al. (2017). Selection of Patients for Intra-arterial Treatment for Acute Ischaemic Stroke: Development and Validation of a Clinical Decision Tool in Two Randomised Trials. *BMJ* 357, j1710. doi:10.1136/bmj.j1710
- World Health Organization (2018). *WHO - The Top 10 Causes of Death*. Geneva: Maggio, 24.
- Zhou, Z. H., and Feng, J. (2019). Deep Forest: Towards an Alternative to Deep Neural Networks. *Natl. Sci. Rev.* 6 (1), 74–86. <https://arxiv.org/pdf/1702.08835v2.pdf>

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Fang, Huang and Wang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.