

Machine Learning for Brain Stroke: A Review

Manisha Sanjay Sirsat,^{*} Eduardo Fermé,^{*†} and Joana Câmara,^{*†‡}

Machine Learning (ML) delivers an accurate and quick prediction outcome and it has become a powerful tool in health settings, offering personalized clinical care for stroke patients. An application of ML and Deep Learning in health care is growing however, some research areas do not catch enough attention for scientific investigation though there is real need of research. Therefore, the aim of this work is to classify state-of-arts on ML techniques for brain stroke into 4 categories based on their functionalities or similarity, and then review studies of each category systematically. A total of 39 studies were identified from the results of ScienceDirect web scientific database on ML for brain stroke from the year 2007 to 2019. Support Vector Machine (SVM) is obtained as optimal models in 10 studies for stroke problems. Besides, maximum studies are found in stroke diagnosis although number for stroke treatment is least thus, it identifies a research gap for further investigation. Similarly, CT images are a frequently used dataset in stroke. Finally SVM and Random Forests are efficient techniques used under each category. The present study showcases the contribution of various ML approaches applied to brain stroke.

Keywords: Support vector machine—Machine learning—Deep learning—Stroke diagnosis—Stroke prevention—Stroke prognostication

© 2020 Elsevier Inc. All rights reserved.

Introduction

Nowadays, stroke is a major health-related challenge [52]. Stroke, also known as cerebrovascular accident, consists of a neurological disease that can result from ischemia or hemorrhage of the brain arteries, and usually leads to heterogeneous motor and cognitive impairments that compromise functionality [34]. Annually, stroke affects about 16 million individuals worldwide and is associated with enormous societal costs. In recent years, ML has rapidly grown and evolved in several applications in a wide variety of health care systems. Figure 1 shows the latest global health estimates by

cause for from the year 2000-2016. It identifies ischaemic heart disease and stroke as the two leading causes of mortality and disability [38] Worldwide. According to the American Heart Association, stroke is considered a severe health issue due to its high mortality rate [9]. Also, the cost of hospitalization for stroke is increasing [17] and consequently, there is an increased need of advanced technologies that can assist in clinical diagnosis, treatment, predictions of clinical events, recommendation of promising therapeutic interventions, rehabilitation programs [2], etc. Early detection of stroke is a crucial step for efficient treatment and ML can be of great value in this process. To be able to do that, Machine Learning (ML) is an ultimate technology which can help health professionals make clinical decisions and predictions. During the past few decades, several studies were conducted on the improvement of stroke diagnosis using ML in terms of accuracy and speed. With that consideration, the current work classifies some of those studies based on their similarity, reviews each classification methodically and provides valuable information regarding the

From the ^{*}NOVA LINCS, Quinta da Torre, Campus Universitário, Caparica 2829-516, Portugal; [†]University of Madeira, Rua Dos Ferreiros 105, Funchal 9000-082 Portugal; and [‡]University of Coimbra, Rua do Colégio Novo, 3000-115, Coimbra.

Received April 11, 2020; revision received July 8, 2020; accepted July 11, 2020.

Corresponding author. E-mail: manisha.sirsat2015@gmail.com.

1052-3057/\$ - see front matter

© 2020 Elsevier Inc. All rights reserved.

<https://doi.org/10.1016/j.jstrokecerebrovasdis.2020.105162>

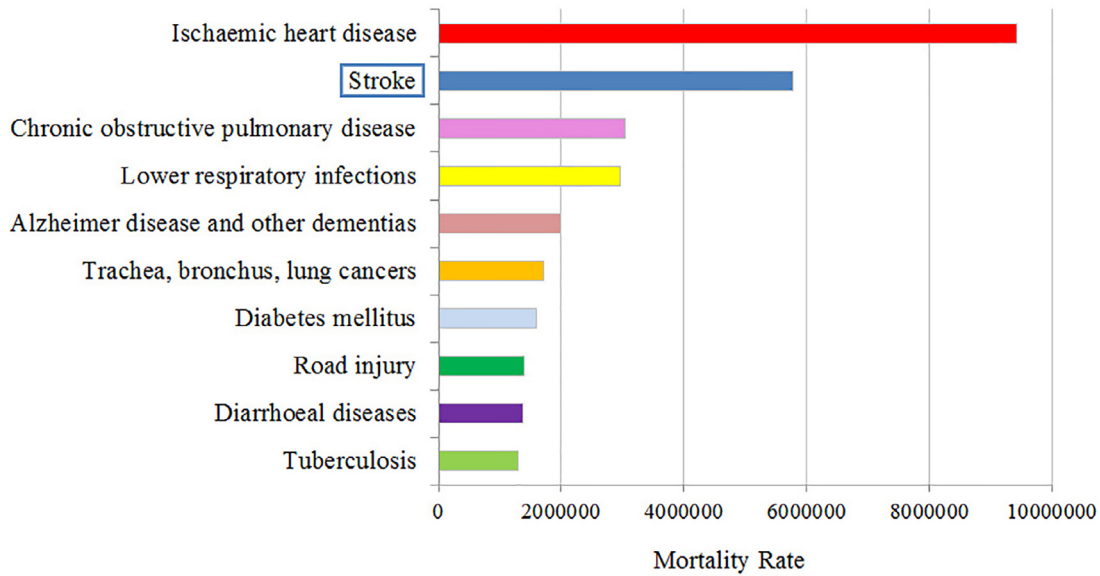


Fig. 1. Figure ranks estimated mortality by cause [69], recorded in 2016.

application of ML-based methods in brain stroke. To the best of our knowledge there is no detailed review about the application of ML for brain stroke. In the Section 2, we review some literature about ML and brain stroke field whereas, Section 3 presents the study design and selection, search strategy, and categorization of the studies per category. Section 4 explains outcomes of the studies by each category. Finally, Section 5 discusses insights of reviewed studies.

Background

Machine Learning

This section provides a brief overview of ML sub types. ML is a crucial branch of Artificial Intelligence (AI). ML is the study of algorithms that have the ability to learn and improve from past data without being explicitly programmed. There are several sub types of ML, however in this review, we focus on supervised learning, unsupervised learning and deep learning are focused types in the current study. Supervised learning trains a model that maps an input to an output based on observations and predicts the output. It is categorised into classification and regression. Classification [23] is to classify discrete target variable using predictors. Logistics Regression (LR) [41], naive bayes classifier [37], Support Vector Machines (SVM) [10] are some classification techniques. Regression investigates the relationship between numerical target variable and predictors. The study [24] has done extensive experimental analysis of 77 regression techniques from 19 ML families over 84 datasets. Unsupervised learning [22] is used to group the observations for making clusters based on their similarity. Clustering [70], association analysis and dimensionality reduction are types of unsupervised

learning [1,20,59]. Deep learning (DL) develops a computational model [44] with multi-processing layers to learn a data progressively from raw input. Some DL architectures are Convolutional Neural Networks (CNN) and Recurrent neural network (RNN) and they are mostly used to solve image processing [63] problems. Some of the top applications of DL are self-driving cars, natural language processing, fraud detection, health care, etc. Eventually, ML has a great implementation in the several fields like medicine, social networking, agriculture, sales and marketing, environmental science, etc. using several ML algorithms.¹

Stroke

Stroke consists of a heterogeneous group of disorders characterized by a sudden and focal disruption of the brain's vascular supply, causing neurologic symptoms that persist more than 24 hours [33]. Broadly, strokes can be classified as either ischaemic, which are the most common, happening in approximately 85% of the cases, or haemorrhagic. Ischaemic strokes occur when blood vessels are occluded by a thrombus or an embolus, resulting in brain ischemia [31]. Haemorrhagic strokes are caused by the rupture and bleeding of a weakened blood vessel into the surrounding brain tissue, which normally leads to intracranial pressure [31]. There are many risk factors that can precipitate a stroke; these can be categorized as non-modifiable and modifiable. The first set of risk factors include age, gender, race and ethnicity. In contrast, modifiable risk factors are related to clinical conditions, such as cardiac diseases (e.g., hypertension, atrial fibrillation,

¹For detailed overview of ML see [27].

hypercholesterolemia) and diabetes mellitus, and life-style factors, namely sedentarism, obesity, poor nutrition, tobacco use and alcohol consumption [16,53]. Early identification and better control of modifiable risk factors is determinant for stroke prevention and to avoid the latter development of vascular dementia [58]. Stroke diagnosis involves a detailed medical history, a physical and neurological examination, and a brain imaging test (e.g., computed tomography (CT) scan or magnetic resonance imaging (MRI)) in order to rule out other stroke mimics (e.g., brain tumors, subdural hematomas) and to determine the type of stroke, its location and the extent of the brain injury [64]. Treatment will depend on the cause and severity of the stroke and it can include surgical procedures (e.g., thrombectomy in case of ischemic stroke), specific medication to dissolve blood clots (e.g., thrombolytic therapy for ischaemic strokes) or to manage cardiovascular risk factors (e.g., statins used in the treatment of hypercholesterolaemia) and rehabilitation (e.g., motor and neurocognitive) [15,42,66]. It is important to note that since poststroke survival rate is increasing, stroke patients tend to live longer with innumerable sequelae, which are not only physical, but also cognitive and emotional, and negatively influence their quality of life and functionality. Despite this fact, rehabilitation after stroke is still primarily directed towards motor function and, therefore, should follow an interdisciplinary and holistic approach in order to overcome and/or mitigate the multidimensional post-stroke impairments [30,33]. Finally, prognosis prediction following stroke is extremely relevant, namely in treatment selection (e.g., identifying which patients will benefit from a specific type of treatment), in determining long-term outcomes (e.g., motor, cognitive, functional) and in planning rehabilitation by establishing appropriate goals, in light of the previous outcome predictions, and redefining unrealistic expectations that could hamper the patient's motivation and engagement in the rehabilitation process [8,61].

Methodology

Study Design

The goal of the present study is to review the contribution of ML in solving some stroke-related problems (e.g., i.e., prevention/risk factor identification, diagnosis, treatment and prognostication). It considers the most systematic approach because it reviews best of studies and writes a result and discussion with some interesting insights, get from all studies which are definitely valuable for further investigation.

Search Strategy

We analysed studies from 2007 to 2019, in fact all available studies. Studies were identified from the results of ScienceDirect web scientific database. Below search term

was searched in the research title, abstract or author-specified keywords of ScienceDirect database.

["machine learning"] AND ["stroke"]

Study Selection

Study selection was done with the help of a Psychology expert. We identified a total of 74 studies, accessed on 7th November 2019; 73 studies were written in English and 1 was written in Spanish. We have only considered studies written in English. We excluded 34 studies that were not related to our topic (e.g., pen stroke, stroke trajectories in a hand drawn pen-and-ink hatching, Chinese character recognition, electric power systems, thai document images, Alzheimer's disease, osteoporosis detection, composition and declarative gesture description, estimation of mass damper parameters, monitor physical activity in children, and more). Therefore, after checking the relevance of each study, we included 39 studies in the review. All included studies were then analyzed individually, considering 8 research questions, specified in the [Table 1](#):

Category

All studies were grouped into 4 categories based on their similarity with the help of a psychologist. We considered the following categories: (a) stroke prevention, (b) stroke diagnosis, (c) stroke treatment, and (d) stroke prognostication/outcome prediction.

Results

After screening all studies by title, abstract and conclusion, we found 8 studies about stroke prevention, 18 studies about stroke diagnosis, 4 studies about stroke treatment, and 9 studies about stroke prognostication. ML applications for stroke problems are several, although we only present the most advanced studies of each category, given in the [Table 3](#), [5](#), [4](#), [6](#), [7](#). The [Table 2](#) briefly explains the list of performance measures used in the reviewed studies.

Table 1. Research questions to analyze each study one-by-one.

-
- (1) Which was the stroke problem addressed?
 - (2) What was the ML or DL approach employed?
 - (3) Which features were used?
 - (4) How many samples were used?
 - (5) What were the classes used?
 - (6) Is there any pre-processing like feature selection, data augmentation, etc.?
 - (7) What were the performance metrics used?
 - (8) What are the clinical implications?
-

Table 2. List of performance measures involved in 39 studies.

-
1. **Classification Accuracy:** It is the fraction of the number of correct predictions by the total number of predictions.
 2. **Precision:** It is also known as positive predictive value, It shows the correctness achieved in positive prediction.
 3. **Recall or sensitivity:** it is the fraction of the total amount of relevant instances that were actually retrieved.
 4. **Precision-recall curve:** It shows the trade-off between the true positive rate and the positive predictive value for a classifier at all classification thresholds.
 5. **Specificity:** It is the measure of negative examples labeled as negative by classifier.
 6. **F-score:** It is a weighted average of the recall and precision.
 7. **Receiver Operating Characteristic (ROC):** It is a trade-off between the true positive rate and the false positive rate for a classifier at all classification thresholds.
 8. **Area Under the Curve (AUC):** It measures the entire 2D area underneath the entire ROC curve.
 9. **Negative predictive value:** It is the fraction of number of true negatives and number of negative calls.
 10. **Paired t-test:** It is used to compare two population means from the same unit or subject.
 11. **Root Mean Square Error (RMSE):** It can be considered as the standard deviation of residual.
 12. **Correlation (R):** It shows the relationship between pairs of variables.
 13. **Coefficient of determination (R^2):** It is the proportion of the variance in the output variable that is predictable from the input variables.
 14. **Computational time:** It is the amount of time required to perform computational processes.
-

Stroke Prevention

In this category, we have considered studies that focus on the early detection or identification of stroke; some studies mention important stroke risk factors and post-stroke complications, namely familiar hypercholesterolaemia, atherosclerosis, body mass index and cerebral edema, respectively, while others refer to specific techniques that can aid stroke detection, such as genome-wide transcriptional profiling and the use of spiking neural networks.. Table 3 shows 8 stroke prevention studies, ML approach to solve those problems, features, number of samples, results achieved by ML approach and lastly, clinical implications. Familial hypercholesterolaemia is a genetic condition characterized by the presence of a high amount of low-density lipoprotein (LDL), which can lead to premature heart disease and stroke. Since most individuals with familial hypercholesterolaemia remain undiagnosed, [55] aimed to predict the familial hypercholesterolaemia score (responsible for early detection of heart-related injuries) using [14] Random Forest (RF). The FIND FH model (or RF model) parameter optimization was done by 5-fold cross validation and got precision of 0.85 and recall 0.45, shown in the Table 3. Similarly, carotid-artery atherosclerosis (CA) can also increase the risk of ischaemic strokes due to the accumulation of plaques in the carotid arteries - the main supply of blood to the brain. SVM was used for 4-class classification problem [11] with the aim of classifying subjects with CA from magnetic resonance images. Cerebral edema is a major poststroke complication that is associated with high mortality. Considering that, RF model was developed [18] for the segmentation of cerebrospinal fluid on Computed Tomography (CT) scan images and concluded that RF is better than Hounsfield Unit thresholding. This approach accomplished an accuracy rate of 97.5%. For stroke monitoring, carotid intima-media thickness (cIMT) is used as

an important biomarker. Study [12] presented 2-level DL based system for cIMT measurements. In the first level, feature extraction and image segmentation were done and second layer consisted of ML regression for smoothing the raw inner lumen and outer interadventitial borders which are used for the cIMT measurements. Additionally, [51] proposed a fully automatic strategy by identifying the Intima-Media Thickness boundaries from ultrasound common carotid artery images using Artificial Neural Networks (ANN) and training computational time is 1.4s. Genome-wide transcriptional profiling can be useful in stroke detection. Study [56] identified a 10-gene pattern of differential expression using ML technique (here, genetic algorithm/kNN) which has enough ability for stroke detection.

The study [40] presented personalised modelling of spatio or spectro-temporal data (SSTD) and early prediction of stroke events for each individual by using Spiking Neural Networks (SNN) with accuracy 94%. Last but not least, Body Mass Index (BMI) is a good indicator of early stroke prediction thus, the study [45] classified BMI status into three classes (normal, overweight and obese) on imbalanced data by LR, bagging and RF ensemble techniques.

Stroke Diagnosis

This section includes a total of 18 studies about stroke diagnosis. Table 5 and 4 present 18 objectives on stroke diagnosis, ML techniques, features, sample size, result of optimal ML model and clinical implications. MRI and CT scans are frequently used in stroke diagnosis. Bio-imaging is an important process in stroke classification or lesion extraction because it identifies the damaged area of the brain, thus more studies focus on this problem to achieve high precision. To target this problem, [57] presented a new approach to the classification of stroke subtypes, specifically haemorrhagic and ischaemic strokes, using SVM,

Table 3. Machine learning techniques for brain stroke identification. Note: Support Vector Machine (SVM), Deep Learning (DL), Random Forest (RF), Logistic Regression (LR), k-nearest neighbors (kNN), Artificial Neural Networks (ANN)

Sr. No	Objective	Reference	ML-based approach	Samples	Features	Optimal Results	Clinical Implications
1	Identify subjects with Carotid-artery Atherosclerosis (CA) from a heterogeneous cohort of magnetic resonance brain images	[11]	SVM	401	MRI images	SVM ACC= 97.5%	Early identification of patients with CA will help prevent ischemic stroke by providing the proper diagnosis and treatment.
2	The carotid intima-media thickness (cIMT), Biomarker for cardiovascular diseases and stroke monitoring. Measuring the cIMT can help prevent and manage cardiovascular diseases (CDV) that may lead to stroke.	[12]	DL for segmentation, ML-based joint coefficient method	203	396 B-mode ultrasound images of the right and left common carotid artery	Error on test set DL1= 0.126 ± 0.134	DL2=0.124 ± 0.100mm
3	Develop an automated technique for cerebrospinal fluid (CSF) segmentation via integration of RF with geodesic active contour (GAC) segmentation	[18]	RF	38	Serial Computed Tomography (CT) images	paired t-test: p<10 ⁻³ in baseline p<10 ⁻⁵ in early follow-up	Early identification of patients at risk of developing malignant cerebral edema, which is a major complication following ischemic stroke.
4	Develop method and system for personalised modelling of Spatio/Spectro-Temporal data (SSTD) and prediction of events.	[40]	Spiking neural networks (SNN)	40	SSTD	ACC = 94%	Developed PMeSNNr framework provides much higher accuracy than conventional ML methods
5	Prediction of body mass index status from voice signals using ML	[45]	LR, bagging and RF	1,568	voice features	LR on imbalanced data AUC = 0.569-0.731	Estimating the body mass index (BMI) through voice signals can enhance early identification and detection of people at risk of having a stroke or other clinical conditions (e.g., CDV and diabetes).
6	An early diagnosis of atherosclerosis	[51]	ANN	55	B-mode ultrasound images	computation time per processed image: time = 1.4s	Assessing the intima-media thickness (IMT) of the common carotid artery (CCA) can contribute to

(Continued)

Table 3 (Continued)

Sr. No	Objective	Reference	ML-based approach	Samples	Features	Optimal Results	Clinical Implications
7	Accelerate early diagnosis and timely intervention for undiagnosed individuals with familial hypercholesterolaemia at high risk	[55]	RF	84,075	health-care encounter data, procedure and diagnostic codes, prescriptions, and laboratory findings	precision= 0.85, recall= 0.45, precision-recall curve= 0.55, ROC= 0.89,	the early diagnosis of atherosclerosis, which is crucial to stroke and heart attack prevention. Precocious diagnosis of familial hypercholesterolemia and implementation of adequate treatment.
8	Identify a ten-gene pattern of differential expression in peripheral blood which may have utility for detection of stroke.	[56]	kNN	23	genome-wide whole blood expression data	levels of sensitivity and specificity upwards of 90%	Identifying molecular biomarkers associated with stroke could aid its early recognition during patient triage.

Multi-Layer Perceptron (MLP), minimal learning machine, Linear Discriminant Analysis (LDA), and Structural Co-Occurrence Matrix (SCM). According to this study, SCM is the best to extract the most discriminant structural information concerning stroke subtype without parameters. These authors used 300 CT scans images (100 are normal, 100 are hemorrhagic stroke, and 100 are ischemic) that were obtained from a tomograph model GE Medical HiSpeed System (GEMSH). The SCM in the frequency domain presented high accuracy i.e. 98% in stroke classification. The study [60] aimed to extract the image features by analyzing CT scans brain tissue densities and used those features for stroke classification. Optimum Path Forest (OPF) classifier with Analysis of Brain Tissue Density (ABTD) feature selection algorithm got 99.30% of accuracy. Like earlier research, [68] worked on perfusion imaging and developed ANN model for identification of CT perfusion deficits with accuracy 85.8%. The CNN was used for anatomy segmentation of the brain lesions. Moreover, [28] studied raw perfusion MRI images to predict the final lesion in acute ischemic stroke using simple SVM technique and it demonstrated 95% accuracy. Since feature vectors size was quite large, in order to overcome dimensional problem, PCA was used and it preserved only the first principal vectors which allowed to explain 90% of total variance, contained in the training set. To access the generalizability of the classifier, K-fold cross-validation technique was implemented. Classifier SVM worked optimal because data points of binary classes are linearly separable. Ischaemic lesion segmentation is used to measure the damaged brain area(s); nonetheless, this procedure is time-consuming and less accurate when done manually. Study [39] presented deep supervised Fully Convolutional Network (FCN) to segment lesion using multimodal MRI imaging datasets with 70% accuracy.

Each dataset contained four modalities, namely T1, T2, DWI, and FLAIR. Intracerebral hemorrhage (ICH) is another one of the causes for high mortality rate therefore, [49] focused on multivariate analysis to predict hematoma expansion in spontaneous ICH with routinely available information using SVM and it scored 83.3% accuracy. Besides, randomized 179 search strategy was implemented for parameter tuning and recursive feature 180 elimination was used for feature selection in this study. Another important aspect is concerned with patient selection for thrombolytic procedures. In this regard, [5] CNN and ANN techniques were used for the prediction of functional thrombolysis outcomes in order to improve clinical decision making. Accuracy by CNN+ANN was 74% for the prediction of 'mRS90' (modified Rankin Scale (mRs) 0-1 in 90 days) and for 'NIHSS24' (National Institutes of Health Stroke Scale of greater than equals to 4 points in 24 hours) scored 71% accuracy.

Cerebral stroke is also caught an attention as a global health problem and if this could predict or classify automatically then that actually contributes to clinical

Table 4. Continued: Machine learning techniques for brain stroke Diagnosis. Note: Natural Language Processing (NLP), Gradient Boosting Machines (GBM), Logistic Regression (LR), recurrent neural network (RNNs), gated recurrent unit (GRU), Magnetic Resonance Imaging (MRI), modified Rankin Scale (mRS), density-based spatial clustering of applications (DBSCAN), Intracerebral Hemorrhage (ICH), Structural Co-Occurrence Matrix (SCM), Linear Discriminant Analysis (LDA)

Sr. No	Objective	Reference	ML-based approach	Samples	Features	Optimal Results	Clinical Implications
10	Segmentation of ischemic lesion using Fully Convolutional Network (FCN)	[39]	FCN	4,284	multimodal MRI images ISLES 2015 dataset	mean ACC= 70%	Enhanced diagnosis and management following ischemic stroke.
11	Classification of different walking conditions for hemiparetic subjects.	[43]	SVM	7	kinematic data by two sensors	ACC= 97.5%	Contribution to pathological gait recognition, analysis and monitoring during the performance of activities of daily living (ADL).
12	Identification of incorrect stroke outcome assessments using density-based outlier detection methods	[47]	DBSCAN, hierarchical DBSCAN (HDBSCAN) and local outlier factor (LOF).	1,00,000	4 different NINDS stroke registry datasets	DBSCAN: high mean ACC= 99.2 ± 0.7 at mRS of 4 and low ACC= 92.0 ± 4.6 at mRS of 3	
13	Prediction of hematoma expansion in spontaneous ICH from routinely available variables by using SVM	[49]	SVM	1,157	CT scan images	sensitivity= 81.3% specificity of 84.8% ACC= 83.3% AUC= 89%	Early identification of hematoma expansion after ICH and implementation of preventive or remediate treatments.
14	Prediction of stroke by hybrid ML technique on incomplete and imbalanced medical data.	[50]	RF for imputation, automated hyperparameter optimization (AutoHPO)	43,400	physiological data	AutoHPO: ACC= 71.6% sensitivity= 67.4%	
15	Classification of hemorrhagic and ischemic strokes by SCM technique using CT images of the brain	[57]	SCM, SVM, multilayer perceptron, minimal learning machine, LDA	300	CT images	SCM: specificity= 99.1% sensitivity= 97% FScore= 98% ACC= 98%	Enhanced stroke classification and diagnosis.
16	Propose novel approach to extract CT scan image features based on human brain tissue densities and classify stroke (healthy brain or ischemic or hemorrhagic	[60]	For feature extraction: Analysis of Brain Tissue Density (ABTD) For classification: MLP, SVM, kNN, OPF and Bayesian	420	CT images	ABTD with OPF: ACC= 99.30%	Early and more efficient stroke detection and diagnosis.

(Continued)

Table 4 (Continued)

Sr. No	Objective	Reference	ML-based approach	Samples	Features	Optimal Results	Clinical Implications
17	stroke) using ML techniques. To determine whether quantitative shape analysis of computed tomography angiography (CTA) can discriminate floating intraluminal thrombus from atherosclerotic plaque.	[67]	LR	23	CTA images	sensitivity= 87.5% specificity= 71.4%	Differentiation between intraluminal thrombus (FFT) and atherosclerotic plaque can impact treatment selection.
18	Predict the presence and laterality of a perfusion deficit on CT perfusion scans using ANN.	[68]	ANN	396	CT perfusion images	ACC= 85.8%	Early identification of perfusion deficits following ischemic stroke can facilitate further management and treatment.

diagnosis. Therefore, a step was taken by [50] and developed a ML model for stroke prediction using a hybrid ML approach on incomplete and imbalanced medical dataset. The RF was used for data imputation and automated hyperparameter optimization (AutoHPO) which is based on deep neural network (DNN) applied for stroke prediction. It obtained 71.6% accuracy and 67.4% sensitivity using 43,400 physiological data points. Another effort was made by [21] for classification of stroke types (healthy brain, an ischemic or a hemorrhagic stroke) from 420 CT scan images by employing Convolutional Neural Networks (CNN) with 100% accuracy. The posturography is a well-known technique to measure the postural control and this was investigated by [29] using time series of data of 38 center of pressure displacements (19 are post stroke adults and 19 are healthy). Findings of the paper were for fractal features 77.6% accuracy, for entropy features 68.4%, and for both features 76.3% accuracy. Likewise, joint disorders happen in hemiparetic patient after a stroke. Study [43] was investigated automated movement detection using SVM classification technique and kinematic sensors data. Total 7 patients walked in different conditions such as stair descent, stair ascent, level ground, downslope, and upslope. The SVM improved accuracy up to 97.5% and it performed superior than ANN and radial basis function neural networks (RBF). Manual classification of stroke sub types is not an easy task thus, to automatize the process of sub typing ischemic stroke, [25] used several advanced ML techniques with NLP of over 1,091 of EHR text records for the classification of TOAST (sub types of ischemic stroke) sub types. Finally, the authors compared their performances with a manual classification. In these experiments, the optimal performing techniques were stacking with LR and XGBoost, followed by GBT, RF, extra trees, SVM and finally KNN. Commonly used measures of clinical outcome after stroke are mRS and Barthel Index (BI). An alternate solution for the above mentioned measures was [47] developed using an unsupervised density-based outlier detection technique which identifies potentially incorrect measurements to assess the measurement quality without manual review.

Optimal classifier Density-based spatial clustering of applications with noise (DBSCAN) obtained 99.2% accuracy at mRs of 4 using 1,00,000 data points, data was gathered from stroke registry in Taiwan. Pneumonia is a general post-stroke consequence thus, the accurate and timely prediction of post-stroke pneumonia is crucial in clinical practice. Three classical ML techniques such as LR, SVM and XGBoost were used along with neural networks MLP and RNN i.e., attention based gated recurrent unit (GRU). Optimal classifier was attention-based GRU, achieved 0.928 AUC with 7 days of post-stroke and 0.905 within 14 days [26]. Ultimately, it leads DL models are more feasible for stroke management and achieve the best accuracy compared to classic ML Techniques. Besides, to discriminate between floating intraluminal thrombus

Table 5. Machine learning techniques for brain stroke Diagnosis. Note: Convolutional Neural Network (CNN), Electrocardiograph (ECG), Computed Tomography (CT), Artificial Neural Networks (ANN), Multivariate Pattern Analysis (MVPA), Random Forest (RF), Bayesian Classifier, Multilayer Perceptron, k-Nearest Neighbor, Random Forest and Support Vector Machines, Digital Imaging and Communications in Medicine (DICOM)

Sr. No	Objective	Reference	ML-based approach	Samples	Features	Optimal Results	Clinical Implications
1	To detect the electrocardiographic signature of Atrial Fibrillation (AF) present during normal sinus rhythm	[4]	CNN	1,80,922	ECG data	AUC= 0.90 Sensitivity= 82.3% specificity to 83.4% F1 score= 45.4% ACC= 83.3%	Proper screening of AF and management of patients with stroke due to an unknown cause, which can help prevent future strokes
2	To develop ML models for the prediction of National Institutes of Health Stroke Scale (NIHSS24) and modified Rankin Scale (mRS90)	[5]	CNN, ANN	204	clinical data CT brain scans	for NIHSS24: ACC=0.71 F1 score=0.74 for mRS90: ACC=0.74 F1 score=0.69	Facilitate the identification and selection of patients that could benefit from intravenous thrombolysis following acute ischemic stroke.
3	To identify disability-relevant spatial neglect based on connectivity.	[7]	MVPA	18	motor-attention network, connectivity data	sensitivity=0.85 specificity= 0.72 precision= 0.75 negative predictive value= 0.83	Important insights for developing combined spatial and motor rehabilitation to improve spatial neglect.
4	To detect undiagnosed post-stroke spasticity cases in United Kingdom	[19]	RF, CART, AdaBoost	45,613	healthcare resource, socioeconomic status, body mass index, age and number of previous stroke events	data-trained algorithm: RF= 75% clinician-trained algorithm: RF=72%	Enhanced identification of post-stroke spasticity could facilitate the management and treatment of this physical deficit.
5	Classification of types of stroke using CT scan images	[21]	CNN combined with RF, MLP, KNN, SVM (Linear), Bayes	420	CT images from DICOM	ACC=100%	Contribution to stroke identification and classification.
6	Classification of Ischemic Stroke Subtype using ML and NLP	[25]	KNN, RF, SVM, Extra Trees, GBM, XGBOOST, stacking using LR	1,091	text notes from electronic health records (EHR)	Kappa for stacking: radiology reports=0.25 clinical notes= 0.57 combined data=0.57	Determining ischemic stroke subtype will enhance outcome prediction and selection of patients for clinical trials (e.g., large-scale epidemiological studies).
7	Prediction of post-stroke pneumonia in the stroke population in China	[26]	LR, SVM, XGBoost, MLP and RNN (i.e., attention based GRU)	13,930	EHR data	within 7 days of post-stroke by GRU: AUC= 0.928	Early detection of post-stroke pneumonia will help to provide

(Continued)

Table 5 (Continued)

Sr. No	Objective	Reference	ML-based approach	Samples	Features	Optimal Results	Clinical Implications
8	Prediction of final lesion in stroke using	[28]	SVM	22,105	MRI images	precision= 95%	necessary treatment and to avoid severe outcomes. Facilitate treatment selection (i.e., pharmacological and non-pharmacological).
9	Identify Postural Control with Statistical methods and ML Decision Tree	[29]	DT	38	posturographic dataset	for fractal features: ACC= 77.6% for entropy features: ACC= 68.4% for both features: ACC= 76.3%	Enhanced diagnosis and treatment of patients with impaired balance control.

(FFT) from atherosclerotic plaque is significant and difficult therefore, [67] employed LR on CT angiography and identified 5 quantitative shape descriptors of carotid FFT. Atrial fibrillation is another clinical issue which can lead to stroke or heart-related complications. Hence, the research [4] was done on the identification of the patient with atrial fibrillation during sinus rhythm using CNN (83.3% accuracy, 45.4% F1 score) and 1,80,922 electrocardiograph data points. Spatial neglect neglect is a common post-stroke sequelae, which leads to asymmetric functional actions. Such patient normally has asymmetric functional actions. Multivariate pattern analysis classifier [7] classified disability-relevant spatial neglect based on 18 motor-attention network and connectivity data of stroke patients with 0.85 sensitivity and 0.72 specificity. In United Kingdom, stroke is one of the three highest mortality causes so [19] the RF, classification and regression tree, adaBoost techniques are applied on 45,613 UK primary care data points to explore potential under recording of post stroke spasticity.

Stroke Treatment

Table 6 follows the same structure like other tables mentioned in the Section 4.2 however, only 4 studies were found under this category. Several classification techniques were applied to modify the real-time display of virtual reality system to monitor a particular state of the patients based on their physiological reactions [6]. However, SVM with RBF achieved 91.43% of accuracy, shown in the Table 6. Improving the functional movements of the arm after stroke is significant part of stroke treatment, basically research idea [13] is to capture the linear acceleration and angular velocity of patient's Upper Extremity (UE) using sensor positioned on patient's wrists and this data is correctly classified with an average accuracy 94.80% in controls and 88.38% in stroke.

Moreover, [46] focuses to maintain and enhance the stroke patient's engagement at time of stroke rehabilitation practicals. The classifiers ANN and Naive Bayes (NB) were able to learn the association between the actual engagement and applied stimulation on 500 data points of Electromyography (EMG), Electroencephalography (EEG), eye movements and facial expression. Remote quantitative Fugl-Meyer Assessment (FMA) is proposed framework for the stroke patients by [71] which is in the clinical and home settings. Ensemble Machine Learning (ELM) was employed to learn the relation between the 2 accelerometer and 7 flex sensors data to clinical FMA scores.

Stroke Prognostication

Last but not least is stroke prognostication. Table 7 presents 9 studies on stroke diagnosis and its structure explained in the Section 4.2. Thrombolysis is a treatment to diffuse wild clots in blood vessels and improves flow of blood although, devastating complication in such

Table 6. Machine learning techniques for brain stroke treatment. Note: Perceptron Learning Algorithm (PLA), K-Center with Radial Basis Functions (RBF), Quadratic discriminant analysis (QDA), Linear discriminant analysis (LDA), Naïve Bayes (NB), Random Forest (RF), artificial neural network (ANN), multilayer perceptron neural network (MLPNN), Extreme Learning Machine (ELM), coefficient of determination (R^2), Activities of Daily Living Clinical Assessment Protocol (ADL-CAP)

Sr. No	Problem	Reference	ML-based approach	Features	Samples	Optimal Results	Clinical Implications
1	Present the classification model to adaptively and dynamically modify the therapy and real-time displays of a virtual reality system with the specific state of each patient using their physiological reactions.	[6]	PLA, KNN, LR, NB, SVM with RBF, Linear SVM, RBF QDA, LDA	physiological signals (relaxed, medium stress level, over-stressed)	7	SVM with RBF ACC= 91.43%	Robot-assisted therapy can be automatically adapted to the patient needs.
2	Method to separate upper extremity functional use from nonfunctional movement after stroke using a single wrist-worn accelerometer.	[13]	RF	sensor data	20	ACC= 91.53% for controls ACC= 70.18% for stroke	Important insights for the improvement of restorative and routine clinical treatments for hemiparesis.
3	Develop and validate of a cyber-physical stroke rehabilitation system (CP-SRS)	[46]	ANN, NB	EMG, EEG, eye movement, facial expression	500	ACC=NB>60%	Enhanced efficiency of rehabilitation exercises for stroke patients.
4	Propose remote quantitative Fugl-Meyer assessment (FMA) model to predict the FMA scores	[71]	ELM, RRelief (for feature selection)	wearable sensor data	24	$R^2= 0.917$	Remote assessment and rehabilitation of various mobility-related functional activities.

Table 7. Machine learning techniques for brain stroke prognostic or outcome prediction. Note: Machine Learning (ML), Computerized Tomography (CT), Area Under receiver-operating-characteristic Curve (AUC), Artificial Neural Network (ANN) and Support Vector Machine (SVM), Residual Neural Network (ResNet), Structured Receptive Fields (RFNN), auto-encoders (AE), Gaussian Process model Regression (GPR), Corticospinal Tract (CST), Regions Of Interest (ROIs)

Sr. No	Objective	Reference	ML-based approach	Samples	Features	Optimal Results	Clinical Implications
1	Identify factors influencing outcome in brain arteriovenous malformations (BAVM) treated with endovascular embolization	[3]	ANN, SVM	199	demographics and clinical details	ACC= 97.5%	Determination of patients with brain arteriovenous malformations that can benefit from endovascular treatments and implementation of individualized interventions.
2	Predict thrombolysis-associated haemorrhage	[10]	SVM	116	CT brain images and clinical variables	AUC= 0.744	Differentiation between patients that will positively and negatively respond to thrombolysis can impact treatment selection.
3	Present a novel scoring system for predicting functional outcome after ICH at 3 and 12 months follow-up	[32]	RF for feature selection, LR	575	intracerebral Hemorrhage Outcomes	for predicting functional status: AUC= 0.89 (3 months) AUC= 0.87 (12 months)	Comprehensive prediction and higher discrimination of functional outcomes in patients with ICH.
4	Predict good reperfusion after endovascular treatment (EVT) and good functional outcome	[35]	ResNet with RFNN	1,301	CT angiography images.	AUC= 0.71	Enhanced functional outcome prediction after acute ischemic stroke can potentially improve treatment selection (e.g., to only consider patients with a good response to endovascular treatment).
5	Predict speech production skills in stroke patients.	[36]	GPR	270	cross-sectional, longitudinal patient data	$R^2 = 0.84$	Realistic long-term behavioral outcome predictions after stroke (e.g., severity of speech production impairments) with the purpose of answering stroke patients' expectations, i.e., when and to what extent they should expect to improve.
6	Predict activities of daily living (Barthel index as the indicator) is crucial for post-stroke patients.	[48]	LR, SVM, RF	313	assessments	for BI status:LR and RF AUC= 0.79 for BI score: SVM, MAE= 9.86 LR, MAE= 9.95	Realistic goal setting and development of a specific and individual discharge plan due to the accurate prediction of

Table 7 (Continued)

Sr. No	Objective	Reference	ML-based approach	Samples	Features	Optimal Results	Clinical Implications
7	Classify patients into good vs. poor functional outcome	[54]	SVM	297	emergency stroke image data	median [IQR] ACC=82.8%	patients' ability to perform ADL. Cost-benefit analysis of thrombolysis and thrombectomy treatment at the acute stage of stroke.
8	Predict post-stroke long term upper limb motor function outcome that can contribute to improve clinical efficacy.	[62]	GPR	50	MRI images	based on different masks to limit subsets of voxels: R= 0.83, RMSE= 0.68	Enhanced prediction of long-term motor outcomes following stroke.
9	Predict motor and cognitive improvement for rehab treatment in the early stage of stroke	[65]	linear SVM	55	functional and clinical data	RMSE: 1. T1 Barthel= 23.38 2. T1 motor FIM= 14.99 3. T1 cognitive FIM= 3.00 4. T1 FIM total= 19.78	Greater understanding of the dynamics of inflammatory biomarkers during rehabilitation may lead to the implementation of innovative treatments.

treatment is Symptomatic intracranial hemorrhage (SICH) can lead mortality. The SVM was used for the prediction of SICH using CT scan images of 116 acute ischaemic stroke patients and has obtained 0.744 AUC. The performance of the SVM was compared with SEDAN and HAT scores (prognostication tools). Similarly, RF was employed [32] to identify factors or feature selection from 575 intracerebral hemorrhage (ICH) outcomes and developed LR model for prognosticating functional status (mRS score, 0 to 3 is good and mRS score, 4 to 6 is poor) based on these factors. LR model reported 0.89 AUC for 3 months and 0.87 for 12 months. Generated results were then compared with ICH score. Brain arteriovenous malformation consist of a set of abnormal blood vessels in the brain. Research [3] focused on identification of outcome which influences outcome in the brain. Two approaches ANN and SVM were employed on 199 demographics an clinical data points and recorded 97.5% accuracy. Moreover, DL approach, structured receptive field neural networks-residual neural network was applied on 1,301 CT angiography images for predicting outcomes of acute ischemic stroke patients [35], reported AUC= 0.71. It got higher accuracy than radiological image biomarkers for outcome prediction.

Besides, MRI images were used to extract lesion information which were injected to Gaussian process model regression [36] for the prediction of behavioural outcomes and recovery of speech production after stroke. The LR, SVM, and RF techniques were applied on 15 rehabilitation assessments from 313 individuals for the prediction of Barthel Index (BI) status and score at discharge of post-stroke patients. Specifically, LR, SVM, and RF were used for BI status and SVM and LR used for BI scores. Stroke topography has widely been mapped with functional outcome using manual lesion segmentation. Therefore, proposed [54] powerful ML based technique (SVM) that improved median accuracy (82.8%) using 297 stroke images as well as provided an important insight on critical brain areas for long-term functional outcome (good or poor). Similarly, linear SVM employed [65] on functional and clinical data of 55 individuals. The authors investigated the exact role of common inflammatory biomarkers in the prediction of motor and cognitive improvement after rehabilitation in the early stages of stroke. Eventually, improved ML or DL based predictions of stroke outcome enhance clinical efficacy. Gaussian Process Regression [62] was used to decode post-stroke motor function from 50 structural brain images of chronic stroke patients and results are showed in the Table 7.

Discussion

In this work, we discussed many stroke related problems from the state-of-art. The reviewed studies were grouped in several categories based on their similarities. We note that it is difficult to compare studies as they

employed different performance metrics for different tasks, considering different datasets, techniques, and tuning parameters. Hence, we only mention the research areas which were targeted in more than one study and the studies which report highest classification accuracy in each section. The stroke prevention category contains 2 studies on carotid-artery atherosclerosis and the stroke diagnosis category consists of 5 studies on classification of stroke types. We identified that 4 out of 39 studies are focused on regression. Hence, the current review has more articles on classification, we present classification accuracy of an optimal model from each category. For example, [60] has achieved the highest 99.30% accuracy in stroke diagnosis. For stroke prevention, [21] mentioned 100% accuracy for classification of ischemic stroke sub types. If we see the results in stroke treatment category, [6] got optimal result, accuracy 91.43% and in prognostication, [3] received 97.5% accuracy. Figure 2 signifies 3 layers, consists of stroke category, dataset and outer layer represents optimal model of particular study. Here, we

try to visualize 39 studies in one figure to get some valuable insights or conclusions.

1. After reviewing all studies, we found 4 main categories of stroke problem such as prevention, diagnosis, treatment and prognostication, mentioned in the first ring of the Figure 2.
2. The second ring comprises names of the dataset of each category. It shows 7 studies from CT images under stroke diagnosis, 2 from prognostication category, and 1 from prevention. After that, MRI images are used in 5 studies, each 2 from diagnosis and prognostication, 1 from prevention, 2 times sensor data and ultrasound images. ML powered by DL for the task of image processing, has emerged rapidly in recent years. Consequently, CT scan image is the most frequently used data in a stroke field.
3. Finally, the third ring presents an optimal model of each study. 10 studies appeared with SVM as optimal model, then RF (6 times), LR (4 times), ANN and CNN (each 3 times) from 4 categories. Thus, SVM is a



Fig. 2. Multi-level pie chart of ML approaches and types of data used for stroke prevention, stroke diagnosis, treatment, and prognostication.

commonly used technique as well as highly efficient to map a data in a stroke field.

4. SVM and RF are applied and obtained as an optimal models in at least 1 study from each category.
5. Since 18 studies are found in the stroke diagnosis category, the rest of the categories require to implement more research ideas. Specifically in stroke treatment.
6. In this review, the majority of research problems are classification and only 4 studies are related to regression. It seems general nature of problem in stroke field is classification.

This review was intended to showcase ML approaches employed on various datasets used for solving various stroke problems for better healthcare system as well as further investigation. We believe that it allows a better understanding and valuable insights into the effectiveness of ML-based stroke problems in enhancing medical practices.

Acknowledgment: We want to thank the BRaNT team for their support and stimulating discussions. This research is supported by BRaNT - Belief Revision applied to Neurorehabilitation Therapy [project number PTDC/CCI-COM/30990/2017], financed by FCT - Fundação o para a Ciência e a Tecnologia. EF is partially supported by UID/CEC/04516/2019. SBB is partially supported by MACBIOIDI: Promoting the cohesion of Macaronesian regions through a common ICT platform for biomedical R - D - i (INTERREG program MAC/1.1.b/098)

References

1. Alawadi S, Fernández-Delgado M, Mera D, Barro S. Polynomial kernel discriminant analysis for 2d visualization of classification problems. *Neural Computing and Applications* 2019;31(8):3515-3531.
2. Almeida Y, Sirsat MS, i Badia SB, Fermé E. Ai-rehab: A framework for ai driven neurorehabilitation training-the profiling challenge. *Healthinf*; 2020845-853.
3. Asadi H, Kok HK, Looby S, Brennan P, O'Hare A, Thornton J. Outcomes and complications after endovascular treatment of brain arteriovenous malformations: a prognostication attempt using artificial intelligence. *World neurosurgery* 2016;96:562-569.
4. Attia ZI, Noseworthy PA, Lopez-Jimenez F, Asirvatham SJ, Deshmukh AJ, Gersh BJ, Carter RE, Yao X, Rabinstein AA, Erickson BJ, et al. An artificial intelligence-enabled ECG algorithm for the identification of patients with atrial fibrillation during sinus rhythm: a retrospective analysis of outcome prediction. *The Lancet* 2019;394(10201):861-867.
5. Bacchi S, Zerner T, Oakden-Rayner L, Kleinig T, Patel S, Jannes J. Deep learning in the prediction of ischaemic stroke thrombolysis functional outcomes: A pilot study. *Academic radiology* 2019.
6. Badesa FJ, Morales R, Garcia-Aracil N, Sabater JM, Casals A, Zollo L. Auto-adaptive robot-aided therapy using machine learning techniques. *Computer methods and programs in biomedicine* 2014;116(2):123-130.
7. Barrett AM, Boukrina O, Saleh S. Ventral attention and motor network connectivity is relevant to functional impairment in spatial neglect after right brain stroke. *Brain and cognition* 2019;129:16-24.
8. Belagaje SR. Stroke rehabilitation. *CONTINUUM: Life-long Learning in Neurology* 2017;23(1):238-253.
9. Benjamin EJ, Virani SS, Callaway CW, Chamberlain AM, Chang AR, Cheng S, Chiuve SE, Cushman M, Delling FN, Deo R, et al. Heart disease and stroke statistics-2018 update: a report from the american heart association. *Circulation* 2018;137(12):e67.
10. Bentley P, Ganesalingam J, Jones ALC, Mahady K, Epton S, Rinne P, Sharma P, Halse O, Mehta A, Rueckert D. Prediction of stroke thrombolysis outcome using CT brain machine learning. *NeuroImage: Clinical* 2014;4:635-640.
11. Bento M, Souza R, Salluzzi M, Rittner L, Zhang Y, Frayne R. Automatic identification of atherosclerosis subjects in a heterogeneous MR brain imaging data set. *Magnetic resonance imaging* 2019.
12. Biswas M, Kuppili V, Araki T, Edla DR, Godia EC, Saba L, Suri HS, Omerzu T, Laird JR, Khanna NN, et al. Deep learning strategy for accurate carotid intima-media thickness measurement: an ultrasound study on japanese diabetic cohort. *Computers in biology and medicine* 2018;98:100-117.
13. Bochniewicz EM, Emmer G, McLeod A, Barth J, Dromerick AW, Lum P. Measuring functional arm movement after stroke using a single wrist-worn sensor and machine learning. *Journal of Stroke and Cerebrovascular Diseases* 2017;26(12):2880-2887.
14. Breiman L. Random forests. *Machine learning* 2001;45(1):5-32.
15. Brenner DA, Zweifler RM, Gomez CR, Kissela BM, Levine D, Howard G, Coull B, Howard VJ. Awareness, treatment, and control of vascular risk factors among stroke survivors. *Journal of stroke and cerebrovascular diseases* 2010;19(4):311-320.
16. Campbell BCV, De Silva DA, Macleod MR, Coutts SB, Schwamm LH, Davis SM, Donnan GA. Ischaemic stroke. *Nature Reviews Disease Primers* 2019;5(1):1-22.
17. Carlo AD. Human and economic burden of stroke. *Age and Aging* 2009;38:4-5. <https://doi.org/10.1093/ageing/afn282>.
18. Chen Y, Dhar R, Heitsch L, Ford A, Fernandez-Cadenas I, Carrera C, Montaner J, Lin W, Shen D, An H, et al. Automated quantification of cerebral edema following hemispheric infarction: application of a machine-learning algorithm to evaluate CSF shifts on serial head CTs. *NeuroImage: Clinical* 2016;12:673-680.
19. Cox AP, Raluy-Callado M, Wang M, Bakheit AM, Moore AP, Dinot J. Predictive analysis for identifying potentially undiagnosed post-stroke spasticity patients in united kingdom. *Journal of biomedical informatics* 2016;60:328-333.
20. van der ML, Hinton G. Visualizing data using t-SNE. *Journal of machine learning research* 2008;9(Nov):2579-2605.
21. Dourado Jr CM, da Silva SPP, da Nóbrega RVM, Barros ACdS, Rebouças Filho PP, de Albuquerque VHC. Deep learning iot system for online stroke detection in skull computed tomography images. *Computer Networks* 2019;152:25-39.
22. Duda RO, Hart PE, Stork DG. *Unsupervised Learning and Clustering. Pattern classification (2nd ed.)*. Wiley; 2001.
23. Fernández-Delgado M, Cernadas E, Barro S, Amorim D. Do we need hundreds of classifiers to solve real world classification problems? *Journal of Machine Learning Research* 2014;15:3133-3181.
24. Fernández-Delgado M, Sirsat MS, Cernadas E, Alawadi S, Barro S, Febrero-Bande M. An extensive experimental

- survey of regression methods. *Neural Networks* 2019;111:11-34. <https://doi.org/10.1016/j.neunet.2018.12.010>.
25. Garg R, Oh E, Naidech A, Kording K, Prabhakaran S. Automating ischemic stroke subtype classification using machine learning and natural language processing. *Journal of Stroke and Cerebrovascular Diseases* 2019;28(7):2045-2051.
 26. Ge Y, Wang Q, Wang L, Wu H, Peng C, Wang J, Xu Y, Xiong G, Zhang Y, Yi Y. Predicting post-stroke pneumonia using deep neural network approaches. *International Journal of Medical Informatics* 2019;132:103986.
 27. Géron A. *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: Concepts, tools, and techniques to build intelligent systems*. O'Reilly Media; 2019.
 28. Giacalone M, Rasti P, Debs N, Frindel C, Cho T-H, Grenier E, Rousseau D. Local spatio-temporal encoding of raw perfusion MRI for the prediction of final lesion in stroke. *Medical image analysis* 2018;50:117-126.
 29. Giovanini LHF, Silva SM, Manffra EF, Nievola JC. Sampling and digital filtering effects when recognizing postural control with statistical tools and the decision tree classifier. *Procedia Computer Science* 2017;108:129-138.
 30. Good DC, Bettermann K, Reichwein RK. Stroke rehabilitation. *CONTINUUM: Lifelong Learning in Neurology* 2011;17(3):545-567.
 31. Greenberg DA, Aminoff MJ, Simon RP. *Clinical neurology*. McGraw-Hill; 2002.
 32. Gupta VP, Garton ALA, Sisti JA, Christophe BR, Lord AS, Lewis AK, Frey H-P, Claassen J, Connolly Jr ES. Prognosticating functional outcome after intracerebral hemorrhage: the ICHOP score. *World neurosurgery* 2017;101:577-583.
 33. Gurd JM, Kischka U, Marshall JC. *Handbook of clinical neuropsychology*. Oxford University Press, USA; 2010.
 34. heart, N., lung, & institute, b. (2015). *Stroke*. <https://www.nhlbi.nih.gov/health-topics/stroke>. [Online; accessed on 24-Jan-2020].
 35. Hilbert A, Ramos LA, van Os H, Olabarriaga SD, Tolhuisen ML, Wermer M, Barros RS, van der Schaaf I, Dippel D, Roos Y, et al. Data-efficient deep learning of radiological image data for outcome prediction after endovascular treatment of patients with acute ischemic stroke. *Computers in biology and medicine* 2019:103516.
 36. Hope TMH, Seghier ML, Leff AP, Price CJ. Predicting outcome and recovery after stroke with lesions extracted from MRI images. *NeuroImage: clinical* 2013;2:424-433.
 37. John GH, & Langley P. (2013). Estimating continuous distributions in bayesian classifiers. *arXiv preprint: 1302.4964*
 38. Johnson W, Onuma O, Owolabi M, Sachdev S. Stroke: a global response is needed. *Bulletin of the World Health Organization* 2016;94. <https://doi.org/10.2471/BLT.16.181636.634-634A>
 39. Karthik R, Gupta U, Jha A, Rajalakshmi R, Menaka R. A deep supervised approach for ischemic lesion segmentation from multimodal MRI using fully convolutional network. *Applied Soft Computing* 2019;84:105685.
 40. Kasabov N, Feigin V, Hou Z-G, Chen Y, Liang L, Krishnamurthi R, Othman M, Parmar P. Evolving spiking neural networks for personalised modelling, classification and prediction of spatio-temporal patterns with a case study on stroke. *Neurocomputing* 2014;134:269-279.
 41. Landwehr N, Hall M, Frank E. Logistic model trees. *Machine learning* 2005;59(1-2):161-205.
 42. Langhorne P, Bernhardt J, Kwakkel G. Stroke rehabilitation. *The Lancet* 2011;377(9778):1693-1702.
 43. Lau H-y, Tong K-y, Zhu H. Support vector machine for classification of walking conditions of persons after stroke with dropped foot. *Human movement science* 2009;28(4):504-514.
 44. LeCun Y, Bengio Y, Hinton G. Deep learning. *nature* 2015;521(7553):436.
 45. Lee BJ, Kim KH, Ku B, Jang J-S, Kim JY. Prediction of body mass index status from voice signals based on machine learning for automated medical applications. *Artificial intelligence in medicine* 2013;58(1):51-61.
 46. Li C, Rusák Z, Horváth I, Ji L. Development of engagement evaluation method and learning mechanism in an engagement enhancing rehabilitation system. *Engineering Applications of Artificial Intelligence* 2016;51:182-190.
 47. Lin C-H, Hsu K-C, Johnson KR, Luby M, Fann YC. Applying density-based outlier identifications using multiple datasets for validation of stroke clinical outcomes. *International journal of medical informatics* 2019;132:103988.
 48. Lin W-Y, Chen C-H, Tseng Y-J, Tsai Y-T, Chang C-Y, Wang H-Y, Chen C-K. Predicting post-stroke activities of daily living through a machine learning-based approach on initiating rehabilitation. *International journal of medical informatics* 2018;111:159-164.
 49. Liu J, Xu H, Chen Q, Zhang T, Sheng W, Huang Q, Song J, Huang D, Lan L, Li Y, et al. Prediction of hematoma expansion in spontaneous intracerebral hemorrhage using support vector machine. *EBioMedicine* 2019;43:454-459.
 50. Liu T, Fan W, Wu C. A hybrid machine learning approach to cerebral stroke prediction based on imbalanced medical dataset. *Artificial Intelligence in Medicine* 2019:101723.
 51. Menchón-Lara R-M, Sancho-Gómez J-L. Fully automatic segmentation of ultrasound common carotid artery images based on machine learning. *Neurocomputing* 2015;151:161-167.
 52. Merino, J. G. (2014). *Clinical stroke challenges: A practical approach*.
 53. Moskowitz MA, Lo EH, Iadecola C. The science of stroke: mechanisms in search of treatments. *Neuron* 2010;67(2):181-198.
 54. Moulton E, Valabregue R, Lehericy S, Samson Y, Rosso C. Multivariate prediction of functional outcome using lesion topography characterized by acute diffusion tensor imaging. *NeuroImage: Clinical* 2019;23:101821.
 55. Myers KD, Knowles JW, Staszak D, Shapiro MD, Howard W, Yadava M, Zuzick D, Williamson L, Shah NH, Banda JM, et al. Precision screening for familial hypercholesterolemia: a machine learning study applied to electronic health encounter data. *The Lancet Digital Health* 2019.
 56. O'Connell GC, Chantler PD, Barr TL. Stroke-associated pattern of gene expression previously identified by machine-learning is diagnostically robust in an independent patient population. *Genomics data* 2017;14:47-52.
 57. Peixoto SA, Rebouças Filho PP. Neurologist-level classification of stroke using a structural co-occurrence matrix based on the frequency domain. *Computers & Electrical Engineering* 2018;71:398-407.
 58. Pendlebury ST, Rothwell PM. Prevalence, incidence, and factors associated with pre-stroke and post-stroke dementia: a systematic review and meta-analysis. *The Lancet Neurology* 2009;8(11):1006-1018.
 59. Postma EO, van den Herik HJ, van der Maaten LJ. Dimensionality reduction: a comparative review. *Journal of Machine Learning Research* 2009;10(1-41):66-71.
 60. Rebouças Filho PP, Sarmiento RM, Holanda GB, de Alencar Lima D. New approach to detect and classify stroke in skull

- CT images via analysis of brain tissue densities. *Computer methods and programs in biomedicine* 2017;148:27-43.
61. Reid JM, Gubitz GJ, Dai D, Kydd D, Eskes G, Reidy Y, Christian C, Counsell CE, Dennis M, Phillips SJ. Predicting functional outcome after stroke by modelling baseline clinical and CT variables. *Age and ageing* 2010;39(3):360-366.
 62. Rondina JM, Filippone M, Girolami M, Ward NS. Decoding post-stroke motor function from structural brain imaging. *NeuroImage: Clinical* 2016;12:372-380.
 63. Russ JC. *The image processing handbook*. CRC press; 2016.
 64. Sacco RL, Kasner SE, Broderick JP, Caplan LR, Connors JJ, Culebras A, Elkind MSV, George MG, Hamdan AD, Higashida RT, et al. An updated definition of stroke for the 21st century: a statement for healthcare professionals from the american heart association/american stroke association. *Stroke* 2013;44(7):2064-2089.
 65. Sale P, Ferriero G, Ciabattini L, Cortese AM, Ferracuti F, Romeo L, Piccione F, Masiero S. Predicting motor and cognitive improvement through machine learning algorithm in human subject that underwent a rehabilitation treatment in the early stage of stroke. *Journal of Stroke and Cerebrovascular Diseases* 2018;27(11):2962-2972.
 66. Smith WS, Sung G, Saver J, Budzik R, Duckwiler G, Liebeskind DS, Lutsep HL, Rymer MM, Higashida RT, Starkman S, et al. Mechanical thrombectomy for acute ischemic stroke: final results of the multi MERCI trial. *Stroke* 2008;39(4):1205-1212.
 67. Thornhill RE, Lum C, Jaber A, Stefanski P, Torres CH, Momoli F, Petrich W, Dowlatshahi D. Can shape analysis differentiate free-floating internal carotid artery thrombus from atherosclerotic plaque in patients evaluated with CTA for stroke or transient ischemic attack? *Academic radiology* 2014;21(3):345-354.
 68. Vargas J, Spiotta A, Chatterjee AR. Initial experiences with artificial neural networks in the detection of computed tomography perfusion deficits. *World neurosurgery* 2019;124:e10-e16.
 69. World Health Organization, G. (2018). Global health estimates 2016: deaths by cause, age, sex, by country and by region. https://www.who.int/healthinfo/global_burden_disease/estimates/en/. [Online; accessed on 11-Dec-2019].
 70. Xu R, Wunsch D. *Clustering*. John Wiley & Sons; 2008.
 71. Yu L, Xiong D, Guo L, Wang J. A remote quantitative fughl-meyer assessment framework for stroke patients based on wearable sensor networks. *Computer methods and programs in biomedicine* 2016;128:100-110.