International Journal on Advanced Science Engineering Information Technology

Vol.14 (2024) No. 5
ISSN: 2088-5334

Machine Learning-Based Stroke Prediction: A Critical Analysis

Agus Byna [a,b], Muhammad Modi Lakulu [a,*], Ismail Yusuf Panessai [a], Nurhaeni [b]

[a] *Faculty of Computing and Meta-Technology, Universiti Pendidikan Sultan Idris, Perak, Malaysia*
[b] *Faculty of Science and Technology, Universitas Sari Mulia, Kalimantan Selatan, Indonesia*
Corresponding author: [*]modi@meta.upsi.edu.my

***Abstract*— Stroke is a critical public health issue that frequently has long-term impairment and negative effects. Devising innovative methods that enable timely and accurate identification and intervention is crucial. In this regard, machine learning (ML) and deep learning (DL) approaches of artificial intelligence (AI) play a crucial role in reducing the incidence of strokes. This study systematically analyzed articles from 2012 to 2022 using the PRISMA Method. PRISMA is a tool that facilitates researchers' access to an online platform for self-directed learning. The cumulative quantity of articles gathered for ten years reached 1405 from five databases. However, only 79 relevant articles were used for identification. The main objective was to provide a thorough taxonomy that classifies using and implementing machine learning approaches for stroke prediction. The results of this experiment confirm that machine-learning techniques have a great deal of potential for accurate stroke prediction. Nevertheless, challenges such as biased data and algorithms and the need for models that can be adjusted to accommodate various demographics and healthcare systems continue to exist. It is essential to recognize the need for additional research projects that thoroughly explore potential data biases, algorithmic biases, and the generalizability of models across various demographics and healthcare systems. More research is necessary to further the literature on the complete assessment of machine learning models in precisely forecasting stroke occurrences.**

***Keywords*— Stroke; artificial intelligence; machine learning; deep learning; prediction.**

Manuscript received 18 Sep. 2023; revised 16 May. 2024; accepted 4 Aug. 2024. Date of publication 31 Oct. 2024.

I. Introduction

The healthcare industry is subject to constant transformation because of continued progress in the fields of medical science and technology [1]. Stroke, a considerable health concern, has the potential to result in long-term incapacity and serious consequences [2]. Consequently, innovative approaches are required to tackle this enduring problem effectively [3]. The complexity of this task is exacerbated by delays in diagnosis and a limited comprehension of the numerous facets of the disease [4], [5]. Therefore, it is crucial to devise innovative methods that enable timely and accurate identification and intervention [6]. In the context of this specific setting, the realm of artificial intelligence (AI) emerges as a highly promising and innovative discipline [7], [8].

In contemporary times, there has been an increasing emphasis on the responsibility of healthcare personnel in the detection and recognition of individuals who have encountered a stroke [9]. The emphasis on good communication of stroke-related information among healthcare practitioners is driven by the knowledge that such communication issues can result in delays in the timely beginning of appropriate therapy [10]. The study highlights the significance of prompt recognition by healthcare practitioners to reduce the incidence of strokes [11]. The prompt detection of a medical issue can exert a substantial influence on the healing process and result in enhanced treatment results [10], [12]. Numerous studies have been undertaken to explore the various components that contribute to stroke, making it a disease of significant complexity that necessitates comprehensive investigation [13], [14], [15].

The utilization of machine learning (ML) and deep learning (DL) methodologies has significant promise in revolutionizing the stroke detection process, hence bringing about a profound transition in healthcare procedures [16]Previous studies have demonstrated a need for prompter evaluation of appropriate management approaches. Therefore, the utilization of artificial intelligence (AI) techniques, particularly machine learning (ML) and deep learning (DL), plays a crucial role in mitigating strokes.

Although patients may receive therapies that are deemed unnecessary during treatment, it is expected that effective approaches would be employed to minimize the resulting consequences. In a recent work [17], researchers employed both methodologies to predict strokes using a data-driven strategy. Nevertheless, the assessment conducted by [18] that underscores the absence of validation for current machine learning-based predictive models using separate datasets. Furthermore, it is worth noting that more research reports need to specifically examine the decision-analytic processes involved in evaluating the clinical usefulness of prediction models.

The subsequent discovery in the study conducted by [19] maintains its emphasis on four key areas of stroke care: stroke prevention, diagnosis, therapy, and prognosis. Nevertheless, the categories above still need to encompass the clinical application of machine learning models in the context of stroke therapy. Similarly, the research undertaken by [20] elucidates that the range of features employed spans from 5 to 200, exhibiting notable overlap in variables. Nevertheless, it is worth noting that a significant deficiency exists in the attention given to external validation and the practicality of the findings in real-world scenarios.

Additional investigation is required to conduct literature evaluations that explicitly examine the precision of machine learning (ML) and deep learning (DL) models in forecasting stroke illness. While several research have conducted systematic evaluations of algorithmic models and their efficacy when used in combination with other statistical techniques, only a limited number of studies have specifically assessed the usefulness of these models in healthcare applications [21]. Considering the necessity for a more methodical understanding of the role of Stroke Prediction in healthcare, it is imperative to emphasize the importance of improved communication and information dissemination to patients [22]. As a result, these matters have been given minimal consideration in scholarly publications.

The main objective of this study is to conduct a comprehensive assessment and investigation of the possible impact of artificial intelligence, specifically machine learning (ML) and deep learning (DL), on the transformation of stroke diagnosis and the improvement of healthcare results. This paper aims to conduct a thorough analysis to evaluate prediction techniques for ischemic and hemorrhagic strokes. Moreover, the primary objective of this study is to construct a comprehensive taxonomy that classifies the utilization and implementation of machine learning (ML) and deep learning (DL) techniques within the domain of stroke prediction. The taxonomy presented in this study encompasses five fundamental categories: the development of reliable prediction models, the formulation of strategic plans for systems, the meticulous evaluation of models, the comprehensive comparison analysis, and the thorough review of findings.

II. Materials and Method

The present study adhered to a systematic review methodology, encompassing many stages such as identifying pertinent information sources, selecting studies, establishing search and eligibility criteria, implementing data collection processes, and analyzing taxonomies. These steps were undertaken to address the research inquiries about the subject matter. Furthermore, this study employed a systematic review technique, utilizing the PRISMA protocol, which is a tool that facilitates researchers' access to an online platform for self-directed learning [23], [24].This study outlines the specific criteria used to determine which papers are included or excluded in the dissemination process [25]. The search database encompasses significant data on applying machine learning techniques to stroke prediction.

A. Search and Analysis Strategy

The search database contains essential information regarding applying machine learning techniques to predict strokes. The current investigation analyzed diverse scientific literature in the database, focusing mainly on artificial intelligence and its integral elements, specifically machine learning and deep learning. These studies' central emphasis was on using artificial intelligence methodologies in stroke prediction. The present study utilized five databases: Scopus, ScienceDirect, Springer, PubMed, and the IEEE Xplore digital library. The databases were chosen based on their ability to handle straightforward and intricate queries. The databases mainly consist of computer science, artificial intelligence, and health science magazines with conference papers. This comprehensive compilation encompasses diverse social and scientific contributions across different fields, providing a holistic representation of the research endeavors conducted by experts. To ensure a thorough analysis of the available literature, the process of selecting pertinent research employed a two-round iterative approach, as outlined in the source above [25].

The procedure was conducted on February 16, 2022, utilizing the search functionality of each respective database. The search methodology encompassed the utilization of targeted terms as search queries, such as "stroke," "(predict* or classify*)," "artificial intelligence," "machine learning," and "deep learning," within a temporal scope ranging from 2012 to 2022. The strategic search was conducted through a series of four consecutive steps. Initially, data was extracted from the five databases dedicated to scholarly publications. Furthermore, duplicate titles were filtered away to speed up the review process.

The eligibility was determined by meticulously examining all papers' titles, abstracts, methods, findings, and comments to verify their conformity with the established inclusion criteria and research objectives. The ultimate stage involved the examination of the complete articles to permit a thorough analysis of each item. The technique utilized in this study facilitated a comprehensive and meticulous assessment of each publication, entailing a meticulous examination of its methodology and findings. Consequently, the study's credibility was augmented. The investigation's findings were examined comprehensively to identify any observable trends, ultimately contributing to developing a clearly defined taxonomy.

III. Results and Discussion

Using the PRISMA framework, the preliminary findings obtained from the five databases were as follows: Scopus contributed 524 articles, ScienceDirect gave 225 articles, Springer supplied 510 articles, PubMed presented 14 items, and IEEE contributed 135 articles. As a result, the cumulative quantity of articles gathered for ten years reached 1405. After proceeding to the second stage, the total number of remaining articles was 1282. Among these, 123 articles were found to have identical titles and were therefore eliminated from further analysis. Upon reaching the third stage of the research process, the number of articles amounted to 270, as 1012 articles were deemed outside the predetermined scope of the desired area.

Ultimately, the conclusion of the procedure yielded a total of 79 articles, while 182 articles were rendered inaccessible owing to various problems such as paywalls, restricted access, or unavailability of sources. Figure 1 shows the selection process using the PRISMA method.

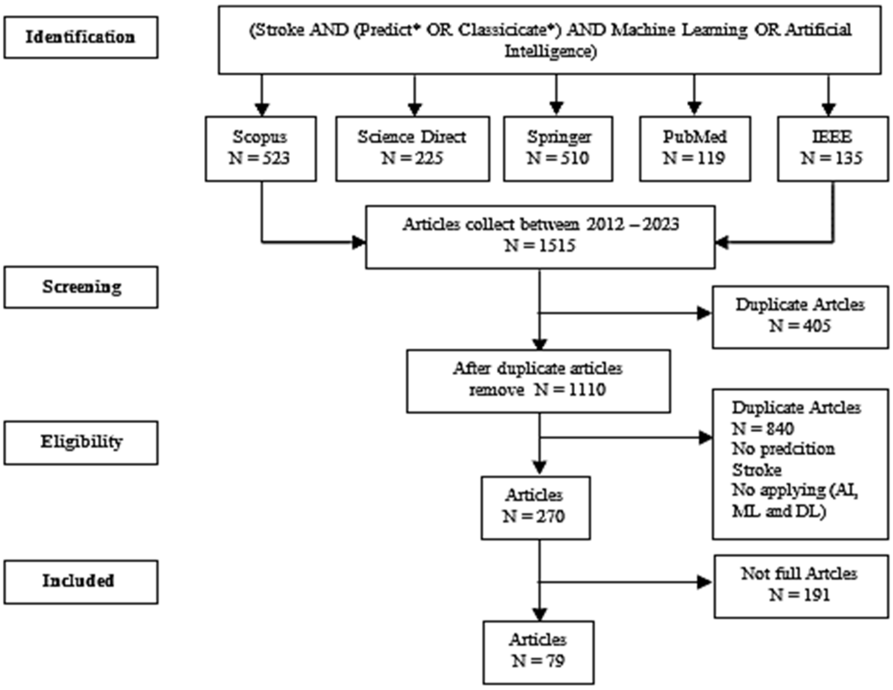

Fig. 1 Result finding articles using PRISMA

A. *Results according to demographics*

Figure 2 depicts the demographic distribution, as expounded upon in this study. Most of the 79 publications were sourced from reputable academic databases such as IEEE Xplore, Science Direct, and Springer. The early findings demonstrate the distribution of publications from various sources: 32% from IEEE Xplore, 24% from Science Direct, 23% from Springer, 15% from Scopus, and 6% from PubMed. The IEEE produced the largest number of publications, with 25 articles (32%). In contrast, PubMed had the lowest number of articles, with just 5 (6%). This discrepancy could be attributed to the health-centric nature of PubMed.

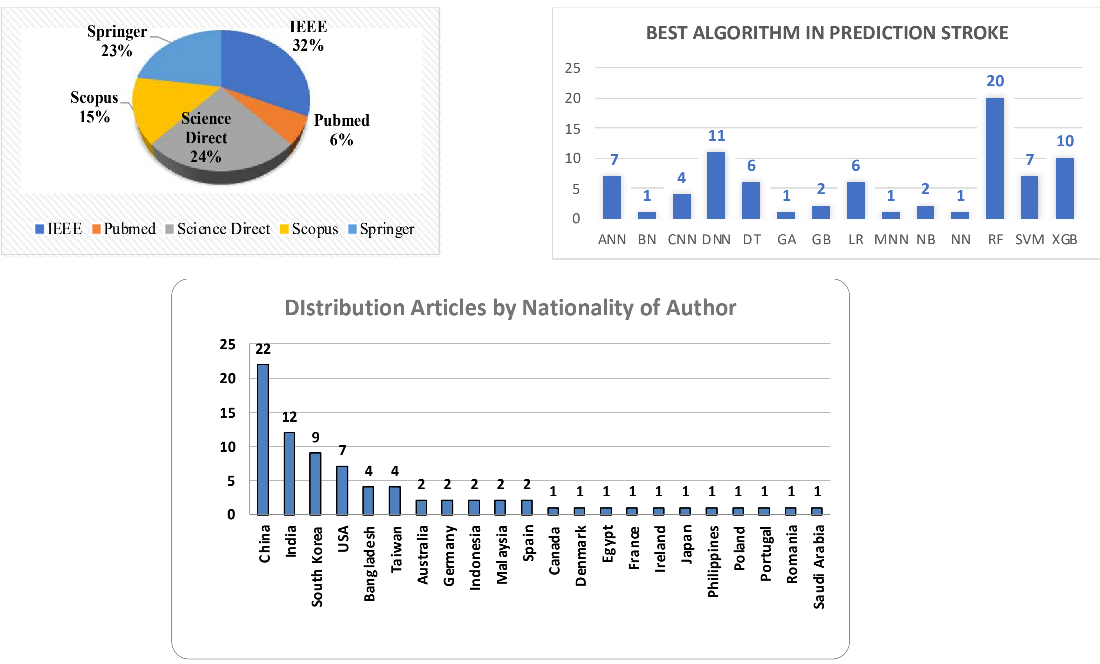

Fig. 2 Result demographics of research articles

Then, we analyze the distribution of stroke prediction articles based on the authors' nationalities. The 79 articles encompassed contributions from authors from 23 distinct nationalities, thereby underscoring the extensive scope of research on stroke prediction. China has emerged as the primary contributor, followed by India, South Korea, and the United States, in terms of occupying the leadership position. The graph presents a graphical depiction of the citation distribution among various databases and countries. The sources that contribute the most citations are the IEEE and Chinese databases. Moreover, this study underscores the crucial use of algorithms in the prognostication of strokes.

The graph illustrates the prevalence of support for the Random Forest (RF) algorithm, as indicated by 20 research studies that have validated its effectiveness. The Deep Neural Network (DNN) and XGBoost (XGB) also exhibit stability. Furthermore, the prediction of stroke is influenced by several methodologies, such as Artificial Neural Network (ANN), Support Vector Machine (SVM), Decision Tree (DT), Logistic Regression (LR), Convolutional Neural Network (CNN), Naive Bayes (NB), and Gradient Boosting (GB). In brief, this study offers valuable insights into the distribution patterns of sources, the global diversity of contributions, and the efficacy of algorithms in stroke prediction. The findings above contribute to the broader comprehension of stroke prediction research, a fundamental basis for developing and improving more effective and enduring prediction techniques.

B. Result in Taxonomy Literature Review of Research Stroke Prediction

This study has developed a comprehensive taxonomy based on analyzing 79 stroke prediction-related articles. The taxonomy presented in this study systematically organizes a wide range of methodological strategies, including Machine Learning (ML) and Deep Learning (DL) approaches. This research aimed to develop a comprehensive and precise taxonomy that effectively classifies and characterizes the various complex elements involved in implementing and using machine learning (ML) and deep learning (DL) methods in stroke prediction. The taxonomy presented here has been carefully developed to provide a well-organized framework that can assist researchers, physicians, and healthcare practitioners in understanding, evaluating, and advancing the field of stroke prediction using artificial intelligence.

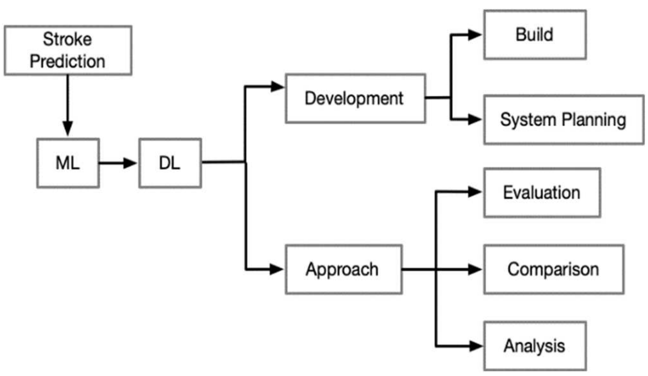

Fig. 3 Result Taxonomy from 79 articles

As illustrated in Figure 3, the taxonomy presented in this study is divided into two primary domains: Development and Approach. The primary focus of the first domain lies in constructing prediction models and systems' strategic planning. The second domain comprises the activities of evaluation, analysis, and model comparison. The combined impact of these specific research areas plays a crucial role in influencing the field of stroke prediction by employing machine learning and deep learning techniques.

Moreover, these areas of concentration encompass the core content of the 79 papers, which have been categorized into five separate study aims for each article, as delineated in Tables 1 to 5.

TABLE I
DEVELOPMENT IN BUILD MODEL

Year	Authors	Method/ Model	Dataset	Best Algorithm	Accuracy
2015	[26]	DL	25	CNN	85%
2016	[27]	LR with selection features combined (CFS, Lasso)	1864	LR	76%
2017	[17]	Data Mining and ML	0	BN	81%
2018	[28]	DL	222	CNN	85%
2019	[29]	ML with DL	43400	XGB	74%
2020	[30]	DL CNN+ANN	204	CNN	75%
2020	[31]	ML	3160	XGB	84%
2020	[32]	ML with Multiclass Classification	0	RF	88%
2020	[33]	ML	85	GB	90%
2020	[34]	ML	1131	LR	95%
2020	[35]	AI	5110	ANN	98%
2021	[36]	ML -DSS	1905	SVM	71%
2021	[37]	ML	1100	RF	71%
2021	[38]	ML	5110	NB	82%
2021	[39]	ML using SMOTE, Borderlines	3035	SVM	83%
2021	[40]	DL	100	DNN	85%
2021	[41]	ML	3127	RF	89%
2021	[42]	ML	19953	XGB	92%
2021	[43]	ML	4861	RF	94%
2021	[44]	DL	338	DNN	94%
2021	[45]	ML	5110	RF	96%
2021	[46]	ML	11	XGB	97%
2021	[47]	ML	5110	DT	98%
2021	[48]	DL ML	300	DNN	98%
2021	[49]	ML -SMOTE	5110	RF	99%
2022	[50]	ML DL - LASSO	3929	DNN	92%
2022	[51]	AI HYBRID	5110	RF	93%

The analysis of Table I reveals that significant progress has been made in developing stroke prediction models. The model represents a synthesis of many studies undertaken from 2015 to 2022. The purpose of these 27 articles was to develop machine learning (ML) and deep learning (DL) models for predicting and assessing the risk of stroke. The current study integrated research undertaken across different nations, wherein each country employed unique methodologies in developing their respective models. The outcomes derived from the execution of these algorithms exhibited a varied spectrum of accuracies, spanning from 71% to 99%. This study emphasizes the notable rise in applying machine learning (ML) and deep learning (DL) techniques within the healthcare sector. The effectiveness of these strategies is especially apparent in the areas of stroke prediction and risk management.

The main aim of constructing a machine learning model for the anticipation of stroke hospitalization, as outlined in Table I, is to establish a prognostic instrument that can accurately discern individuals with a heightened likelihood of being hospitalized for stroke [38], [49]. This model utilizes machine learning approaches to evaluate observable traits and symptoms, enabling the generation of predictions regarding the probability of stroke incidents. The main objective is to offer prompt identification and intervention for those who might be susceptible to stroke, to enhance patient results, and to alleviate the healthcare system's stroke-related load [48].

The study conducted by [34] did not specifically address the precise accuracy of machine learning models in predicting stroke within the senior population in China, particularly in the context of unbalanced data. While previous research has provided comprehensive information on different machine learning models and their predictive performance in various scenarios [35], the specific accuracy of these models in predicting stroke among the elderly Chinese population with imbalanced data has not been explicitly addressed [36].

TABLE II
DEVELOPMENT IN SYSTEM PLANNING TO BUILD A MODEL

Years	Authors	Method/Model	Dataset	Best Algorithm	Accuracy
2018	[52]	ML	350	ANN	98%
2019	[53]	Application ANN an NB	108	ANN	88%
2019	[54]	ML	287	DT	91%
2020	[55]	ML with analysis	227	DT	91%
2020	[56]	ML for tool extraction	507	ANN	95%
2021	[57]	ML	5110	LR	90%
2021	[58]	AI - ML	10000	XGB	99%

Table II presents a compilation of works that advance stroke prediction systems using machine learning methodologies. Many scholarly investigations have examined various models and datasets to improve the precision of stroke prediction. The table incorporates studies conducted in multiple countries, such as India [52], [56], [57], [58], Romania [53], and South Korea [54] [55]The experiments utilized various machine learning algorithms, including artificial neural networks (ANN), logistic regression (LR), and decision trees (DT). The accuracy of these models varied from 88% to 99%. The results shown in Table 2 highlight the importance of machine learning in stroke prediction and its potential to support healthcare practitioners in making educated and knowledgeable judgments.

Furthermore, the studies presented in Table II include a range of machine learning algorithms, such as artificial neural networks (ANN) [52], [58], decision trees (DT) [54], [55], and logistic regression (LR) [57]. It may be worthwhile to investigate and compare the performance of these algorithms. Understanding the advantages and disadvantages of each algorithm can provide researchers and practitioners with the knowledge necessary to make informed decisions regarding their utilization in various contexts. The findings underscore the use of physiological indicators and risk factors as inputs for predictive algorithms. Assessing the resilience and dependability of these indicators across diverse populations and healthcare contexts would yield significant benefits. Moreover, it is crucial to consider the ethical ramifications associated with using sensitive patient data for predictive systems.

Although these studies show promise, it is imperative to thoroughly evaluate the procedures and validation processes used to ensure that accuracy is neither overstated nor influenced by bias. To ascertain the reliability and generalizability of these prediction systems, replication studies, and external validation are essential. In conclusion, the findings presented in Table II provide significant contributions to understanding the development of stroke prediction systems using machine learning algorithms. Nevertheless, it is crucial to consider the inherent limitations present in this research, including geographical constraints, the performance of comparing algorithms, the resilience of parameters, and the necessity for rigorous validation. By taking into account these factors, researchers have the potential to optimize the effectiveness and practicality of stroke prediction algorithms within the context of clinical practice.

TABLE III
APPROACH IN ANALYSIS MODEL

Years	Authors	Method/Model	Dataset	Best Algorithm	Accuracy
2019	[59]	ML with XGB	6070	XGB	78%
2019	[60]	DL. DNN with PCA	15099	DNN	83%
2019	[61]	ML with PLS Optimation hybrid	1145	GA	86%
2019	[62]	ML	215	NB	88%
2019	[63]	Three cross-validation methods were examined: K- fold, Shuffle-Split, and Leave-One-Out. Using PCA	0	RF	93%
2019	[64]	ML Tenfold validation	119	RF	96%
2020	[65]	ML with Recursive Feature Elimination with Cross-Validation	1828	RF	80%
2020	[66]	ML	0	XGB	88%
2020	[67]	ML cross-validation and feature selection	58493	SVM	97%
2020	[68]	ML multi data fusion	79	MNN	98%
2020	[69]	ML with analysis	62001	RF	99%
2021	[70]	ML	660	ANN	80%
2021	[71]	ML	474	DNN	83%
2021	[72]	DL ABC Optimized	43400	DNN	87%
2021	[73]	ML	1056	DNN	90%
2021	[74]	ML -SMOTE	5110	SVM	94%
2021	[75]	ML	50	SVM	95%
2021	[76]	ML	6022	RF	98%
2022	[77]	ML - PCA	29072	DT	71%
2022	[78]	ML XGB,LR	41970	XGB	86%

Table III provides a detailed summary of multiple research studies that have utilized machine learning approaches to analyze and predict stroke outcomes. The studies above examine different categories of stroke, namely ischemic (12 articles) and hemorrhagic (8 articles). They employ various algorithms and datasets to attain a notable level of precision in forecasting hospitalization [59] stroke type, mortality/morbidity, and the requirement for ankle-foot orthosis in stroke patients [71]. The objective of this endeavor is to enhance the quality of patient care, treatment, and safety through the identification of risk factors, analysis of health behavior data, and investigation of the relationship between laboratory-based measurements and practical home usage among individuals who have experienced a stroke [63].

Table 3 provides a comprehensive overview of the many studies undertaken on stroke prediction and analysis using machine learning (ML) and deep learning (DL) approaches. This research aims to improve stroke diagnosis, treatment, and patient outcomes using various countries, datasets, and algorithms. A notable characteristic shared by these 20 publications is the incorporation of machine learning (ML) and deep learning (DL) models. These models have shown promising results in effectively forecasting instances of stroke and analyzing the associated risk variables. The potential of specific models in aiding healthcare practitioners in stroke prevention and management is underscored by the significant accuracy they have achieved, with reported rates ranging from 71% [77] to 98% [76].

Furthermore, despite the favorable results exhibited by machine learning (ML) and deep learning (DL) models, interpreting these models can be complex. Understanding the fundamental components and variables that contribute to the prediction and analysis of strokes can present difficulties, which may hinder the smooth integration of these models into clinical settings. Moreover, empirical investigations demonstrate considerable disparities in the sample sizes, spanning from a few hundred to tens of thousands of individuals. In general, increasing the sample size leads to more reliable results, which supports the need to conduct more research with larger groups of participants to strengthen the basis of evidence.

Nevertheless, it is essential to acknowledge that these 20 publications do not comprehensively examine the limitations and inherent biases linked to the utilized machine learning (ML) and deep learning (DL) methodologies. Investigating potential biases inherent in datasets, algorithmic biases, and the generalizability of models across varied populations and healthcare systems would provide advantageous outcomes. In summary, the papers included in this compilation provide light on the potential of machine learning (ML) and deep learning (DL) approaches in predicting and analyzing strokes. However, it is important to recognize and admit certain limitations.

TABLE IV
THE APPROACH IN EVALUATION MODEL

Years	Authors	Method/Model	Dataset	Best Algorithm	Accuracy
2014	[79]	DL SNN	82	ANN	95%
2019	[80]	ML and DL with Astral score	2604	DNN	88%
2019	[81]	Step-forward/backward ,MPNN	30082	RF	91%
2020	[82]	ML	614	LR	66%
2020	[83]	ML	328	XGB	80%
2020	[84]	AI	514	LR	81%
2020	[85]	ML Features extraction	450	LR	84%
2020	[86]	ML Improve Base Algorithm with Weighted Voting classifier	5110	XGB	96%
2020	[87]	ML	4799	RF	96%
2020	[88]	AI	92	RF	97%
2020	[89]	ML	43400	DT	98%
2020	[90]	ML subtype classification dataset	16636	RF	98%
2020	[91]	ML with DL	5110	DNN	99%
2021	[92]	DL	109	CNN	87%
2021	[93]	ML	5110	DT	95%

The 15 publications in Table IV explore the utilization of machine learning and artificial intelligence methodologies in predicting and classifying various stroke subtypes. These studies evaluate the accuracy and effectiveness of several algorithms in predicting the occurrence of strokes, categorizing different types of strokes, and predicting long-term results. The main aim of this study is to identify key characteristics that have a significant impact, develop models that can predict outcomes, and improve the decision-making process in stroke care. This compilation highlights the potential of machine learning approaches in enhancing stroke detection and improving patient outcomes.

Furthermore, this study introduces a data-driven methodological framework that aims to assess several feature selection methods within the specific context of predicting outcomes for patients with ischemic stroke. The table provides a comprehensive overview of studies completed in different nations, utilizing machine learning (ML) and deep learning (DL) methodologies to predict stroke outcomes. One notable attribute of these 15 articles is their extensive geographical scope, which enables a broader understanding of the subject matter. The incorporation of research conducted in various countries, including Taiwan [79], South Korea [80], the United States [81], Indonesia [88], China [90], India [93], and France [92], highlights the global interest in utilizing machine learning (ML) and deep learning (DL) algorithms to predict stroke outcomes.

Furthermore, a wide range of machine learning (ML) and deep learning (DL) techniques have been utilized in these studies. These techniques include artificial neural networks (JST) [79], deep neural networks (DNN) [80] random forests (RF) [81], [87], [88], [94], decision trees (DT), [89], [93], and logistic regression (LR))[82], [84], [85]. Utilizing these various algorithms contributes to a more comprehensive understanding of prediction accuracy. However, it is essential to note that this compilation may have a possible disadvantage because it lacks a detailed examination or comparison of the methodology employed in the studies. Although the chart provides an overview of the maximum algorithmic accuracy attained in each study, it does not thoroughly examine the distinct merits and limitations of different approaches. The limitations imposed on the 15 articles hinder their ability to evaluate the various methodologies comprehensively.

TABLE V
THE APPROACH IN MODEL COMPARISON

Years	Authors	Method/Model	Dataset	Best Algorithm	Accuracy
2014	[95]	ML	107	SVM	70%
2014	[96]	SMOTE With ML,Chi-Square	778	SVM	87%
2017	[97]	DL Comparison with DL	900000	DNN	87%
2017	[98]	A.I with PCA	1824	NN	95%
2018	[99]	ML Tenfold validation using 5 times experiments	425	RF	94%
2019	[100]	ML	5116	RF	90%
2021	[101]	ML	503842	GB	83%
2021	[102]	ML using feature selection	2555	ANN	87%
2022	[94]	DL ML RF DNN	16403	RF	83%
2022	[103]	ML - SMOTE	5110	RF	97%

Table V presents a comprehensive comparative review of the predictive capabilities of several machine learning algorithms about stroke risk and outcomes. This compilation comprises studies conducted in several nations that utilize different datasets and models. The primary aim of this study was to determine the most accurate approach for predicting the occurrence of strokes and assessing functional results. This was achieved by examining the utilization of structured data and deep learning techniques in various research investigations. Overall, the research findings indicate favorable results in utilizing machine learning techniques to predict strokes. This collection of ten articles is a compendium of studies that compare and contrast various methodologies and models for predicting outcomes and hazards associated with ischemic stroke. The research encompasses a range of machine learning (ML) and deep learning (DL) algorithms, such as Support Vector Machines (SVM), Random Forest (RF), Artificial Neural Networks (ANN), and Deep Neural Networks (DNN), among other methods.

One significant advantage of these studies is their utilization of extensive datasets, which enables a more thorough examination and rigorous assessment of model performance. For example, the investigations carried out by Ho et al. [96] and Chun et al. [101] utilized extensive datasets with many patients, establishing a robust basis for their conclusions. Another notable advantage is the ability to compare different algorithms and models directly. This methodology enables researchers to identify the most precise and effective methodologies for forecasting stroke outcomes. Gupta et al. [103] conducted a comparative analysis of various machine learning models, wherein they obtained a notable accuracy rate of 97% using the Random Forest classifier. Nevertheless, it is crucial to consider several constraints. The main emphasis of these studies was primarily directed towards the prediction of ischemic stroke, which may restrict the applicability of the results to other forms of stroke or different populations. Furthermore, using diverse datasets from various studies introduces complexity when attempting direct comparisons between models.

IV. Conclusion

The substantial influence of stroke on health outcomes highlights the pressing need for the implementation of effective interventions to mitigate its effects. The utilization of artificial intelligence (AI) techniques, particularly machine learning (ML) and deep learning (DL), exhibits significant potential in addressing this particular difficulty. This study utilized the PRISMA Method to examine 79 scholarly publications published between 2012 and 2022. The primary objective was to investigate the application of machine learning algorithms in the context of stroke prediction. The results highlight the potential of machine learning methods in providing accurate predictions for stroke outcomes. Nevertheless, challenges such as biased data and algorithms and the need for models that can be adjusted to accommodate various demographics and healthcare systems continue to exist.

To fully harness the capabilities of artificial intelligence (AI) in stroke prevention, it is imperative to adopt multidisciplinary frameworks, maintain unshakable dedication, and continuously develop our approach. The results of this study offer insight into the progression of advancements and stimulate additional investigation aimed at enhancing the precision and influence of stroke prediction by integrating machine learning and artificial intelligence. Continuing research endeavors necessitate addressing the issues of data bias and algorithmic behavior to establish prediction models that are grounded in principles of fairness, neutrality, and suitability across diverse circumstances. This undertaking will provide a valuable contribution to advancing scholarly literature and developing reliable and robust machine-learning models that can accurately predict the incidence of strokes.

Acknowledgment

The authors wish to extend their gratitude to Universitas Sari Mulia and Universiti Pendidikan Sultan Idris for their sponsorship and financial support.

References

[1] B. Hofstra, V. V. Kulkarni, S. M. N. Galvez, B. He, D. Jurafsky, and D. A. McFarland, "The diversity–innovation paradox in science," *Proc Natl Acad Sci U S A*, vol. 117, no. 17, 2020, doi:10.1073/pnas.1915378117.

[2] E. S. Donkor, "Stroke in the 21st Century: A Snapshot of the Burden, Epidemiology, and Quality of Life," *Stroke Res Treat*, vol. 2018, 2018, doi: 10.1155/2018/3238165.

[3] D. Kuriakose and Z. Xiao, "Pathophysiology and treatment of stroke: Present status and future perspectives," *International Journal of Molecular Sciences*, vol. 21, no. 20. 2020. doi: 10.3390/ijms21207609.

[4] J. Benito-Lozano *et al.*, "Diagnostic Process in Rare Diseases: Determinants Associated with Diagnostic Delay," *Int J Environ Res Public Health*, vol. 19, no. 11, 2022, doi: 10.3390/ijerph19116456.

[5] D. Richards, J. A. Morren, and E. P. Pioro, "Time to diagnosis and factors affecting diagnostic delay in amyotrophic lateral sclerosis," *Journal of the Neurological Sciences*, vol. 417. 2020. doi:10.1016/j.jns.2020.117054.

[6] J. L. Clarke, S. Bourn, A. Skoufalos, E. H. Beck, and D. J. Castillo, "An Innovative Approach to Health Care Delivery for Patients with Chronic Conditions," *Popul Health Manag*, vol. 20, no. 1, 2017, doi:10.1089/pop.2016.0076.

[7] J. Bajwa, U. Munir, A. Nori, and B. Williams, "Artificial intelligence in healthcare: transforming the practice of medicine," *Future Healthc J*, vol. 8, no. 2, 2021, doi: 10.7861/fhj.2021-0095.

[8] R. T.-A. Rahman, M.-M. Lakulu, and I.-Y. Panessai, "Advancing Preeclampsia Prediction with Machine Learning: A Comprehensive Systematic Literature Review", *Int J Intell Syst Appl Eng*, vol. 11, no. 3, pp. 13–23, Jul. 2023.

[9] W. J. Powers *et al.*, "2018 Guidelines for the Early Management of Patients With Acute Ischemic Stroke: A Guideline for Healthcare Professionals From the American Heart Association/American Stroke Association," *Stroke*, 2018, doi: 10.1161/STR.0000000000000158.

[10] K. Paul Buchan Jr, "Using machine learning to predict super-utilizers of healthcare," *Legacy Theses & Dissertations (2009-2024)*, 2021.

[11] Syarfaini, Nildawati, S. Aeni, Surahmawati, A. S. Adha, and M. Amansyah, "Risk factors preparation of stroke incidence in health institution employees who check up at the Health Service EXPO Event Indonesia," *Gac Sanit*, vol. 35, pp. S49–S52, 2021, doi:10.1016/j.gaceta.2020.12.014.

[12] Y. Wang, "Toward Better Health Care Service: Statistical and Machine Learning Based Analysis of Swedish Patient Satisfaction Survey," 2017.

[13] S. E. Chiuve, K. M. Rexrode, D. Spiegelman, G. Logroscino, J. E. Manson, and E. B. Rimm, "Primary prevention of stroke by healthy lifestyle," *Circulation*, vol. 118, no. 9, 2008, doi:10.1161/circulationaha.108.781062.

[14] E. S. Connolly *et al.*, "Guidelines for the management of aneurysmal subarachnoid hemorrhage: A guideline for healthcare professionals from the american heart association/american stroke association," *Stroke*, vol. 43, no. 6. 2012. doi: 10.1161/str.0b013e3182587839.

[15] H. J. Lee, E. K. Choi, S. H. Lee, Y. J. Kim, K. Do Han, and S. Oh, "Risk of ischemic stroke in metabolically healthy obesity: A nationwide population-based study," *PLoS One*, vol. 13, no. 3, 2018, doi: 10.1371/journal.pone.0195210.

[16] Sunarti, "Agenda Setting Pemberitaan Covid 19 dan Pelarangan Mudik pada Media Online di Indonesia," 2021.

[17] A. R. Al Taleb, M. Hoque, A. Hasanat, and M. B. Khan, "Application of data mining techniques to predict length of stay of stroke patients," in *2017 International Conference on Informatics, Health & Technology (ICIHT)*, IEEE, Feb. 2017, pp. 1–5. doi:10.1109/iciht.2017.7899004.

[18] F. Wang *et al.*, "Personalized risk prediction of symptomatic intracerebral hemorrhage after stroke thrombolysis using a machine-learning model," *Ther Adv Neurol Disord*, vol. 13, p. 175628642090235, Jan. 2020, doi: 10.1177/1756286420902358.

[19] M. S. Sirsat, E. Fermé, and J. Câmara, "Machine Learning for Brain Stroke: A Review.," *J Stroke Cerebrovasc Dis*, vol. 29, no. 10, p. 105162, Oct. 2020, doi: 10.1016/j.jstrokecerebrovasdis.2020.105162.

[20] L. Schwartz, R. Anteby, E. Klang, and S. Soffer, "Stroke mortality prediction using machine learning: systematic review," *J Neurol Sci*, vol. 444, p. 120529, Jan. 2023, doi: 10.1016/j.jns.2022.120529.

[21] A. Panesar, *Machine Learning and AI for Healthcare: Big Data for Improved Health Outcomes*. 2019. doi: 10.1007/978-1-4842-3799-1.

[22] F. Jiang *et al.*, "Artificial intelligence in healthcare: Past, present and future," Stroke *and Vascular Neurology*, vol. 2, no. 4. BMJ Publishing Group, pp. 230–243, Dec. 01, 2017. doi: 10.1136/svn-2017-000101.

[23] A. A. Selcuk, "A Guide for Systematic Reviews: PRISMA," *Turk Arch Otorhinolaryngol*, vol. 57, no. 1, pp. 57–58, May 2019, doi:10.5152/tao.2019.4058.

[24] M. L. Rethlefsen *et al.*, "PRISMA-S: an extension to the PRISMA Statement for Reporting Literature Searches in Systematic Reviews," *Syst Rev*, vol. 10, no. 1, p. 39, Jan. 2021, doi: 10.1186/s13643-020-01542-z.

[25] R. Mohamed, M. Ghazali, and M. A. Samsudin, "A Systematic Review on Mathematical Language Learning Using PRISMA in Scopus Database," *Eurasia Journal of Mathematics, Science and Technology Education*, vol. 16, no. 8, p. em1868, May 2020, doi:10.29333/ejmste/8300.

[26] N. Stier, N. Vincent, D. Liebeskind, and F. Scalzo, "Deep learning of tissue fate features in acute ischemic stroke," in *2015 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 2015, pp. 1316–1321. doi: 10.1109/BIBM.2015.7359869.

[27] X. Li *et al.*, "Integrated Machine Learning Approaches for Predicting Ischemic Stroke and Thromboembolism in Atrial Fibrillation," *AMIA Annual Symposium Proceedings*, vol. 2016, pp. 799–807, Feb. 2017.

[28] A. Nielsen, M. B. Hansen, A. Tietze, and K. Mouridsen, "Prediction of Tissue Outcome and Assessment of Treatment Effect in Acute Ischemic Stroke Using Deep Learning," *Stroke*, vol. 49, no. 6, pp. 1394–1401, Jun. 2018, doi: 10.1161/STROKEAHA.117.019740.

[29] T. Liu, W. Fan, and C. Wu, "A hybrid machine learning approach to cerebral stroke prediction based on imbalanced medical dataset," *Artif Intell Med*, vol. 101, p. 101723, 2019, doi:10.1016/j.artmed.2019.101723.

[30] S. Bacchi, T. Zerner, L. Oakden-Rayner, T. Kleinig, S. Patel, and J. Jannes, "Deep Learning in the Prediction of Ischaemic Stroke Thrombolysis Functional Outcomes: A Pilot Study," *Acad Radiol*, vol. 27, no. 2, 2020, doi: 10.1016/j.acra.2019.03.015.

[31] X. Li *et al.*, "Using machine learning to predict stroke-associated pneumonia in Chinese acute ischaemic stroke patients," *Eur J Neurol*, vol. 27, no. 8, pp. 1656–1663, Aug. 2020, doi: 10.1111/ene.14295.

[32] N. J. Apao, L. S. Feliscuzo, C. L. C. Sta. Romana, and J. Tagaro, "Multiclass Classification Using Random Forest Algorithm to Prognosticate The Level Of Activity Of Patients With Stroke," *International Journal of Scientific & Technology Research*, vol. 9, pp. 1233–1240, 2020.

[33] X. Chen, C. Wei, W. Wu, L. Guo, C. Liu, and G. Lu, "Based on Machine Learning Algorithm: Construction of an Early Prediction Model of Integrated Traditional Chinese and Western Medicine for Cognitive Impairment after Ischemic Stroke," in *2020 5th International Conference on Universal Village (UV)*, IEEE, Oct. 2020, pp. 1–7. doi: 10.1109/UV50937.2020.9426200.

[34] Y. Wu and Y. Fang, "Stroke Prediction with Machine Learning Methods among Older Chinese," *Int J Environ Res Public Health*, vol. 17, no. 6, p. 1828, Mar. 2020, doi: 10.3390/ijerph17061828.

[35] C.-C. Peng, S.-H. Wang, S.-J. Liu, Y.-K. Yang, and B.-H. Liao, "Artificial Neural Network Application to the Stroke Prediction," in *2020 IEEE 2nd Eurasia Conference on Biomedical Engineering, Healthcare and Sustainability (ECBIOS)*, IEEE, May 2020, pp. 130–133. doi: 10.1109/ecbios50299.2020.9203638.

[36] X. Lin *et al.*, "Prediction-Driven Decision Support for Patients with Mild Stroke: A Model Based on Machine Learning Algorithms," *Front Neurol*, vol. 12, Dec. 2021, doi: 10.3389/fneur.2021.761092.

[37] J. Cui *et al.*, "Machine Learning-Based Model for Predicting Incidence and Severity of Acute Ischemic Stroke in Anterior Circulation Large Vessel Occlusion," *Front Neurol*, vol. 12, Dec. 2021, doi:10.3389/fneur.2021.749599.

[38] G. Sailasya and G. L. A. Kumari, "Analyzing the Performance of Stroke Prediction using ML Classification Algorithms," *International Journal of Advanced Computer Science and Applications*, vol. 12, no. 6, pp. 539–545, 2021, doi: 10.14569/ijacsa.2021.0120662.

[39] Q. Qin, X. Zhou, and Y. Jiang, "Prognosis Prediction of Stroke based on Machine Learning and Explanation Model," *International Journal of Computers Communications & Control*, vol. 16, no. 2, pp. 1–13, Mar. 2021, doi: 10.15837/ijccc.2021.2.4108.

[40] Y. Xie *et al.*, "Stroke prediction from electrocardiograms by deep neural network," *Multimed Tools Appl*, vol. 80, no. 11, pp. 17291–17297, May 2021, doi: 10.1007/s11042-020-10043-z.

[41] K. Uchida *et al.*, "Development of Machine Learning Models to Predict Probabilities and Types of Stroke at Prehospital Stage: the Japan Urgent Stroke Triage Score Using Machine Learning (JUST-ML)," *Transl Stroke Res*, Aug. 2021, doi: 10.1007/s12975-021-00937-x.

[42] Y. Yang, J. Zheng, Z. Du, Y. Li, and Y. Cai, "Accurate prediction of stroke for hypertensive patients based on medical big data and machine learning algorithms: Retrospective study," *JMIR Med Inform*, vol. 9, no. 11, p. e30277, Nov. 2021, doi: 10.2196/30277.

[43] N. S. Adi, R. Farhany, R. Ghina, and H. Napitupulu, "Stroke Risk Prediction Model Using Machine Learning," in *2021 International Conference on Artificial Intelligence and Big Data Analytics*, IEEE, Oct. 2021, pp. 56–60. doi: 10.1109/icaibda53487.2021.9689731.

[44] L. Jiang *et al.*, "A deep learning-based model for prediction of hemorrhagic transformation after stroke," *Brain Pathology*, Oct. 2021, doi: 10.1111/bpa.13023.

[45] T. Tazin, M. N. Alam, N. N. Dola, M. S. Bari, S. Bourouis, and M. Monirujjaman Khan, "Stroke Disease Detection and Prediction Using Robust Learning Approaches.," *J Healthc Eng*, vol. 2021, p. 7633381, Nov. 2021, doi: 10.1155/2021/7633381.

[46] P.-W. Chen, N. A. Baune, I. Zwir, J. Wang, V. Swamidass, and A. W. K. Wong, "Measuring Activities of Daily Living in Stroke Patients with Motion Machine Learning Algorithms: A Pilot Study," *Int J Environ Res Public Health*, vol. 18, no. 4, p. 1634, Feb. 2021, doi:10.3390/ijerph18041634.

[47] R. K. Kavitha, W. Jaisingh, and S. R. Sujithra, "Applying Machine Learning Techniques for Stroke Prediction in Patients," in *2021 International Conference on Advancements in Electrical, Electronics, Communication, Computing and Automation (ICAECA)*, IEEE, Oct. 2021, pp. 1–4. doi: 10.1109/icaeca52838.2021.9675652.

[48] H. Zhao *et al.*, "The Construction of a Risk Prediction Model Based on Neural Network for Pre-operative Acute Ischemic Stroke in Acute Type A Aortic Dissection Patients," *Front Neurol*, vol. 12, Dec. 2021, doi: 10.3389/fneur.2021.792678.

[49] Ferdib-Al-Islam and M. Ghosh, "An Enhanced Stroke Prediction Scheme Using SMOTE and Machine Learning Techniques," in *2021 12th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, IEEE, Jul. 2021, pp. 1–6. doi:10.1109/icccnt51525.2021.9579648.

[50] X. Zheng *et al.*, "Using machine learning to predict atrial fibrillation diagnosed after ischemic stroke," *Int J Cardiol*, vol. 347, pp. 21–27, Jan. 2022, doi: 10.1016/j.ijcard.2021.11.005.

[51] A. Patil and G. Suresh Kumar, "An Artificial Intelligence Enabled Framework with Hybrid Feature Selection Method For Efficient Early Detection Of Stroke," *J Theor Appl Inf Technol*, vol. 15, no. 3, pp. 813–824, 2022.

[52] R. S. Jeena and S. Kumar, "Machine intelligence in stroke prediction," *Int J Bioinform Res Appl*, vol. 14, no. 1/2, p. 29, 2018, doi:10.1504/ijbra.2018.089192.

[53] A. Albu, L. Stanciu, M.-S. Pasca, and C.-G. Zimbru, "Choosing Between Artificial Neural Networks and Bayesian Inference in Stroke Risk Prediction," in *2019 E-Health and Bioengineering Conference (EHB)*, IEEE, Nov. 2019, pp. 1–4. doi:10.1109/ehb47216.2019.8970035.

[54] J. Yu *et al.*, "Semantic Analysis of NIH Stroke Scale using Machine Learning Techniques," in *2019 International Conference on Platform Technology and Service, PlatCon 2019 - Proceedings*, IEEE, Jan. 2019, pp. 1–5. doi: 10.1109/platcon.2019.8668961.

[55] J. Yu, S. Park, H. Lee, C.-S. Pyo, and Y. S. Lee, "An Elderly Health Monitoring System Using Machine Learning and In-Depth Analysis Techniques on the NIH Stroke Scale," *Mathematics*, vol. 8, no. 7, p. 1115, Jul. 2020, doi: 10.3390/math8071115.

[56] P. Govindarajan, R. K. Soundarapandian, A. H. Gandomi, R. Patan, P. Jayaraman, and R. Manikandan, "Classification of stroke disease using machine learning algorithms," *Neural Comput Appl*, vol. 32, no. 3, pp. 817–828, Feb. 2020, doi: 10.1007/s00521-019-04041-y.

[57] C. D. Anisha and K. G. Saranya, "Early diagnosis of stroke disorder using homogenous logistic regression ensemble classifier," *International Journal of Nonlinear Analysis and Applications*, vol. 12, no. Special Issue, pp. 1649–1654, 2021, doi:10.22075/ijnaa.2021.5851.

[58] P. A, N. G, V. K. R, P. P, and S. R. R.V.T, "Stroke Prediction System Using Artificial Neural Network," in *2021 6th International Conference on Communication and Electronics Systems (ICCES)*,

IEEE, Jul. 2021, pp. 1898–1902. doi:10.1109/icces51350.2021.9489055.

[59] Y. Xu *et al.*, "Extreme Gradient Boosting Model Has a Better Performance in Predicting the Risk of 90-Day Readmissions in Patients with Ischaemic Stroke," *Journal of Stroke and Cerebrovascular Diseases*, vol. 28, no. 12, p. 104441, Dec. 2019, doi:10.1016/j.jstrokecerebrovasdis.2019.104441.

[60] S. Cheon, J. Kim, and J. Lim, "The use of deep learning to predict stroke patient mortality," *Int J Environ Res Public Health*, vol. 16, no. 11, 2019, doi: 10.3390/ijerph16111876.

[61] C. Shih, W. Cheng-Chung Chu, and Y.-W. Chang, "The Causes Analysis of Ischemic Stroke Transformation into Hemorrhagic Stroke using PLS (partial Least Square)-GA and Swarm Algorithm," in *2019 IEEE 43rd Annual Computer Software and Applications Conference (COMPSAC)*, IEEE, Jul. 2019, pp. 720–729. doi:10.1109/compsac.2019.00108.

[62] M. Zdrodowska, "Attribute Selection for Stroke Prediction," *Acta Mechanica et Automatica*, vol. 13, no. 3, pp. 200–204, Sep. 2019, doi:10.2478/ama-2019-0026.

[63] I. Almubark *et al.*, "Machine Learning Approaches to Predict Functional Upper Extremity Use in Individuals with Stroke," *Proceedings - 2018 IEEE International Conference on Big Data, Big Data 2018*, pp. 5291–5294, 2019, doi: 10.1109/bigdata.2018.8622054.

[64] L. García-Terriza *et al.*, "Comparison of Different Machine Learning Approaches to Model Stroke Subtype Classification and Risk Prediction," in *Modeling and Simulation in Medicine (MSM 2019)*, Society for Modeling and Simulation International (SCS), Apr. 2019, pp. 1–10. doi: 10.22360/SpringSim.2019.MSM.006.

[65] G. Fang, W. Liu, and L. Wang, "A machine learning approach to select features important to stroke prognosis," *Comput Biol Chem*, vol. 88, p. 107316, Oct. 2020, doi: 10.1016/j.compbiolchem.2020.107316.

[66] M. Grosser *et al.*, "Improved multi-parametric prediction of tissue outcome in acute ischemic stroke patients using spatial features," *PLoS One*, vol. 15, no. 1, p. e0228113, Jan. 2020, doi:10.1371/journal.pone.0228113.

[67] C.-L. C.-H. Lin *et al.*, "Evaluation of machine learning methods to stroke outcome prediction using a nationwide disease registry," *Comput Methods Programs Biomed*, vol. 190, p. 105381, Jul. 2020, doi: 10.1016/j.cmpb.2020.105381.

[68] Y. Liu, B. Yin, and Y. Cong, "The Probability of Ischaemic Stroke Prediction with a Multi-Neural-Network Model," *Sensors*, vol. 20, no. 17, p. 4995, Sep. 2020, doi: 10.3390/s20174995.

[69] Md. Shafiul Azam, Md. Habibullah, and H. Kabir Rana, "Performance Analysis of Various Machine Learning Approaches in Stroke Prediction," *Int J Comput Appl*, vol. 175, no. 21, pp. 11–15, Sep. 2020, doi: 10.5120/ijca2020920740.

[70] F. H. Hassan and M. A. Omar, "Recurrent Stroke Prediction using Machine Learning Algorithms with Clinical Public Datasets: An Empirical Performance Evaluation," *Baghdad Science Journal*, vol. 18, no. 4(Suppl.), p. 1406, Dec. 2021, doi:10.21123/bsj.2021.18.4(Suppl.).1406.

[71] Y. J. Choo, J. H. J. K. Kim, J. H. J. K. Kim, M. C. Chang, and D. Park, "Machine learning analysis to predict the need for ankle foot orthosis in patients with stroke," *Sci Rep*, vol. 11, no. 1, p. 8499, Dec. 2021, doi: 10.1038/s41598-021-87826-3.

[72] A. Dev and S. K. Malik, "Artificial Bee Colony Optimized Deep Neural Network Model for Handling Imbalanced Stroke Data," *International Journal of E-Health and Medical Communications*, vol. 12, no. 5, pp. 67–83, Sep. 2021, doi: 10.4018/IJEHMC.20210901.oa5.

[73] J. K. Kim, Y. J. Choo, and M. C. Chang, "Prediction of Motor Function in Stroke Patients Using Machine Learning Algorithm: Development of Practical Models," *Journal of Stroke and Cerebrovascular Diseases*, vol. 30, no. 8, p. 105856, Aug. 2021, doi:10.1016/j.jstrokecerebrovasdis.2021.105856.

[74] C. Rana, N. Chitre, B. Poyekar, and P. Bide, "Stroke Prediction Using Smote-Tomek and Neural Network," in *2021 12th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, IEEE, Jul. 2021, pp. 1–5. doi:10.1109/icccnt51525.2021.9579763.

[75] X. Zhao *et al.*, "Machine Learning Analysis of MicroRNA Expression Data Reveals Novel Diagnostic Biomarker for Ischemic Stroke," *Journal of Stroke and Cerebrovascular Diseases*, vol. 30, no. 8, p. 105825, Aug. 2021, doi: 10.1016/j.jstrokecerebrovasdis.2021.105825.

[76] C. Fernandez-Lozano *et al.*, "Random forest-based prediction of stroke outcome," *Sci Rep*, vol. 11, no. 1, p. 10071, Dec. 2021, doi:10.1038/s41598-021-89434-7.

[77] S. Dev, H. Wang, C. S. Nwosu, N. Jain, B. Veeravalli, and D. John, "A predictive analytics approach for stroke prediction using machine learning and neural networks," *Healthcare Analytics*, vol. 2, no. February, p. 100032, Nov. 2022, doi: 10.1016/j.health.2022.100032.

[78] J. Maharjan *et al.*, "Enriching the Study Population for Ischemic Stroke Therapeutic Trials Using a Machine Learning Algorithm," *Front Neurol*, vol. 12, Jan. 2022, doi: 10.3389/fneur.2021.784250.

[79] C.-A. Cheng, Y.-C. Lin, and H.-W. Chiu, "Prediction of the prognosis of ischemic stroke patients after intravenous thrombolysis using artificial neural networks," in *Studies in Health Technology and Informatics*, Athens: IOS Press, 2014, pp. 115–118. doi: 10.3233/978-1-61499-423-7-115.

[80] J. H. J. Heo, J. G. Yoon, H. Park, Y. D. Kim, H. S. Nam, and J. H. J. Heo, "Machine Learning–Based Model for Prediction of Outcomes in Acute Stroke," *Stroke*, vol. 50, no. 5, pp. 1263–1265, May 2019, doi:10.1161/strokeaha.118.024293.

[81] T. Clifford, J. Bruce, T. Obafemi-Ajayi, and J. Matta, "Comparative Analysis of Feature Selection Methods to Identify Biomarkers in a Stroke-Related Dataset," in *2019 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB)*, D. C. B. F. S. G. E. D. N. G. Baruzzo G. Daberdaku S., Ed., IEEE, Jul. 2019, pp. 1–8. doi: 10.1109/CIBCB.2019.8791457.

[82] S. A. Alaka *et al.*, "Functional Outcome Prediction in Ischemic Stroke: A Comparison of Machine Learning Algorithms and Regression Models," *Front Neurol*, vol. 11, Aug. 2020, doi:10.3389/fneur.2020.00889.

[83] C. Kim *et al.*, "Impact of 25-Hydroxyvitamin D on the Prognosis of Acute Ischemic Stroke: Machine Learning Approach," *Front Neurol*, vol. 11, Jan. 2020, doi: 10.3389/fneur.2020.00037.

[84] E. Zihni *et al.*, "Opening the black box of artificial intelligence for clinical decision support: A study predicting stroke outcome.," *PLoS One*, vol. 15, no. 4, p. e0231166, Apr. 2020, doi:10.1371/journal.pone.0231166.

[85] L. Yang, Q. Liu, Q. Zhao, X. Zhu, and L. Wang, "Machine learning is a valid method for predicting prehospital delay after acute ischemic stroke," *Brain Behav*, vol. 10, no. 10, Oct. 2020, doi:10.1002/brb3.1794.

[86] M. U. Emon, M. S. Keya, T. I. Meghla, M. M. Rahman, M. S. Al Mamun, and M. S. Kaiser, "Performance Analysis of Machine Learning Approaches in Stroke Prediction," in *2020 4th International Conference on Electronics, Communication and Aerospace Technology (ICECA)*, IEEE, Nov. 2020, pp. 1464–1469. doi:10.1109/iceca49313.2020.9297525.

[87] V. Bandi, D. Bhattacharyya, and D. Midhunchakkravarthy, "Prediction of brain stroke severity using machine learning," *Revue d'Intelligence Artificielle*, vol. 34, no. 6, pp. 753–761, Dec. 2020, doi:10.18280/RIA.340609.

[88] G. S. Saragih, Z. Rustam, D. Aldila, R. Hidayat, R. E. Yunus, and J. Pandelaki, "Ischemic Stroke Classification using Random Forests Based on Feature Extraction of Convolutional Neural Networks," *Int J Adv Sci Eng Inf Technol*, vol. 10, no. 5, p. 2177, Oct. 2020, doi:10.18517/ijaseit.10.5.13000.

[89] Y. C. Choi, "Stroke Prediction Using Machine Learning based on Artificial Intelligence," *International Journal of Advanced Trends in Computer Science and Engineering*, vol. 9, no. 5, pp. 8916–8921, Oct. 2020, doi: 10.30534/ijatcse/2020/289952020.

[90] G. Fang, P. Xu, and W. Liu, "Automated Ischemic Stroke Subtyping Based on Machine Learning Approach," *IEEE Access*, vol. 8, pp. 118426–118432, 2020, doi: 10.1109/ACCESS.2020.3004977.

[91] T. R. G *et al.*, "Antlion re-sampling based deep neural network model for classification of imbalanced multimodal stroke dataset," *Multimed Tools Appl*, Oct. 2020, doi: 10.1007/s11042-020-09988-y.

[92] N. Debs *et al.*, "Impact of the reperfusion status for predicting the final stroke infarct using deep learning," *Neuroimage Clin*, vol. 29, p. 102548, 2021, doi: 10.1016/j.nicl.2020.102548.

[93] V. JalajaJayalakshmi, V. Geetha, and M. M. Ijaz, "Analysis and Prediction of Stroke using Machine Learning Algorithms," in *2021 International Conference on Advancements in Electrical, Electronics, Communication, Computing and Automation (ICAECA)*, IEEE, Oct. 2021, pp. 1–5. doi: 10.1109/ICAECA52838.2021.9675545.

[94] G. Fang, Z. Huang, and Z. Wang, "Predicting Ischemic Stroke Outcome Using Deep Learning Approaches," *Front Genet*, vol. 12, no. January, pp. 1–8, Jan. 2022, doi: 10.3389/fgene.2021.827522.

[95] H. Asadi, R. Dowling, B. Yan, and P. Mitchell, "Machine Learning for Outcome Prediction of Acute Ischemic Stroke Post Intra-Arterial Therapy," *PLoS One*, vol. 9, no. 2, p. e88225, Feb. 2014, doi:10.1371/journal.pone.0088225.

[96] K. C. Ho *et al.*, “Predicting Discharge Mortality after Acute Ischemic Stroke Using Balanced Data,” *AMIA Annual Symposium Proceedings*, vol. 2014, pp. 1787–1796, Nov. 2014.

[97] Chen-Ying Hung *et al.*, “Comparing deep neural network and other machine learning algorithms for stroke prediction in a large-scale population-based electronic medical claims database,” *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, vol. 2017, pp. 3110–3113, Jul. 2017, doi: 10.1109/EMBC.2017.8037515.

[98] M. S. Singh and P. Choudhary, “Stroke prediction using artificial intelligence,” in *2017 8th Annual Industrial Automation and Electromechanical Engineering Conference (IEMECON)*, IEEE, Aug. 2017, pp. 158–161. doi: 10.1109/IEMECON.2017.8079581.

[99] M. Monteiro *et al.*, “Using Machine Learning to Improve the Prediction of Functional Outcome in Ischemic Stroke Patients.,” *IEEE/ACM Trans Comput Biol Bioinform*, vol. 15, no. 6, pp. 1953–1959, Nov. 2018, doi: 10.1109/TCBB.2018.2811471.

[100] H. A. S. F. A.- el ghany, E. M. Nahla F.Omran, Abdelmgeid A.Ali, “Stroke Prediction using Distributed Machine Learning Based on Apache Spark”, *IJAST*, vol. 28, no. 15, pp. 89 - 97, Nov. 2019.

[101] M. Chun *et al.*, “Stroke risk prediction using machine learning: a prospective cohort study of 0.5 million Chinese adults,” *Journal of the American Medical Informatics Association*, vol. 28, no. 8, pp. 1719–1727, Jul. 2021, doi: 10.1093/jamia/ocab068.

[102] J.-M. Choi *et al.*, “Prediction of Hemorrhagic Transformation after Ischemic Stroke Using Machine Learning,” *J Pers Med*, vol. 11, no. 9, p. 863, Aug. 2021, doi: 10.3390/jpm11090863.

[103] S. Gupta and S. Raheja, “Stroke Prediction using Machine Learning Methods,” in *2022 12th International Conference on Cloud Computing, Data Science & Engineering (Confluence)*, IEEE, Jan. 2022, pp. 553–558. doi: 10.1109/Confluence52989.2022.9734197.